Neural Information Processing Systems Conference 2010

Vancouver,
British Columbia
December 6 - 9th

Neural Information
Processing Systems
Foundation

2010 Conference Book

NIPS 2010

NEURAL INFORMATION PROCESSING SYSTEMS CONFERENCE

TUTORIALS
December 6, 2010
Hyatt Regency
Vancouver, BC, Canada

CONFERENCE SESSIONS
December 6-9, 2010
Hyatt Regency
Vancouver, BC, Canada

SAM ROWEIS SYMPOSIUM
December 9, 2010
Hyatt Regency
Vancouver, BC, Canada

WORKSHOP
December 10-11, 2010
The Westin Resort & Spa
The Hilton Whistler Resort & Spa
Whistler, BC Canada

The technical program includes six invited talks and 292 accepted papers, selected from a total of 1,219 submissions considered by the program committee. Because the conference stresses interdisciplinary interactions, there are no parallel sessions. Papers presented at the conference will appear in “Advances in Neural Information Processing Systems 23,” edited by John Lafferty, Chris Williams, John Shawe- Taylor, Richard Zemel, and Aron Culotta.

Neural Information
Processing Systems
Foundation

TABLE OF CONTENTS

Neural Information
Processing Systems
Foundation

ORGANIZING COMMITTEE

General Chairs	**John Lafferty**, Carnegie Mellon University, **Chris Williams**, University of Edinburgh
Program Chairs	**Richard Zemel**, University of Toronto, **John Shawe-Taylor**, University College London
Tutorials Chair	**Kevin Murphy**, University of British Columbia
Workshop Chair	**Neil D Lawrence**, University of Manchester
Demonstration Chair	**Isabelle Guyon**, Clopinet
Publications Chair & Electronic Proceedings Chair	**Aron Culotta**, Southeastern Louisiana University
Workflow Master	**Laurent Charlin**, University of Toronto

PROGRAM COMMITTEE

Maria Florina Balcan, Georgia Institute of Technology
Mikhail Belkin, The Ohio State University
Shai Ben-David, University of Waterloo
Andrew Blake . Microsoft Research
David Blei. Princeton University
Rich Caruana, Microsoft Research
Olivier Chapelle, Yahoo Research
Corinna Cortes, Google, Inc
Hal Daume III, University of Utah
Tijl De Bie, University of Bristol
Nando de Freitas, UBC
Ofer Dekel, Microsoft Research
David Dunson, Duke University
Li Fei-Fei, Stanford University
Rob Fergus, New York University
Kenji Fukumizu, Institute of Statistical Mathematics
Sally Goldman, Google, Inc. & Washington University in St. Louis
Matthias Hein, Saarland University
Katherine Heller, University of Cambridge
Tony Jebara, Columbia University
Charles Kemp, Carnegie Mellon University
John Langford, Yahoo Research
Tai Sing Lee, Carnegie Mellon University
Wee Sun Lee, National University of Singapore
Mate Lengyel, University of Cambridge
Chih-Jen Lin, National Taiwan University
Bernabe Linares-Barranco, Instituto de Microelectronica de Sevilla
Phil Long, Google, Inc
Jon McAuliffe, UC Berkeley
Remi Munos, INRIA Lille - Nord Europe
Iain Murray, University of Toronto/Edinburgh
Yael Niv, Princeton University
William Noble, University of Washington
Rob Nowak, University of Wisconsin-Madison
Jan Peters, Max-Planck Institute of Biological Cybernetics
Jonathan Pillow, University of Texas at Austin
Pascal Poupart, University of Waterloo
Pradeep Ravikumar, University of Texas at Austin
Irina Rish, IBM T.J. Watson Research Center
Matthias Seeger, Saarland University
Nathan Srebro, TTI-Chicago
Ben Taskar, University of Pennsylvania
Bill Triggs, CNRS
Raquel Urtasun, TTI-Chicago
Kilian Weinberger, Washington University in St. Louis
Tong Zhang, Rutgers
Dengyong Zhou, Microsoft Research
Xiaojin (Jerry) Zhu, University of Wisconsin-Madison

NIPS would like to especially thank Microsoft Research for their donation of Conference Management Toolkit (CMT) software and server space.

NIPS FOUNDATION OFFICERS & BOARD MEMBERS

President	**Terrence Sejnowski**, The Salk Institute
Treasurer	**Marian Stewart Bartlett**, University of California, San Diego
Secretary	**Michael Mozer**, University of Colorado, Boulder
Legal Advisor	**Phil Sotel**, Pasadena, CA
Executive	**Dale Schuurmans**, University of Alberta, Canada, **Yoshua Bengio**, University of Montreal, Canada; **Daphne Koller**, Stanford University; **John C. Platt**, Microsoft Research; **Yair Weiss**, Hebrew University of Jerusalem; **Lawrence Saul**, University of Pennsylvania; **Bernhard Scholkopf**, Max Planck Institute
Advisory Board	**Gary Blasdel**, Harvard Medical School; **Jack Cowan**, University of Chicago; **Stephen Hanson**, Rutgers University; **Michael I. Jordan**, University of California, Berkeley; **Michael Kearns**, University of Pennsylvania; **Scott Kirkpatrick**, Hebrew University, Jerusalem; **Richard Lippmann**, Massachusetts Institute of Technology; **Todd K. Leen**, Oregon Graduate Institute; **Bartlett Mel**, University of Southern California; **John Moody**, International Computer Science Institute, Berkeley and Portland; **Gerald Tesauro**, IBM Watson Labs; **Dave Touretzky**, Carnegie Mellon University; **Sebastian Thrun**, Stanford University; **Thomas G. Dietterich**, Oregon State University; **Sara A. Solla**, Northwestern University Medical; School; **Sue Becker**, McMaster University, Ontario, Canada
Emeritus Members	**Terrence L. Fine**, Cornell University; **Eve Marder**, Brandeis University

CORE LOGISTICS TEAM

The running of NIPS would not be possible without the help of many volunteers, students, researchers and administrators who donate their valuable time and energy to assist the conference in various ways. However, there is a core team at the Salk Institute whose tireless efforts make the conference run smoothly and efficiently every year. This year, NIPS would particularly like to acknowlege the exceptional work of:

Lee Campbell - IT Manager
Chris Hiestand - Webmaster
Sheri Leone - Volunteer Coordinator
Mary Ellen Perry - Executive Director

AWARDS

OUTSTANDING STUDENT PAPER AWARDS
Construction of Dependent Dirichlet Processes based on Poisson Processes
Dahua Lin*, Eric Grimson and John Fisher

A Theory of Multiclass Boosting
Indraneel Mukherjee* and Robert E Schapire

STUDENT PAPER HONORABLE MENTIONS
MAP estimation in Binary MRFs via Bipartite Multi-cuts
Sashank Jakkam Reddi*, Sunita Sarawagi and Sundar Vishwanathan

The Multidimensional Wisdom of Crowds
Peter Welinder*, Steve Branson, Serge Belongie and Pietro Perona

* Winner

SPONSORS

NIPS gratefully acknowledges the generosity of those individuals and organizations who have provided financial support for the NIPS 2010 conference. The financial support enabled us to sponsor student travel and participation, the outstanding student paper awards, the demonstration track and the opening buffet.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
UNITED STATES AIR FORCE

Microsoft® Research

ELSEVIER
Artificial Intelligence

Google™

IBM Research

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

BRAIN PRODUCTS
Solutions for neurophysiological research

TWO 2σ SIGMA

TOYOTA

D E Shaw & Co

YAHOO!® LABS

intel®

Springer
Machine Learning Journal

PROGRAM HIGHLIGHTS

Monday, December 6th

8:00am–6:00pm	Registration Desk Open	Third Floor	
8:00am–6:00pm	Internet Access Room Open	Oxford	
9:30pm–5:30pm	Tutorials	Regency Ballroom D and E/F	pg 8
6:15pm–6:30pm	Opening Remarks, Awards and Reception	Regency Ballroom	
6:30pm–6:50pm	Spotlights Session 1	Regency Ballroom	pg 12
7:00pm–11:59pm	Poster Session	Second and Third Floors	pg 12

Tuesday, December 7th

7:30am–9:00am	Breakfast	Perspectives Level, 34th floor	
8:00am–9:00am	Breakfast	also in Regency A Conference Level	
8:00am–6:00pm	Registration Desk Open	Third Floor	
8:00am–6:00pm	Internet Access Room Open	Oxford	
8:30am–9:40am	Oral Session 1: **Antonio Rangel** *Invited Talk: How Does the Brain Compute and Compare Values at the Time of Decision-Making?*	Regency Ballroom	pg 47
9:40am–10:00am	Spotlights Session 2	Regency Ballroom	pg 47
10:00am–10:20pm	Oral Session 2	Regency Ballroom	pg 47
10:20am–10:50am	Break		
10:50am–11:10am	Oral Session 3	Regency Ballroom	pg 47
11:10am–11:30am	Spotlights Session 3	Regency Ballroom	pg 48
11:30am–11:50am	Oral Session 4	Regency Ballroom	pg 48
11:50am–12:10am	Spotlights Session 4	Regency Ballroom	pg 48
12:10am–2:00pm	Lunch		
2:00pm–3:10pm	Oral Session 5: **Meng Xiao-Li** *Invited Talk: Machine Learning Human Intelligence: Principled Corner Cutting (PC2)*	Regency Ballroom	pg 48
3:10pm–3:30pm	Spotlights Session 5	Regency Ballroom	pg 49
3:30pm–3:50pm	Oral Session 6	Regency Ballroom	pg 49
3:50pm–4:20pm	Break		
4:20pm–5:00pm	Oral Session 7	Regency Ballroom	pg 49
5:00pm–5:20pm	Spotlights Session 6	Regency Ballroom	pg 50
5:20pm–5:40pm	Oral Session 8	Regency Ballroom	pg 50
5:40pm–6:00pm	Spotlights Session 7	Regency Ballroom	pg 50
1:00pm–6:00pm	Poster/Demo Setup and Preview	Second & Third Floor	
7:00pm–11:59pm	Poster Session	Second & Third Floor	pg 51
7:30pm–11:59pm	Demonstrations 1	Georgia A	pg 71

PROGRAM HIGHLIGHTS

Wednesday, December 8th

7:30am–9:00am	Breakfast	Perspectives Level, 34th floor	
8:00am–9:00am	Breakfast	also in Regency A Conference Level	
8:00am–6:00pm	Registration Desk Open	Third Floor	
8:00am–6:00pm	Internet Access Room Open	Oxford	
8:30am–9:40am	Oral Session 9: **David Parkes** Invited Talk: The Interplay of Machine Learning and Mechanism Design	Regency Ballroom	pg 76
9:40am–10:00am	Spotlights Session 8	Regency Ballroom	pg 76
10:00am–10:20am	Oral Session 10	Regency Ballroom	pg 77
10:20am–10:50am	Break		
10:50am–11:10am	Oral Session 11	Regency Ballroom	
11:10am–11:30am	Spotlights Session 9	Regency Ballroom	pg 77
11:30am–11:50am	Oral Session 12	Regency Ballroom	pg 77
11:50am–12:10am	Spotlights Session 10	Regency Ballroom	pg 77
12:10pm–2:00pm	Lunch		
2:00pm–3:10pm	Oral Session 13: **Michael Jordan** *Statistical Inference of Protein Structure and Function*	Regency Ballroom	pg 77
3:10pm–3:30pm	Spotlights Session 11	Regency Ballroom	pg 78
3:30pm–3:50pm	Oral Session 14	Regency Ballroom	pg 78
3:50pm–4:20pm	Break		
4:20pm–5:00pm	Oral Session 15	Regency Ballroom	pg 78
5:00pm–5:20pm	Spotlights Session 12	Regency Ballroom	pg 79
5:20pm–5:40pm	Oral Session 16	Regency Ballroom	pg 79
5:40pm–6:00pm	Spotlights Session 13	Regency Ballroom	pg 79
7:00pm–11:59pm	Poster Session	Second and Third Floors	pg 79
7:30pm–11:59pm	Demonstrations 2	Georgia A	pg 107

Thursday, December 9th

7:30am–9:00am	Breakfast	Perspectives Level, 34th floor	
8:00am–9:00am	Breakfast	also in Regency A Conference Level	
8:00am–11:00am	Internet Cafe	Oxford	
8:00am–11:00am	Registration Desk	Third Floor	
9:00am–10:10am	Oral Session 17: **Martin Banks** Invited Talk: *Perceptual Bases for Rules of Thumb in Photography*	Regency Ballroom	pg 109
10:10am–10:40am	Break		
10:40am–11:50am	Oral Session 18: **Josh Tenenbaum** Invited Talk: *How to Grow a Mind: Statistics, Structure and Abstraction*	Regency Ballroom	pg 109
2:00pm–5:00pm	The Sam Roweis Symposium	Regency Ballroom	pg 112
2:00pm–6:00pm	Buses depart for Workshops		

MONDAY TUTORIALS

9:30AM–11:30AM - Tutorial Session 1

High-dimensional Statistics: Prediction, Association and Causal Inference
Peter Buhlmann buhlmann@stat.math.ethz.ch
Location: Regency E/F

Reinforcement Learning for Embodied Cognition
Dana Ballard dana@cs.utexas.edu
Location: Regency D

1:00PM–3:00PM - Tutorial Session 2

Optimization Algorithms in Machine Learning
Stephen Wright swright@cs.wisc.edu
Location: Regency E/F

Reinforcement Learning in Humans and Other Animals
Nathaniel Daw daw@gatsby.ucl.ac.uk
Location: Regency D

3:30PM–5:30PM - Tutorial Session 3

Latent Factor Models for Relational Arrays and Network Data
Peter Hoff hoff@stat.washington.edu
Location: Regency E/F

Vision-Based Control, Control-Based Vision, and the Information Knot That Ties Them
Stefano Soatto soatto@ucla.edu
Location: Regency D

ABSTRACTS OF TUTORIALS

Tutorial Session 1, 9:30am–11:30am

Tutorial: ***High-dimensional Statistics: Prediction, Association and Causal Inference***
Peter Buhlmann
buhlmann@stat.math.ethz.ch
ETH Zurich

This tutorial surveys methodology and theory for high-dimensional statistical inference when the number of variables or features greatly exceeds sample size. Particular emphasis will be placed on problems of model and feature selection. This includes variable selection in regression models or estimation of the edge set in graphical modeling. While the former is concerned with association, the latter can be used for causal analysis. In the highdimensional setting, major challenges include designing computational algorithms that are feasible for large-scale problems, assigning statistical error rates (e.g., p-values), and developing theoretical insights about the limits of what is possible. We will present some of the most important recent developments and discuss their implications for prediction, association analysis and some exciting new directions in causal inference.

***Peter Buhlmann** is Professor of Statistics at the ETH Zurich. His research interests are in computational statistics, high-dimensional statistical inference, machine learning and applications in the life sciences. He is a Fellow of the Institute of Mathematical Statistics, an elected Member of the International Statistical Institute and he presented a Medallion Lecture at the JSM 2009. He has served as editorial member of the Journal of the Royal Statistical Society (Series B), Journal of Machine Learning Research, Biometrical Journal and he is currently the Editor of the Annals of Statistics.*

Tutorial Session 1, 9:30am–11:30am

Tutorial: ***Reinforcement Learning for Embodied Cognition***
Dana Ballard
dana@cs.utexas.edu
University of Texas, Austin

The enormous progress in instrumentation for measuring brain states has made it possible to tackle the large issue of an overall model of brain computation. The intrinsic complexity of the brain can lead one to set aside issues related to its relationships with the body, but the field of Embodied Cognition stresses that understanding of brain function at the system level requires one to address the role of the brain-body interface. While it is obvious that the brain receives all its input through the senses and directs its outputs through the motor system, it has only recently been appreciated that the body interface performs huge amounts of computation that does not have to be repeated by the brain, and thus affords the brain great simplifications in its representations. In effect the brain's abstract states can explicitly or implicitly refer to coded representations of the world created by the body. Even if the brain can communicate with the world through abstractions, the severe speed limitations in its neural circuitry means that vast amounts of indexing must be performed during development so that appropriate behavioral responses can be rapidly accessed. One way this could happen would be if the brain used some kind of decomposition whereby behavioral primitives could be quickly accessed and combined. Such a factorization has huge synergies with embodied cognition models, which can use the natural filtering imposed by the body in directing behavior to select relevant primitives. These advantages can be explored with virtual environments replete with humanoid avatars. Such settings allow the manipulation of experimental parameters in systematic ways. Our test settings are those of everyday natural settings such as walking and driving in a small town, and sandwich making and looking for lost items in an apartment. The issues we focus on center around the programming of the individual behavioral primitives using reinforcement learning (RL). Central issues are eye fixation programming, credit assignment to individual behavioral modules, and learning the value of behaviors via inverse reinforcement learning. Eye fixations are the central information gathering method used by humans, yet the protocols for programming them are still unsettled. We show that information gain in an RL setting can potentially explain experimental data. Credit assignment. If behaviors are to be decomposed into individual modules, then dividing up received reward amongst them becomes a major issue. We show that Bayesian estimation techniques, used in the RL setting, resolve this issue efficiently. Inverse Reinforcement Learning. One way to learn new behaviors would be if a human agent could imitate them and learn their value. We show that an efficient algorithm developed by Rothkopf can estimate value of behaviors from observed data using Bayesian RL techniques.

***Dana H. Ballard** obtained his undergraduate degree in Aeronautics and Astronautics from M.I.T. in 1967. Subsequently he obtained MS and PhD degrees in information engineering from the University of Michigan and the University of California at Irvine in 1969 and 1974 respectively. He is the author of two books, Computer Vision (with Christopher Brown) and An Introduction to Natural Computation. His main research interest is in computational theories of the brain with emphasis on human vision. His research places emphasis on Embodied Cognition. Starting in 1985, he and Chris Brown designed and built the first high-speed binocular camera control system capable of simulating human eye movements in real time. Currently he pursues this research at the University of Texas at Austin by using model humans in virtual reality environments. His focus is on the use of machine learning as a model for human behavior with an emphasis on reinforcement learning.*

Tutorial Session 2, 1:00pm–3:00pm

Tutorial: ***Optimization Algorithms in Machine Learning***
Stephen Wright
swright@cs.wisc.edu
University of Wisconsin-Madison

Optimization provides a valuable framework for thinking about, formulating, and solving many problems in machine learning. Since specialized techniques for the quadratic programming problem arising in support vector classification were developed in the 1990s, there has been more and more cross-fertilization between optimization and machine learning, with the large size and computational demands of machine learning applications driving much recent algorithmic research in optimization. This tutorial reviews the major computational paradigms in machine learning that are amenable to optimization algorithms, then discusses the algorithmic tools that are being brought to bear on such applications. We focus particularly on such algorithmic tools of recent interest as stochastic and incremental gradient methods, online optimization, augmented Lagrangian methods, and the various tools that have been applied recently in sparse and regularized optimization.

Steve Wright *is a Professor of Computer Sciences at the University of Wisconsin-Madison. His research interests lie in computational optimization and its applications to science and engineering. Prior to joining UW-Madison in 2001, Wright was a Senior Computer Scientist (1997-2001) and Computer Scientist (1990-1997) at Argonne National Laboratory, and Professor of Computer Science at the University of Chicago (2000-2001). He is the past Chair of the Mathematical Optimization Society (formerly the Mathematical Programming Society), the leading professional society in optimization, and a member of the Board of the Society for Industrial and Applied Mathematics (SIAM). Wright is the author or co-author of four widely used books in numerical optimization, including ”Primal Dual Interior-Point Methods” (SIAM, 1997) and ”Numerical Optimization” (with J. Nocedal, Second Edition, Springer, 2006). He has also authored over 85 refereed journal papers on optimization theory, algorithms, software, and applications. He is coauthor of widely used interior-point software for linear and quadratic optimization. His recent research includes algorithms, applications, and theory for sparse optimization (including applications in compressed sensing and machine learning).*

Tutorial Session 2, 1:00pm–3:00pm

Tutorial: ***Reinforcement Learning in Humans and Other Animals***
Nathaniel D Daw
daw@gatsby.ucl.ac.uk
New York University

Algorithms from computer science can serve as detailed process-level hypotheses for how the brain might approach difficult information processing problems. This tutorial reviews how ideas from the computational study of reinforcement learning have been used in biology to conceptualize the brain’s mechanisms for trial-and-error decision making, drawing on evidence from neuroscience, psychology, and behavioral economics. We begin with the much-debated relationship between temporal-difference learning and the neuromodulator dopamine, and then consider how more sophisticated methods and concepts from RL – including partial observability, hierarchical RL, function approximation, and various model-based approaches – can provide frameworks for understanding additional issues in the biology of adaptive behavior. In addition to helping to organize and conceptualize data from many different levels, computational models can be employed more quantitatively in the analysis of experimental data. The second aim of this tutorial is to review and demonstrate, again using the example of reinforcement learning, recent methodological advances in analyzing experimental data using computational models. An RL algorithm can be viewed as generative model for raw, trial-by-trial experimental data such as a subject’s choices or a dopaminergic neuron’s spiking responses; the problems of estimating model parameters or comparing candidate models then reduce to familiar problems in Bayesian inference. Viewed this way, the analysis of neuroscientific data is ripe for the application of many of the same sorts of inferential and machine learning techniques well studied by the NIPS community in other problem domains.

***Nathaniel Daw** is Assistant Professor of Neural Science and Psychology and Affiliated Assistant Professor of Computer Science at New York University. Prior to this he completed his PhD in Computer Science at Carnegie Mellon University and pursued postdoctoral research at the Gatsby Computational Neuroscience Unit at UCL. His research concerns reinforcement learning and decision making from a computational approach, and particularly the application of computational models to the analysis of behavioral and neural data. He is the recipient of a McKnight Scholar Award, a NARSAD Young Investigator Award, and a Royal Society USA Research Fellowship.*

Tutorial Session 3, 3:30pm–5:30pm

Tutorial: ***Latent Factor Models for Relational Arrays and Network Data***

Peter Hoff
hoff@stat.washington.edu
University of Washington

Network and relational data structures have increasingly played a role in the understanding of complex biological, social and other relational systems. Statistical models of such systems can give descriptions of global relational features, characterize local network structure, and provide predictions for missing or future relational data. Latent variable models are a popular tool for describing network and relational patterns. Many of these models are based on well-known matrix decomposition methods, and thus have a rich mathematical framework upon which to build. Additionally, the parameters in these models are easy to interpret: Roughly speaking, a latent variable model posits that the relationship between two nodes is a function of observed and unobserved (latent) characteristics, potentially in addition to contextual factors. In this tutorial I give an introduction to latent variable models for relational and network data. I first provide a mathematical justification for a general latent factor model based on exchangeability considerations. I then describe and illustrate several latent variable models in the context of the statistical analysis of several network datasets. I also compare several such models in terms of what network features they can, and cannot, represent. A particularly flexible class of models are the ”latent factor” models, based on singular value and eigen-decompositions of a relational matrix. These models generalize in a natural way to accommodate more complicated relational data, such as datasets that are described by multiway arrays, such as a network measured over time or the measurement of several relational variables on a common nodeset. I will close the tutorial by showing how tools from multiway data analysis (such as the higher order SVD and PARAFAC decomposition) can be used to build statistical models of multiway networks and relational data.

***Peter Hoff** is an Associate Professor of Statistics and Biostatistics at the University of Washington. He has developed a variety of Bayesian methods for multivariate data, including covariance and copula estimation, cluster analysis, mixture modeling and social network analysis. He is on the editorial board of the Annals of Applied Statistics, JRSSB and SIAM Classics.*

Tutorial Session 3, 3:30pm–5:30pm

Tutorial: ***Vision-Based Control, Control-Based Vision, and the Information Knot That Ties Them***

Stefano Soatto
soatto@ucla.edu
UCLA

The purpose of this tutorial is to explore the interplay between sensing and control, to highlight the ”information knot” that ties them, and to design inference and learning algorithms to compute ”representations” from data that are optimal, by design, for decision and control tasks. We will focus on visual sensing, but the analysis developed extends to other modalities. We will first review various notions of information proposed in different fields from economic theory to perception psychology, and adapt them to decision and control tasks, as opposed to transmission and storage of data. We will see that for complex sensing phenomena, such as vision, nuisance factors play an important role, especially those that are not ”invertible” such as occlusions of line-of-sight and quantization-scale. Handling of the nuisances brings forward a notion of ”representation,” whose complexity measures the amount of ”actionable information” contained in the data. We will discuss how to build representations that are optimal by design, in the sense of retaining all and only the statistics that matter to the task. For ”invertible” nuisances, such representations can be made lossless (not in the classical sense of distortion, but in the sense of optimal performance in a decision or control task). In some cases, these representations are supported on a thin-set, which can help elucidate the ”signal-to-symbol barrier” problem, and relate to a topology-based notion of ”sparsity”. However, non-invertible nuisances spoil the picture, requiring the introduction of a notion of ”stability” of the representation with respect to non-invertible nuisances. This is not the classical notion of (bounded-input-bounded-output) stability from control theory, but instead relates to ”structural stability” from catastrophe theory. The design of maximally stable statistics brings forward a notion of ”proper sampling” of the data. However, this is not the traditional notion of proper sampling from Nyquist, but one related to persistent topology. Once an optimal representation is constructed, a bound on the risk or control functional can be derived, analog to distortion in communications. The ”currency” that trades off this error (the equivalent of the bit-rate in communication) is not the amount of data, but instead the ”control authority” over the sensing process. Thus, sensing and control are intimately tied: Actionable information drives the control process, and control of the sensing process is what allows computing a representation. We will present case studies in which formulating visual decision problems (e.g. detection, localization, recognition, categorization) in the context of vision-based control leads to improved performance and reduced computational burden. They include established low-level vision tools (e.g. tracking, local invariant descriptors), robotic exploration, and action and activity recognition. We will describe some of these in detail and distribute source code at the workshop, together with course notes.

***Stefano Soatto** received his Ph.D. in Control and Dynamical Systems from the California Institute of Technology in 1996; he joined UCLA in 2000 after being Assistant and then Associate Professor of Electrical Engineering and Biomedical Engineering at Washington University, and Research Associate in Applied Sciences at Harvard University. Between 1995 and 1998 he was also Ricercatore in the Department of Mathematics and Computer Science at the University of Udine - Italy. He received his D.Ing. degree (highest honors) from the University of Padova- Italy in 1992. His general research interests are in Computer Vision and Nonlinear Estimation and Control Theory. In particular, he is interested in ways for computers to use sensory information to interact with humans and the environment. Dr. Soatto is the recipient of the David Marr Prize for work on Euclidean reconstruction and reprojection up to subgroups. He also received the Siemens Prize with the Outstanding Paper Award from the IEEE Computer Society for his work on optimal structure from motion. He received the National Science Foundation Career Award and the Okawa Foundation Grant. He is a Member of the Editorial Board of the International Journal of Computer Vision (IJCV) and Foundations and Trends in Computer Graphics and Vision. He is the founder and director of the UCLA Vision Lab; more information is available at http://vision.ucla.edu.*

MONDAY CONFERENCE

MONDAY, DECEMBER 6TH

6:30PM–7:45PM - OPENING REMARKS, AWARDS & RECEPTION

Dinner buffet open until 9pm; dessert buffet until 10pm.

6:30–6:50PM - SPOTLIGHTS SESSION 1
Session Chair: Xiaojin (Jerry) Zhu

- ***Learning from Candidate Labeling Sets***
  Jie Luo, Idiap/EPF Lausanne, and Francesco Orabona, University of Milano.

- ***Why Are Some Word Orders More Common Than Others? A uniform information density account***
  Luke Maurits, Dan Navarro and Amy Perfors, Univ. of Adelaide.

- ***Layered Image Motion with Explicit Occlusions, temporal consistency, and depth ordering***
  Deqing Sun, Erik Sudderth and Michael Black, Brown Univ.

- ***b-Bit Minwise Hashing for Estimating Three-Way Similarities***
  Ping J Li, Cornell, Arnd C Konig, Microsoft Research; Wenhao Gui

- ***Getting Lost in Space: Large Sample Analysis of the Resistance Distance***
  Ulrike von Luxburg and Agnes Radl, Max Planck Institute for Biological Cybernetics, and Matthias Hein, Saarland University.

7:00–11:59PM - POSTER SESSION AND RECEPTION

M1 **Sufficient Conditions for Generating Group Level Sparsity in a Robust Minimax Framework**, Hongbo Zhou, Southern Illinois University, and Qiang Cheng, Southern Illinois University

M2 **Multitask Learning without Label Correspondences**, Novi Quadrianto, Tiberio Caetano and James Petterson, NICTA and ANU, S.V.N; Alexander J Smola, Yahoo! Research; Vishwanathan, Purdue University

M3 **Generative Local Metric Learning for Nearest Neighbor Classification**, Yung-Kyun Noh and Daniel Lee, University of Pennsylvania, Byoung-Tak Zhang, Seoul National University

M4 **Gated Softmax Classification**, Roland Memisevic, Christopher Zach, ETH Zurich, Geoffrey Hinton, University of Toronto, and Marc Pollefeys, ETH Zurich

M5 **Convex Multiple-Instance Learning by Estimating Likelihood Ratio**, Fuxin Li and Cristian Sminchisescu, University of Bonn

M6 **Boosting Classifier Cascades**, Mohammad Saberian and Nuno Vasconcelos, UC San Diego

M7 **Efficient algorithms for learning kernels from multiple similarity matrices with general convex loss functions**, Achintya Kundu, vikram M Tankasali and Chiranjib Bhattacharyya, Indian Institute of Science Bangalore; Aharon Ben-Tal, Technion- Israel Institute of Technology

M8 **Learning Kernels with Radiuses of Minimum Enclosing Balls**, *Kun Gai, Guangyun Chen and Changshui Zhang, Tsinghua University*

M9 **Multi-label Multiple Kernel Learning by Stochastic Approximation: Application to Visual Object Recognition**, Serhat S Bucak and Rong Jin, Michigan State University, and Anil K Jain

M10 **Spectral Regularization for Support Estimation**, Ernesto De Vito, DIMA, Lorenzo Rosasco, MIT and IIT, and Alessandro Toigo, Unimi

M11 **Direct Loss Minimization for Structured Prediction**, David A McAllester, Tamir Hazan and Joseph Keshet, TTI Chicago

M12 **Sidestepping Intractable Inference with Structured Ensemble Cascades**, David Weiss, Benjamin J Sapp and Ben Taskar, University of Pennsylvania

M13 **Layer-wise analysis of deep networks with Gaussian kernels**, Gregoire Montavon, Mikio L Braun and Klaus-Robert Muller, TU Berlin

M14 **Scrambled Objects for Least-Squares Regression**, Odalric Maillard, and Remi Munos, INRIA Lille

M15 **Block Variable Selection in Multivariate Regression and High-dimensional Causal Inference**, Aurelie C Lozano and Vikas Sindhwani, IBM Research

M16 **Why are some word orders more common than others? A uniform information density account**, Luke Maurits, Dan Navarro and Amy Perfors, Univ. of Adelaide

M17 **Parametric Bandits: The Generalized Linear Case**, Sarah Filippi, Olivier Cappe and Aurelien Garivier, Telecom ParisTech; Csaba Szepesvari, Univ. of Alberta

M18 **Spike timing-dependent plasticity as dynamic filter**, Joscha T Schmiedt and Christian Albers, University of Bremen; Klaus Pawelzik,

M19 **Short-term memory in neuronal networks through dynamical compressed sensing**, Surya Ganguli, UCSF; Haim Sompolinsky, Hebrew Univ. & Harvard

M20 **Hallucinations in Charles Bonnet Syndrome Induced by Homeostasis: a Deep Boltzmann Machine Model,** David P Reichert, Peggy Series and Amos J Storkey, University of Edinburgh

M21 **Mixture of time-warped trajectory models for movement decoding**, Elaine A Corbett, Eric J Perreault and Konrad Koerding, Northwestern University

M22 **Decoding Ipsilateral Finger Movements from ECoG Signals in Humans**, Yuzong Liu, Mohit Sharma, Charles M Gaona, Zachary V Freudenburg, Kilian Q Weinberger, Jonathan D Breshears, Jarod Roland and Eric C Leuthardt, Washington University in St. Louis

M23 **b-Bit Minwise Hashing for Estimating Three-Way Similarities**, Ping J Li, Cornell, Arnd C Konig, Microsoft Research, and Wenhao Gui

M24 **A Computational Decision Theory for Interactive Assistants**, Alan Fern and Prasad Tadepalli, Oregon State University

M25 **Multi-Stage Dantzig Selector**, Ji Liu, Peter Wonka and Jieping Ye, Arizona State University

M26 **Estimation of Renyi Entropy and Mutual Information Based on Generalized Nearest-Neighbor Graphs**, David Pal and Csaba Szepesvari, University of Alberta, Barnabas Poczos, Carnegie Mellon University

M27 **Probabilistic Belief Revision with Structural Constraints**, Peter Jones and Sanjoy K Mitter, MIT; Venkatesh Saligrama, Boston University

M28 **Getting lost in space: Large sample analysis of the resistance distance**, Ulrike von Luxburg, Agnes Radl, Max Planck Institute for Biological Cybernetics, and Matthias Hein, Saarland University

M29 **Bayesian Action-Graph Games**, Albert Xin Jiang and Kevin Leyton-Brown, University of British Columbia

M30 **Random Conic Pursuit for Semidefinite Programming**, Ariel Kleiner and Michael I Jordan, University of California Berkeley; Ali Rahimi, Intel,

M31 **Generalized roof duality and bisubmodular functions**, Vladimir Kolmogorov, UC London

M32 **Sparse Inverse Covariance Selection via Alternating Linearization Methods**, Katya Scheinberg, Shiqian Ma, Columbia University, and Donald Goldfarb

M33 **Optimal Web-Scale Tiering as a Flow Problem**, Gilbert Leung, eBay Inc., Novi Quadrianto, SML-NICTA and RSISE-ANU; Alexander J Smola, Yahoo! Research; Kostas Tsioutsiouliklis, Yahoo! Labs

M34 **Parallelized Stochastic Gradient Descent**, Martin Zinkevich, Markus Weimer, Alex Smola, and Lihong Li, Yahoo! Research

M35 **Non-Stochastic Bandit Slate Problems**, Satyen Kale, Yahoo!; Lev Reyzin, Georgia Institute of Technology; Robert E Schapire, Princeton University

M36 **Repeated Games against Budgeted Adversaries**, Jacob Abernethy, UC Berkeley; Manfred Warmuth, UC Santa Cruz

M37 **Two-Layer Generalization Analysis for Ranking Using Rademacher Average**, Wei Chen and Zhi-Ming Ma, Chinese Academy of Sciences; Tie-Yan Liu, Microsoft Research

M38 **Empirical Bernstein Inequalities for U-Statistics, Thomas Peel**, Sandrine Anthoine and Liva Ralaivola, Universite Aix-Marseille I / CNRS

M39 **On the Theory of Learnining with Privileged Information**, Dmitry Pechyony, NEC Labs, and Vladimir Vapnik

M40 **Interval Estimation for Reinforcement-Learning Algorithms in Continuous-State Domains**, Martha White and Adam M White, University of Alberta

M41 **Monte-Carlo Planning in Large POMDPs**, David Silver, MIT, and Joel Veness, Univ. of New South Wales

M42 **Basis Construction from Power Series Expansions of Value Functions**, Sridhar Mahadevan and Bo Liu, UMass Amherst

M43 **Reward Design via Online Gradient Ascent**, Jonathan D Sorg, Satinder Singh and Richard L Lewis, University of Michigan

M44 **Bootstrapping Apprenticeship Learning**, Abdeslam Boularias, MPI for Biological Cybernetics, and Brahim Chaib-draa

M45 **PAC-Bayesian Model Selection for Reinforcement Learning**, Mahdi Milani Fard and Joelle Pineau, McGill University

M46 **An Approximate Inference Approach to Temporal Optimization in Optimal Control**, Konrad C Rawlik and Sethu Vijayakumar, University of Edinburgh; Marc Toussaint, TU Berlin

M47 **Nonparametric Bayesian Policy Priors for Reinforcement Learning**, Finale Doshi-Velez, David Wingate, Nicholas Roy, and Joshua Tenenbaum, Massachusetts Institute of Technology

M48 **Predictive State Temporal Difference Learning**, Byron Boots and Geoff Gordon, Carnegie Mellon Univ.

M49 **Double Q-learning**, Hado P van Hasselt, Center for Mathematics and Computer Science

M50 **Error Propagation for Approximate Policy and Value Iteration**, Amir-massoud Farahmand and Csaba Szepesvari, University of Alberta, Remi Munos, INRIA Lille - Nord Europe

M51 **Multiparty Differential Privacy via Aggregation of Locally Trained Classifiers**, Manas A Pathak and Bhiksha Raj, CMU; Shantanu Rane, Mitsubishi Electric Research Labs,

M52 **A New Probabilistic Model for Rank Aggregation**, Tao Qin, Xiubo Geng, Chinese Academy of Sciences, and Tie-Yan Liu, Microsoft Research

M53 **The Maximal Causes of Natural Scenes are Edge Filters**, Jose G Puertas, FIAS, Jorg Bornschein and Jorg Lucke, University of Frankfurt

M54 **Inference with Multivariate Heavy-Tails in Linear Models**, Danny Bickson and Carlos Guestrin, CMU

M55 **A Bayesian Approach to Concept Drift**, Stephen H Bach and Mark Maloof, Georgetown University

M56 **Auto-Regressive HMM Inference with Incomplete Data for Short- Horizon Wind Forecasting**, Chris Barber, Joseph Bockhorst and Paul Roebber, UW Milwaukee

M57 **Switching state space model for simultaneously estimating state transitions and nonstationary firing rates**, Ken Takiyama and Masato Okada, The Univ. of Tokyo

M58 **Computing Marginal Distributions over Continuous Markov Networks for Statistical Relational Learning**, Matthias Broecheler, Univ. of Maryland CP, and Lise Getoor

M59 **Heavy-Tailed Process Priors for Selective Shrinkage**, Fabian L Wauthier and Michael I Jordan, UC Berkeley

M60 **MAP Estimation for Graphical Models by Likelihood Maximization**, Akshat Kumar & Shlomo Zilberstein, UMass

M61 **Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models**, Han Liu, Kathryn Roeder and Larry Wasserman, CMU

M62 **Fast Large-scale Mixture Modeling with Component-specific Data Partitions**, Bo Thiesson, Microsoft Research; Chong Wang, Princeton University

M63 **Deterministic Single-Pass Algorithm for LDA**, Issei Sato, Tokyo University; Kenichi Kurihara, Google; Hiroshi Nakagawa

M64 **Efficient Relational Learning with Hidden Variable Detection**, Ni Lao, Jun Zhu, Liu Xinwang, Yandong Liu and William W Cohen, Carnegie Mellon University

M65 **Fast detection of multiple change-points shared by many signals using group LARS**, Jean-Philippe Vert, Mines ParisTech, and Kevin Bleakley, Institut Curie

M66 **A VLSI Implementation of the Adaptive Exponential Integrate-and-Fire Neuron Model,** Sebastian C Millner, Andreas Grubl, Karlheinz Meier, Johannes Schemmel and Marc-Olivier Schwartz, Kirchhoff-Institut fur Physik

M67 **Structural epitome: a way to summarize one's visual experience**, Nebojsa Jojic, Microsoft Research, Alessandro Perina, Universtiy of Verona, and Vittorio Murino, University of Verona / Italian Institute of Tech.

M68 **A unified model of short-range and long-range motion perception**, Shuang Wu, Xuming He, Hongjing Lu and Alan L Yuille, UCLA

M69 **Object Bank: A High-Level Image Representation for Scene Classification & Semantic Feature Sparsification**, Li-Jia Li, Li Fei-Fei and Hao Su, Stanford University; Eric Xing, Carnegie Mellon Univ.

M70 **Size Matters: Metric Visual Search Constraints from Monocular Metadata**, Mario J Fritz and Trevor Darrell, UC Berkeley; Kate Saenko, MIT,

M71 **Occlusion Detection and Motion Estimation with Convex Optimization**, Alper Ayvaci, Michalis Raptis and Stefano Soatto, UCLA

M72 **Functional form of motion priors in human motion perception**, Hongjing Lu, UCLA, Tungyou Lin, Alan L Lee, Luminita Vese, and Alan L Yuille, UCLA

M73 **Layered image motion with explicit occlusions, temporal consistency, and depth ordering**, Deqing Sun, Erik Sudderth and Michael Black, Brown University

M74 **Sparse Instrumental Variables (SPIV) for Genome-Wide Studies**, Felix V Agakov and Amos J Storkey, Univ. of Edinburgh; Paul McKeigue and Jon Krohn, Oxford

M75 **Predicting Execution Time of Computer Programs Using Sparse Polynomial Regression**, Ling Huang, Byung-Gon Chun, Petros Maniatis and Mayur Naik, Intel; Jinzhu Jia, Bin Yu, UC Berkeley,

M76 **Cross Species Expression Analysis using a Dirichlet Process Mixture Model with Latent Matchings**, Hai-Son P Le and Ziv Bar-Joseph, Carnegie Mellon Univ.

M77 **Adaptive Multi-Task Lasso: with Application to eQTL Detection**, Seunghak Lee, Jun Zhu and Eric Xing, CMU

M78 **Reverse Multi-Label Learning**, James Petterson, NICTA, and Tiberio Caetano, NICTA and ANU

M79 **Empirical Risk Minimization with Approximations of Probabilistic Grammars**, Shay Cohen and Noah Smith, Carnegie Mellon University

M80 **An Alternative to Low-level-Sychrony-Based Methods for Speech Detection**, Paul L Ruvolo, UC San Diego; Javier R Movellan, Machine Perception Laboratory

M81 **Phone Recognition with the Mean-Covariance Restricted Boltzmann Machine**, George E Dahl, Marc'Aurelio Ranzato, Abdel-rahman Mohamed and Geoffrey Hinton, University of Toronto

M82 **Beyond Actions: Discriminative Models for Contextual Group Activities**, Tian Lan, Yang Wang, Weilong Yang, and Greg Mori, Simon Fraser University

M83 **Group Sparse Coding with a Laplacian Scale Mixture Prior**, Pierre J Garrigues, IQ Engines Inc., and Bruno A Olshausen, UC Berkeley

M84 **Regularized estimation of image statistics by Score Matching**, Diederik P Kingma, Universiteit Utrecht, and Yann Le Cun, New York University

M85 **Learning from Candidate Labeling Sets**, Jie Luo, Idiap/EPF Lausanne, and Francesco Orabona, Univ. of Milano

M86 **Co-regularization Based Semi-supervised Domain Adaptation**, Hal Daume III, Abhishek Kumar, University of Utah, and Avishek Saha

M87 **Batch Bayesian Optimization via Simulation Matching**, Javad Azimi, Alan Fern and Xiaoli Fern, Oregon State U.

M88 **Active Estimation of F-Measures**, Christoph Sawade, Niels Landwehr and Tobias Scheffer, Univ. of Potsdam

M89 **Agnostic Active Learning Without Constraints**, Alina Beygelzimer, IBM Research; Zhang Tong and Daniel Hsu, Rutgers Univ. and Univ. of Pennsylvania; John Langford, Yahoo Research

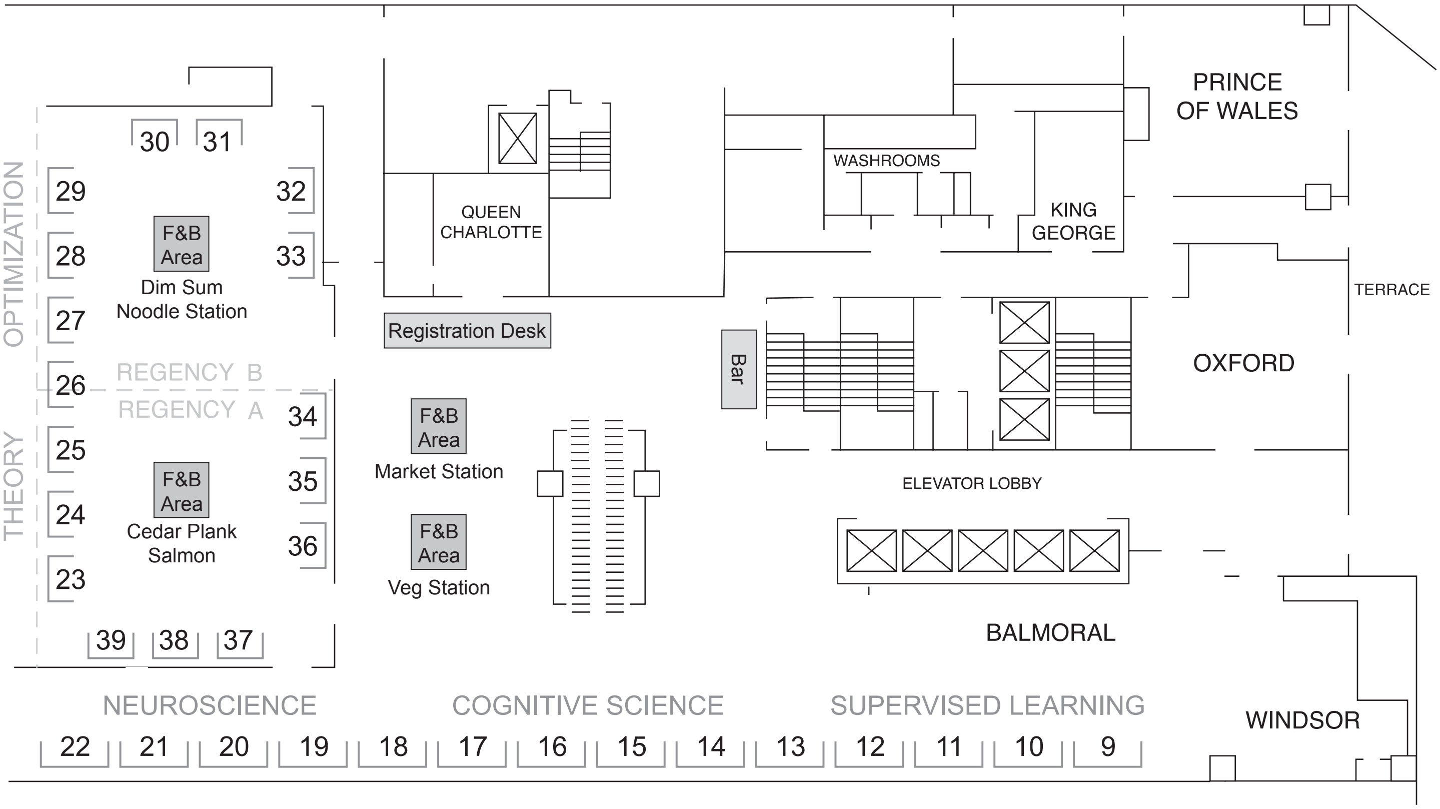
M90 **CUR from a Sparse Optimization Viewpoint**, Jacob Bien, Ya Xu and Michael Mahoney, Stanford

M91 **Towards Property-Based Classification of Clustering Paradigms**, Margareta Ackerman, Shai Ben-David and David R Loker, University of Waterloo

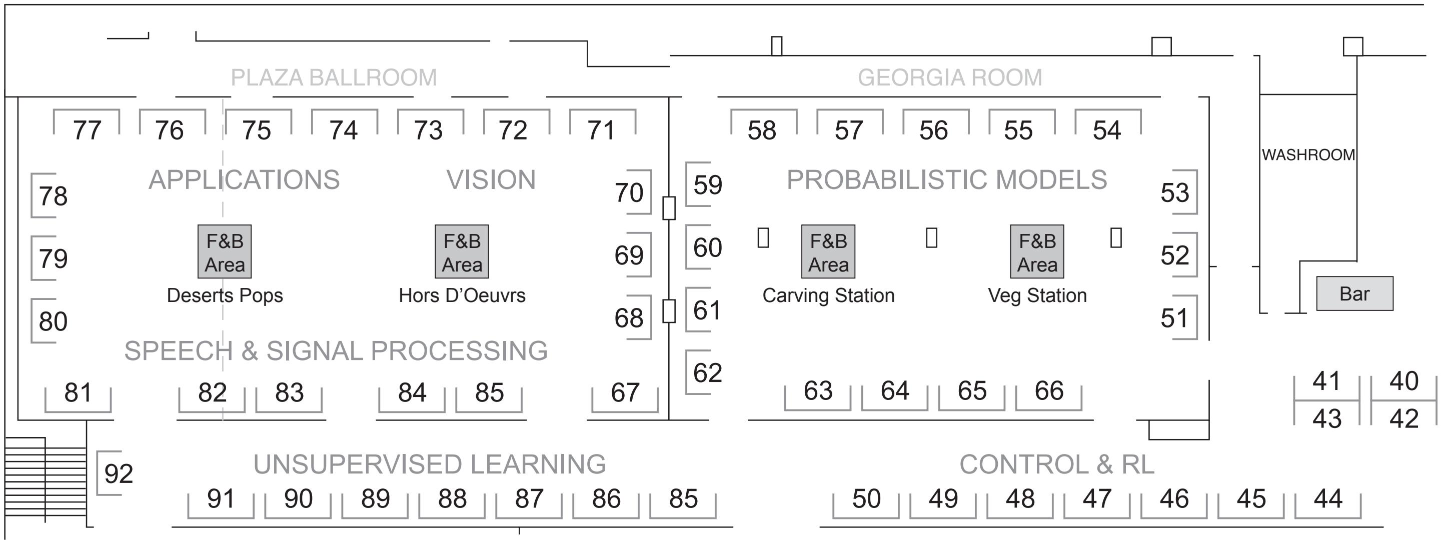
M92 **Energy Disaggregation via Discriminative Sparse Coding**, J. Zico Kolter, MIT, Siddharth Batra and Andrew Ng, Stanford

NIPS POSTER BOARDS - MONDAY, DECEMBER 6TH

CONVENTION LEVEL (3RD FLOOR)



PLAZA LEVEL (2ND FLOOR)



MONDAY, DECEMBER 6TH

SPOTLIGHTS SESSION (6:30–6:55PM)

6:30–6:50PM - SPOTLIGHTS SESSION 1
Session Chair: Xiaojin (Jerry) Zhu

- ***Learning from Candidate Labeling Sets***
  Jie Luo, Idiap/EPF Lausanne, and Francesco Orabona, University of Milano.
  Subject Area: Unsupervised & Semi-supervised Learning
  See abstract, page 37

- ***Why Are Some Word Orders More Common Than Others? A uniform information density account***
  Luke Maurits, Dan Navarro and Amy Perfors, Univ. of Adelaide.
  Subject Area: Cognitive Science
  See abstract, page 20

- ***Layered Image Motion with Explicit Occlusions, temporal consistency, and depth ordering***
  Deqing Sun, Erik Sudderth and Michael Black, Brown Univ.
  Subject Area: Vision
  See abstract, page 34

- ***b-Bit Minwise Hashing for Estimating Three-Way Similarities***
  Ping J Li, Cornell, Arnd C Konig, Microsoft Research; Wenhao Gui
  Subject Area: Theory
  See abstract, page 22

- ***Getting Lost in Space: Large Sample Analysis of the Resistance Distance***
  Ulrike von Luxburg and Agnes Radl, Max Planck Institute for Biological Cybernetics, and Matthias Hein, Saarland University.
  Subject Area: Theory
  See abstract, page 23

POSTER SESSION AND RECEPTION (7:00–11:59PM)

M1 Sufficient Conditions for Generating Group Level Sparsity in a Robust Minimax Framework

Hongbo Zhou hongboz@siu.edu
Qiang Cheng qcheng@cs.siu.edu
Southern Illinois Univeristy

Regularization technique has become a principle tool for statistics and machine learning research and practice. However, in most situations, these regularization terms are not well interpreted, especially on how they are related to the loss function and data. In this paper, we propose a robust minimax framework to interpret the relationship between data and regularization terms for a large class of loss functions. We show that various regularization terms are essentially corresponding to different distortions to the original data matrix. This minimax framework includes ridge regression, lasso, elastic net, fused lasso, group lasso, local coordinate coding, multiple kernel learning, etc., as special cases. Within this minimax framework, we further gave mathematically exact definition for a novel representation called sparse grouping representation (SGR), and proved sufficient conditions for generating such group level sparsity. Under these sufficient conditions, a large set of consistent regularization terms can be designed. This SGR is essentially different from group lasso in the way of using class or group information, and it outperforms group lasso when there appears group label noise. We also gave out some generalization bounds in a classification setting.
Subject Area: Supervised Learning

M2 Multitask Learning without Label Correspondences

Novi Quadrianto novi.quad@gmail.com
Tiberio Caetano Tiberio.Caetano@nicta.com.au
James Petterson james.petterson@nicta.com.au
SML-NICTA and RSISE-ANU
Alexander J Smola alex@smola.org
Yahoo! Research
Vishwanathan vishy@stat.purdue.edu
Purdue University

We propose an algorithm to perform multitask learning where each task has potentially distinct label sets and label correspondences are not readily available. This is in contrast with existing methods which either assume that the label sets shared by different tasks are the same or that there exists a label mapping oracle. Our method directly maximizes the mutual information among the labels, and we show that the resulting objective function can be efficiently optimized using existing algorithms. Our proposed approach has a direct application for data integration with different label spaces for the purpose of classification, such as integrating Yahoo! and DMOZ web directories.
Subject Area: Supervised Learning

M3 Generative Local Metric Learning for Nearest Neighbor Classification

Yung-Kyun Noh yungkyun.noh@gmail.com
Daniel Lee ddlee@seas.upenn.edu
University of Pennsylvania
Byoung-Tak Zhang btzhang@bi.snu.ac.kr
Seoul National University

We consider the problem of learning a local metric to enhance the performance of nearest neighbor classification. Conventional metric learning methods attempt to separate data distributions in a purely discriminative manner; here we show how to take advantage of information from parametric generative models. We focus on the bias in the information-theoretic error arising from finite sampling effects, and find an appropriate local metric that maximally reduces the bias based upon knowledge from generative models. As a byproduct, the asymptotic theoretical analysis in this work relates metric learning with dimensionality reduction, which was not understood from previous discriminative approaches. Empirical experiments show that this learned local metric enhances the discriminative nearest neighbor performance on various datasets using simple class conditional generative models.
Subject Area: Supervised Learning

M4 Gated Softmax Classification

Roland Memisevic roland@cs.toronto.edu
Geoffrey Hinton hinton@cs.toronto.edu
University of Toronto
Christopher Zach chzach@inf.ethz.ch
Marc Pollefeys marc.pollefeys@inf.ethz.ch
ETH Zurich

We describe a ”log-bilinear” model that computes class probabilities by combining an input vector multiplicatively with a vector of binary latent variables. Even though the latent variables can take on exponentially many possible combinations of values, we can efficiently compute the exact probability of each class by marginalizing over the latent variables. This makes it possible to get the exact gradient of the log likelihood. The bilinear score-functions are defined using a three-dimensional weight tensor, and we show that factorizing this tensor allows the model to encode invariances inherent in a task by learning a dictionary of invariant basis functions. Experiments on a set of benchmark problems show that this fully probabilistic model can achieve classification performance that is competitive with (kernel) SVMs, backpropagation, and deep belief nets.
Subject Area: Supervised Learning

M5 Convex Multiple-Instance Learning by Estimating Likelihood Ratio

Fuxin Li fuxin.li@ins.uni-bonn.de
Cristian Sminchisescu cristian.sminchisescu@ins.uni-bonn.de
University of Bonn

Multiple-Instance learning has been long known as a hard non-convex problem. In this work we propose an approach that recasts it as a convex likelihood ratio estimation problem. Firstly the constraint in multiple-instance learning is reformulated into a convex constraint on the likelihood ratio. Then we show that a joint estimation of a likelihood ratio function and the likelihood on training instances can be learned convexly. Theoretically we prove a quantitative relationship between the risk estimated under the 0-1 classification loss and under a loss function for likelihood ratio estimation. It is shown that our likelihood ratio estimation is generally a good surrogate for the 0-1 loss and separates positive and negative instances well. However with the joint estimation it tends to underestimate the likelihood of an example to be positive. We propose to use these likelihood ratio estimates as features and learn a linear combination on them to classify the bags. Experiments on synthetic and real datasets show the superiority of the approach.
Subject Area: Supervised Learning

M6 Boosting Classifier Cascades

Mohammad Saberian saberian@ucsd.edu
Nuno Vasconcelos nuno@ucsd.edu
UC San Diego

The problem of optimal and automatic design of a detector cascade is considered. A novel mathematical model is introduced for a cascaded detector. This model is analytically tractable, leads to recursive computation, and accounts for both classification and complexity. A boosting algorithm, FCBoost, is proposed for fully automated cascade design. It exploits the new cascade model, minimizes a Lagrangian cost that accounts for both classification risk and complexity. It searches the space of cascade configurations to automatically determine the optimal number of stages and their predictors, and is compatible with bootstrapping of negative examples and cost sensitive learning. Experiments show that the resulting cascades have state-of-the-art performance in various computer vision problems.
Subject Area: Supervised Learning

M7 Efficient algorithms for learning kernels from multiple similarity matrices with general convex loss functions

Achintya Kundu achintya.ece@gmail.com
Indian Institute of Science
vikram M Tankasali vikram@csa.iisc.ernet.in
Chiranjib Bhattacharyya chiru@csa.iisc.ernet.in
Indian Institute of Science Bangalore
Aharon Ben-Tal abental@ie.technion.ac.il
Technion- Israel Institute of Technology

In this paper we consider the problem of learning an n x n Kernel matrix from m similarity matrices under general convex loss. Past research have extensively studied the m =1 case and have derived several algorithms which require sophisticated techniques like ACCP, SOCP, etc. The existing algorithms do not apply if one uses arbitrary losses and often can not handle m ¿ 1 case. We present several provably convergent iterative algorithms, where each iteration requires either an SVM or a Multiple Kernel Learning (MKL) solver for m ¿ 1 case. One of the major contributions of the paper is to extend the well known Mirror Descent(MD) framework to handle Cartesian product of psd matrices. This novel extension leads to an algorithm, called EMKL, which solves the problem in O(m2 log n) iterations; in each iteration one solves an MKL involving m kernels and m eigendecomposition of n x n matrices. By suitably defining a restriction on the objective function, a faster version of EMKL is proposed, called REKL, which avoids the eigendecomposition. An alternative to both EMKL and REKL is also suggested which requires only an SVM solver. Experimental results on real world protein data set involving several similarity matrices illustrate the efficacy of the proposed algorithms.

Subject Area: Supervised Learning

M8 Learning Kernels with Radiuses of Minimum Enclosing Balls

Kun Gai gaikun.gk@gmail.com
Guangyun Chen cgy08@mails.tsinghua.edu.cn
Changshui Zhang zcs@mail.tsinghua.edu.cn
Tsinghua University

In this paper, we point out that there exist scaling and initialization problems in most existing multiple kernel learning (MKL) approaches, which employ the large margin principle to jointly learn both a kernel and an SVM classifier. The reason is that the margin itself can not well describe how good a kernel is due to the negligence of the scaling. We use the ratio between the margin and the radius of the minimum enclosing ball to measure the goodness of a kernel, and present a new minimization formulation for kernel learning. This formulation is invariant to scalings of learned kernels, and when learning linear combination of basis kernels it is also invariant to scalings of basis kernels and to the types (e.g., L1 or L2) of norm constraints on combination coefficients. We establish the differentiability of our formulation, and propose a gradient projection algorithm for kernel learning. Experiments show that our method significantly outperforms both SVM with the uniform combination of basis kernels and other state-of-art MKL approaches.

Subject Area: Supervised Learning

M9 Multi-label Multiple Kernel Learning by Stochastic Approximation: Application to Visual Object Recognition

Serhat S Bucak bucakser@msu.edu
Rong Jin rongjin@cse.msu.edu
Michigan State University
Anil K Jain jain@cse.msu.edu

Recent studies have shown that multiple kernel learning is very effective for object recognition leading to the popularity of kernel learning in computer vision problems. In this work we develop an efficient algorithm for multi-label multiple kernel learning (ML-MKL). We assume that all the classes under consideration share the same combination of kernel functions and the objective is to find the optimal kernel combination that benefits all the classes. Although several algorithms have been developed for ML-MKL their computational cost is linear in the number of classes making them unscalable when the number of classes is large a challenge frequently encountered in visual object recognition. We address this computational challenge by developing a framework for ML-MKL that combines the worst-case analysis with stochastic approximation. Our analysis shows that the complexity of our algorithm is $O(m^{1/3}\sqrt{lnm})$ where m is the number of classes. Empirical studies with object recognition show that while achieving similar classification accuracy the proposed method is significantly more efficient than the state-of-the-art algorithms for ML-MKL.

Subject Area: Supervised Learning

M10 Spectral Regularization for Support Estimation

Ernesto De Vito devito@dima.unige.it
DIMA Lorenzo Rosasco lrosasco@mit.edu
MIT and IIT
alessandro toigo Alessandro.Toigo@ge.infn.it
Unimi

In this paper we consider the problem of learning from data the support of a probability distribution when the distribution does not have a density (with respect to some reference measure). We propose a new class of regularized spectral estimators based on a new notion of reproducing kernel Hilbert space which we call "completely regular". Completely regular kernels allow to capture the relevant geometric and topological properties of an arbitrary probability space. In particular they are the key ingredient to prove the universal consistency of the spectral estimators and in this respect they are the analogue of universal kernels for supervised problems. Numerical experiments show that spectral estimators compare favorably to state of the art machine learning algorithms for density support estimation.

Subject Area: Supervised Learning

M11 Direct Loss Minimization for Structured Prediction

David A McAllester mcallester@ttic.edu
Tamir Hazan tamir@ttic.edu
Joseph Keshet jkeshet@ttic.edu
TTI Chicago

In discriminative machine learning one is interested in training a system to optimize a certain desired measure of performance, or loss. In binary classification one typically tries to minimizes the error rate. But in structured prediction each task often has its own measure of performance such as the BLEU score in machine translation or the intersection-over-union score in PASCAL segmentation. The most common approaches to structured prediction, structural SVMs and CRFs, do not minimize the task loss: the former minimizes a surrogate loss with no guarantees for task loss and the latter minimizes log loss independent of task loss. The main contribution of this paper is a theorem stating that a certain perceptron-like learning rule, involving features vectors derived from loss-adjusted inference, directly corresponds to the gradient of task loss. We give empirical results on phonetic alignment of a standard test set from the TIMIT corpus, which surpasses all previously reported results on this problem.
Subject Area: Supervised Learning

M12 Sidestepping Intractable Inference with Structured Ensemble Cascades

David Weiss djweiss@cis.upenn.edu
Benjamin J Sapp bensapp@cis.upenn.edu
Ben Taskar taskar@cis.upenn.edu
University of Pennsylvania

For many structured prediction problems, complex models often require adopting approximate inference techniques such as variational methods or sampling, which generally provide no satisfactory accuracy guarantees. In this work, we propose sidestepping intractable inference altogether by learning ensembles of tractable sub-models as part of a structured prediction cascade. We focus in particular on problems with high-treewidth and large state-spaces, which occur in many computer vision tasks. Unlike other variational methods, our ensembles do not enforce agreement between sub-models, but filter the space of possible outputs by simply adding and thresholding the max-marginals of each constituent model. Our framework jointly estimates parameters for all models in the ensemble for each level of the cascade by minimizing a novel, convex loss function, yet requires only a linear increase in computation over learning or inference in a single tractable sub-model. We provide a generalization bound on the filtering loss of the ensemble as a theoretical justification of our approach, and we evaluate our method on both synthetic data and the task of estimating articulated human pose from challenging videos. We find that our approach significantly outperforms loopy belief propagation on the synthetic data and a state-of-the-art model on the pose estimation/tracking problem.
Subject Area: Supervised Learning

M13 Layer-wise analysis of deep networks with Gaussian kernels

Gregoire Montavon g.montavon@gmail.com
Mikio L Braun braun@cs.tu-berlin.de
Klaus-Robert Müller krm@cs.tu-berlin.de
TU Berlin

Deep networks can potentially express a learning problem more efficiently than local learning machines. While deep networks outperform local learning machines on some problems, it is still unclear how their nice representation emerges from their complex structure. We present an analysis based on Gaussian kernels that measures how the representation of the learning problem evolves layer after layer as the deep network builds higher-level abstract representations of the input. We use this analysis to show empirically that deep networks build progressively better representations of the learning problem and that the best representations are obtained when the deep network discriminates only in the last layers.
Subject Area: Supervised Learning

M14 Scrambled Objects for Least-Squares Regression

Odalric Maillard odalric.maillard@inria.fr
Remi Munos remi.munos@inria.fr
INRIA Lille - Nord Europe

We consider least-squares regression using a randomly generated subspace $G_P \subset F$ of finite dimension P, where F is a function space of infinite dimension, e.g.$L_2([0, 1]^d)$. G_P is defined as the span of P random features that are linear combinations of the basis functions of F weighted by random Gaussian i.i.d. coefficients. In particular we consider multi-resolution random combinations at all scales of a given mother function such as a hat function or a wavelet. In this latter case the resulting Gaussian objects are called scrambled wavelets and we show that they enable to approximate functions in Sobolev spaces $H^s([01]^d)$. As a result given N data the least-squares estimate $\hat{g}$ built from P scrambled wavelets has excess risk $||f^* - \hat{g}||^2_{¶} = O(|| f^* ||^2_{H^s([01]^d)} (\log N)/P + P(\log N)/N$ for target functions $f^* \in H^s([01]^d)$ of smoothness order $s > d/2$. An interesting aspect of the resulting bounds is that they do not depend on the distribution ¶ from which the data are generated which is important in a statistical regression setting considered here. Randomization enables to adapt to any possible distribution. We conclude by describing an efficient numerical implementation using lazy expansions with numerical complexity $\tilde{O}(2^d N^{3/2} \log N + N^2)$ where d is the dimension of the input space.
Subject Area: Supervised Learning

M15 Block Variable Selection in Multivariate Regression and High-dimensional Causal Inference

Aurelie C Lozano aclozano@us.ibm.com
Vikas Sindhwani vsindhw@us.ibm.com
IBM Research

We consider multivariate regression problems involving high-dimensional predictor and response spaces. To efficiently address such problems, we propose a variable selection method, Multivariate Group Orthogonal Matching Pursuit, which extends the standard Orthogonal Matching Pursuit technique to account for arbitrary sparsity patterns induced by domain-specific groupings over both input and output variables, while also taking advantage of the correlation that may exist between the multiple outputs. We illustrate the utility of this framework for inferring causal relationships over a collection of highdimensional time series variables. When applied to time-evolving social media content, our models yield a new family of causality-based influence measures that may be seen as an alternative to PageRank. Theoretical guarantees, extensive simulations and empirical studies confirm the generality and value of our framework.
Subject Area: Supervised Learning

M16 Why are some word orders more common than others? A uniform information density account

Luke Maurits luke.maurits@adelaide.edu.au
Dan Navarro daniel.navarro@adelaide.edu.au
Amy Perfors amy.perfors@adelaide.edu.au
University of Adelaide

Languages vary widely in many ways, including their canonical word order. A basic aspect of the observed variation is the fact that some word orders are much more common than others. Although this regularity has been recognized for some time, it has not been wellexplained. In this paper we offer an information-theoretic explanation for the observed word-order distribution across languages, based on the concept of Uniform Information Density (UID). We suggest that object-first languages are particularly disfavored because they are highly non-optimal if the goal is to distribute information content approximately evenly throughout a sentence, and that the rest of the observed word-order distribution is at least partially explainable in terms of UID. We support our theoretical analysis with data from child-directed speech and experimental work.
Subject Area: Cognitive Science
Spotlight presentation, Monday, 6:30.

M17 Parametric Bandits: The Generalized Linear Case

Sarah Filippi filippi@telecom-paristech.fr
Aur´elien Garivier garivier@telecom-paristech.fr
Olivier Cappe cappe@telecom-paristech.fr
Telecom ParisTech stochastic optimization
Csaba Szepesvari szepesva@ualberta.ca
University of Alberta

We consider structured multi-armed bandit tasks in which the agent is guided by prior structural knowledge that can be exploited to efficiently select the optimal arm(s) in situations where the number of arms is large, or even infinite. We propose a new optimistic, UCB-like, algorithm for non-linearly parameterized bandit problems using the Generalized Linear Model (GLM) framework. We analyze the regret of the proposed algorithm, termed GLM-UCB, obtaining results similar to those recently proved in the literature for the linear regression case. The analysis also highlights a key difficulty of the non-linear case which is solved in GLM-UCB by focusing on the reward space rather than on the parameter space. Moreover, as the actual efficiency of current parameterized bandit algorithms is often deceiving in practice, we provide an asymptotic argument leading to significantly faster convergence. Simulation studies on real data sets illustrate the performance and the robustness of the proposed GLM-UCB approach.
Subject Area: Cognitive Science

M18 Spike timing-dependent plasticity as dynamic filter

Joscha T Schmiedt schmiedt@uni-bremen.de
University of Bremen
Christian Albers calbers@neuro.uni-bremen.de
Institute for Theoretical Physiks University of Bremen
Klaus Pawelzik pawelzik@neuro.uni-bremen.de

When stimulated with complex action potential sequences synapses exhibit spike timing-dependent plasticity (STDP) with attenuated and enhanced pre- and postsynaptic contributions to long-term synaptic modifications. In order to investigate the functional consequences of these contribution dynamics (CD) we propose a minimal model formulated in terms of differential equations. We find that our model reproduces a wide range of experimental results with a small number of biophysically interpretable parameters. The model allows to investigate the susceptibility of STDP to arbitrary time courses of pre- and postsynaptic activities, i.e. its nonlinear filter properties. We demonstrate this for the simple example of small periodic modulations of pre- and postsynaptic firing rates for which our model can be solved. It predicts synaptic strengthening for synchronous rate modulations. For low baseline rates modifications are dominant in the theta frequency range, a result which underlines the well known relevance of theta activities in hippocampus and cortex for learning. We also find emphasis of low baseline spike rates and suppression for high baseline rates. The latter suggests a mechanism of network activity regulation inherent in STDP. Furthermore, our novel formulation provides a general framework for investigating the joint dynamics of neuronal activity and the CD of STDP in both spike-based as well as rate-based neuronal network models.
Subject Area: Neuroscience

M19 Short-term memory in neuronal networks through dynamical compressed sensing

Surya Ganguli surya@phy.ucsf.edu
UCSF
Haim Sompolinsky haim@fiz.huji.ac.il
Hebrew University and Harvard University

Recent proposals suggest that large, generic neuronal networks could store memory traces of past input sequences in their instantaneous state. Such a proposal raises important theoretical questions about the duration of these memory traces and their dependence on network size, connectivity and signal statistics. Prior work, in the case of gaussian input sequences and linear neuronal networks, shows that the duration of memory traces in a network cannot exceed the number of neurons (in units of the neuronal time constant), and that no network can out-perform an equivalent feedforward network. However a more ethologically relevant scenario is that of sparse input sequences. In this scenario, we show how linear neural networks can essentially perform compressed sensing (CS) of past inputs, thereby attaining a memory capacity that exceeds the number of neurons. This enhanced capacity is achieved by a class of "orthogonal" recurrent networks and not by feedforward networks or generic recurrent networks. We exploit techniques from the statistical physics of disordered systems to analytically compute the decay of memory traces in such networks as a function of network size, signal sparsity and integration time. Alternately, viewed purely from the perspective of CS, this work introduces a new ensemble of measurement matrices derived from dynamical systems, and provides a theoretical analysis of their asymptotic performance.
Subject Area: Neuroscience

M20 Hallucinations in Charles Bonnet Syndrome Induced by Homeostasis: a Deep Boltzmann Machine Model

David P Reichert d.p.reichert@sms.ed.ac.uk
Peggy Series pseries@inf.ed.ac.uk
Amos J Storkey a.storkey@ed.ac.uk
University of Edinburgh

The Charles Bonnet Syndrome (CBS) is characterized by complex vivid visual hallucinations in people with, primarily, eye diseases and no other neurological pathology. We present a Deep Boltzmann Machine model of CBS, exploring two core hypotheses: First, that the visual cortex learns a generative or predictive model of sensory input, thus explaining its capability to generate internal imagery. And second, that homeostatic mechanisms stabilize neuronal activity levels, leading to hallucinations being formed when input is lacking. We reproduce a variety of qualitative findings in CBS. We also introduce a modification to the DBM that allows us to model a possible role of acetylcholine in CBS as mediating the balance of feed-forward and feed-back processing. Our model might provide new insights into CBS and also demonstrates that generative frameworks are promising as hypothetical models of cortical learning and perception.
Subject Area: Neuroscience

M21 Mixture of time-warped trajectory models for movement decoding

Elaine A Corbett ecorbett@u.northwestern.edu
Eric J Perreault e-perreault@northwestern.edu
Konrad Koerding kk@northwestern.edu
Northwestern University

Applications of Brain-Machine-Interfaces typically estimate user intent based on biological signals that are under voluntary control. For example, we might want to estimate how a patient with a paralyzed arm wants to move based on residual muscle activity. To solve such problems it is necessary to integrate obtained information over time. To do so, state of the art approaches typically use a probabilistic model of how the state, e.g. position and velocity of the arm, evolves over time – a so-called trajectory model. We wanted to further develop this approach using two intuitive insights: (1) At any given point of time there may be a small set of likely movement targets, potentially identified by the location of objects in the workspace or by gaze information from the user. (2) The user may want to produce movements at varying speeds. We thus use a generative model with a trajectory model incorporating these insights. Approximate inference on that generative model is implemented using a mixture of extended Kalman filters. We find that the resulting algorithm allows us to decode arm movements dramatically better than when we use a trajectory model with linear dynamics.
Subject Area: Neuroscience

M22 Decoding Ipsilateral Finger Movements from ECoG Signals in Humans

Yuzong Liu yuzongliu@wustl.edu
Mohit Sharma mohit.sharma@wustl.edu
Charles M Gaona cgaona@wustl.edu
Jonathan D Breshears jbreshears@wustl.edu
jarod Roland jarod.roland@gmail.com
Zachary V Freudenburg voges78@wustl.edu
Kilian Q Weinberger kilian@wustl.edu
Eric C Leuthardt leuthardte@wustl.edu
Washington University St. Louis

Several motor related Brain Computer Interfaces (BCIs) have been developed over the years that use activity decoded from the contralateral hemisphere to operate devices. Many recent studies have also talked about the importance of ipsilateral activity in planning of motor movements. For successful upper limb BCIs, it is important to decode finger movements from brain activity. This study uses ipsilateral cortical signals from humans (using ECoG) to decode finger movements. We demonstrate, for the first time, successful finger movement detection using machine learning algorithms. Our results show high decoding accuracies in all cases which are always above chance. We also show that significant accuracies can be achieved with the use of only a fraction of all the features recorded and that these core features also make sense physiologically. The results of this study have a great potential in the emerging world of motor neuroprosthetics and other BCIs.
Subject Area: Neuroscience

M23 b-Bit Minwise Hashing for Estimating Three-Way Similarities

Ping J Li pingli@cornell.edu
Cornell Arnd C Konig chrisko@microsoft.com
Microsoft Research
Wenhao Gui wg58@cornell.edu

Computing two-way and multi-way set similarities is a fundamental problem. This study focuses on estimating 3-way resemblance (Jaccard similarity) using b-bit minwise hashing. While traditional minwise hashing methods store each hashed value using 64 bits, b-bit minwise hashing only stores the lowest b bits (where b¿= 2 for 3-way). The extension to 3-way similarity from the prior work on 2-way similarity is technically non-trivial. We develop the precise estimator which is accurate and very complicated; and we recommend a much simplified estimator suitable for sparse data. Our analysis shows that b-bit minwise hashing can normally achieve a 10 to 25-fold improvement in the storage space required for a given estimator accuracy of the 3-way resemblance.
Subject Area: Theory
Spotlight presentation, Monday, 6:30.

M24 A Computational Decision Theory for Interactive Assistants

Alan Fern afern@eecs.orst.edu
Prasad Tadepalli tadepalli@cs.orst.edu
Oregon State University

We study several classes of interactive assistants from the points of view of decision theory and computational complexity. We first introduce a class of POMDPs called hiddengoal MDPs (HGMDPs), which formalize the problem of interactively assisting an agent whose goal is hidden and whose actions are observable. In spite of its restricted nature, we show that optimal action selection in finite horizon HGMDPs is PSPACE-complete even in domains with deterministic dynamics. We then introduce a more restricted model called helper action MDPs (HAMDPs), where the assistant's action is accepted by the agent when it is helpful, and can be easily ignored by the agent otherwise. We show classes of HAMDPs that are complete for PSPACE and NP along with a polynomial time class. Furthermore, we show that for general HAMDPs a simple myopic policy achieves a regret, compared to an omniscient assistant, that is bounded by the entropy of the initial goal distribution. A variation of this policy is shown to achieve worst-case regret that is logarithmic in the number of goals for any goal distribution.
Subject Area: Theory

M25 Multi-Stage Dantzig Selector

Ji Liu ji.liu@asu.edu
Peter Wonka peter.wonka@asu.edu
Jieping Ye jieping.ye@asu.edu
Arizona State University

We consider the following sparse signal recovery (or feature selection) problem: given a design matrix $X \in \mathbb{R}^{n\times m}$ $(m \gg n)$ and a noisy observation vector $y \in \mathbb{R}^n$ satisfying $y = X\beta^* + \epsilon$ where ϵ is the noise vector following a Gaussian distribution $N(0, \sigma^2 I)$, how to recover the signal (or parameter vector) β^* when the signal is sparse? The Dantzig selector has been proposed for sparse signal recovery with strong theoretical guarantees. In this paper we propose a multi-stage Dantzig selector method which iteratively refines the target signal β^*. We show that if X obeys a certain condition then with a large probability the difference between the solution β estimated by the proposed method and the true solution β^* measured in terms of the l_p norm $(p \geq 1)$ is bounded as

$$||\hat{\beta} - \hat{\beta}^*||_p \leq (C_{(s - N)}^{1/p} \sqrt{log\ m} + \Delta)\sigma$$

C is a constant s is the number of nonzero entries in β^* Δ is independent of m and is much smaller than the first term and N is the number of entries of β^* larger than a certain value in the order of $\mathcal{O}(\sigma \sqrt{log\ m})$. The proposed method improves the estimation bound of the standard Dantzig selector approximately from $Cs^{1/p} \sqrt{log\ m}\sigma$ to $C_{(s - N)}^{1/p} \sqrt{log\ m}\sigma$ where the value N depends on the number of large entries in β^*. When N = s the proposed algorithm achieves the oracle solution with a high probability. In addition with a large probability the proposed method can select the same number of correct features under a milder condition than the Dantzig selector.
Subject Area: Theory

M26 Estimation of Renyi Entropy and Mutual Information Based on Generalized Nearest-Neighbor Graphs

David Pal dpal@cs.ualberta.ca
Csaba Szepesvari szepesva@ualberta.ca
University of Alberta
Barnabas Poczos poczos@ualberta.ca
Carnegie Mellon University

We present simple and computationally efficient non-parametric estimators of R´enyi entropy and mutual information based on an i.i.d. sample drawn from an unknown, absolutely continuous distribution over R^d. The estimators are calculated as the sum of p-th powers of the Euclidean lengths of the edges of the 'generalized nearest-neighbor' graph of the sample and the empirical copula of the sample respectively. For the first time, we prove the almost sure consistency of these estimators and upper bounds on their rates of convergence, the latter of which under the assumption that the density underlying the sample is Lipschitz continuous. Experiments demonstrate their usefulness in independent subspace analysis.
Subject Area: Theory

M27 Probabilistic Belief Revision with Structural Constraints

Peter Jones jonep923@mit.edu
Sanjoy K Mitter mitter@mit.edu
MIT
Venkatesh Saligrama srv@bu.edu
Boston University

Experts (human or computer) are often required to assess the probability of uncertain events. When a collection of experts independently assess events that are structurally interrelated, the resulting assessment may violate fundamental laws of probability. Such an assessment is termed incoherent. In this work we investigate how the problem of incoherence may be affected by allowing experts to specify likelihood models and then update their assessments based on the realization of a globally-observable random sequence.
Subject Area: Theory

M28 Getting lost in space: Large sample analysis of the resistance distance

Ulrike von Luxburg ulrike.luxburg@tuebingen.mpg.de
Agnes Radl agnes. radl@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics
Matthias Hein hein@cs.uni-sb.de
Saarland University

The commute distance between two vertices in a graph is the expected time it takes a random walk to travel from the first to the second vertex and back. We study the behavior of the commute distance as the size of the underlying graph increases. We prove that the commute distance converges to an expression that does not take into account the structure of the graph at all and that is completely meaningless as a distance function on the graph. Consequently the use of the raw commute distance for machine learning purposes is strongly discouraged for large graphs and in high dimensions. As an alternative we introduce the amplified commute distance that corrects for the undesired large sample effects.
Subject Area: Theory
Spotlight presentation, Monday, 6:30.

M29 Bayesian Action-Graph Games

Albert Xin Jiang jiang@cs.ubc.ca
Kevin Leyton-Brown kevinlb@cs.ubc.ca
University of British Columbia

Games of incomplete information, or Bayesian games, are an important game-theoretic model and have many applications in economics. We propose Bayesian action-graph games (BAGGs) a novel graphical representation for Bayesian games. BAGGs can represent arbitrary Bayesian games and furthermore can compactly express Bayesian games exhibiting commonly encountered types of structure including symmetry actionand type-specific utility independence and probabilistic independence of type distributions. We provide an algorithm for computing expected utility in BAGGs and discuss conditions under which the algorithm runs in polynomial time. Bayes-Nash equilibria of BAGGs can be computed by adapting existing algorithms for complete-information normal form games and leveraging our expected utility algorithm. We show both theoretically and empirically that our approaches improve significantly on the state of the art.
Subject Area: Theory

M30 Random Conic Pursuit for Semidefinite Programming

Ariel Kleiner akleiner@cs.berkeley.edu
Michael I Jordan jordan@eecs.berkeley.edu
University of California Berkeley
Ali Rahimi ali.rahimi@intel.com
Intel

We present a novel algorithm, Random Conic Pursuit, that solves semidefinite programs (SDPs) via repeated optimization over randomly selected two-dimensional subcones of the PSD cone. This scheme is simple, easily implemented, applicable to very general SDPs, scalable, and theoretically interesting. Its advantages are realized at the expense of an ability to readily compute highly exact solutions, though useful approximate solutions are easily obtained. This property renders Random Conic Pursuit of particular interest for machine learning applications, in which the relevant SDPs are generally based upon random data and so exact minima are often not a priority. Indeed, we present empirical results to this effect for various SDPs encountered in machine learning; these experiments demonstrate the potential practical usefulness of Random Conic Pursuit. We also provide a preliminary analysis that yields insight into the theoretical properties and convergence of the algorithm.
Subject Area: Optimization

M31 Generalized roof duality and bisubmodular functions

Vladimir Kolmogorov v.kolmogorov@cs.ucl.ac.uk
UC London

Consider a convex relaxation f of a pseudo-boolean function f. We say that the relaxation is totally half-integral if $f(x)$ is a polyhedral function with half-integral extreme points x, and this property is preserved after adding an arbitrary combination of constraints of the form $x_i = x_j$, $x_i = 1 - x_j$, and $x_i = \gamma$ where $\gamma \in \{0, 1, \frac{1}{2}\}$ is a constant. A well-known example is the roof duality relaxation for quadratic pseudo-boolean functions f. We argue that total half-integrality is a natural requirement for generalizations of roof duality to arbitrary pseudo-boolean functions. Our contributions are as follows. First we provide a complete characterization of totally half-integral relaxations f by establishing a one-to one correspondence with bisubmodular functions. Second we give a new characterization of bisubmodular functions. Finally we show some relationships between general totally half-integral relaxations and relaxations based on the roof duality.
Subject Area: Optimization

M32 Sparse Inverse Covariance Selection via Alternating Linearization Methods

Katya Scheinberg ks79@columbia.edu
Shiqian Ma sm2756@columbia.edu
Donald Goldfarb goldfarb@columbia.edu
Columbia University

Gaussian graphical models are of great interest in statistical learning. Because the conditional independencies between different nodes correspond to zero entries in the inverse covariance matrix of the Gaussian distribution one can learn the structure of the graph by estimating a sparse inverse covariance matrix from sample data by solving a convex maximum likelihood problem with an ℓ_1-regularization term. In this paper we propose a first-order method based on an alternating linearization technique that exploits the problem's special structure; in particular the subproblems solved in each iteration have closed-form solutions. Moreover our algorithm obtains an $\in$-optimal solution in O(1/ $\in$) iterations. Numerical experiments on both synthetic and real data from gene association networks show that a practical version of this algorithm outperforms other competitive algorithms.
Subject Area: Optimization

M33 Optimal Web-Scale Tiering as a Flow Problem

Gilbert Leung gleung@alum.mit.edu
eBay Inc.
Novi Quadrianto novi.quad@gmail.com
SML-NICTA and RSISE-ANU
Alexander J Smola alex@smola.org
Yahoo! Research
Kostas Tsioutsiouliklis kostas@yahoo-inc.com
Yahoo! Labs

We present a fast online solver for large scale maximum-flow problems as they occur in portfolio optimization, inventory management, computer vision, and logistics. Our algorithm solves an integer linear program in an online fashion. It exploits total unimodularity of the constraint matrix and a Lagrangian relaxation to solve the problem as a convex online game. The algorithm generates approximate solutions of max-flow problems by performing stochastic gradient descent on a set of flows. We apply the algorithm to optimize tier arrangement of over 80 Million web pages on a layered set of caches to serve an incoming query stream optimally. We provide an empirical demonstration of the effectiveness of our method on real query-pages data.
Subject Area: Optimization

M34 Parallelized Stochastic Gradient Descent

Martin Zinkevich maz@yahoo-inc.com
Markus Weimer weimer@yahoo-inc.com
Alex Smola smola@yahoo-inc.com
Lihong Li lihong@yahoo-inc.com
Yahoo!

Research With the increase in available data parallel machine learning has become an increasingly pressing problem. In this paper we present the first parallel stochastic gradient descent algorithm including a detailed analysis and experimental evidence. Unlike prior work on parallel optimization algorithms our variant comes with parallel acceleration guarantees and it poses no overly tight latency constraints which might only be available in the multicore setting. Our analysis introduces a novel proof technique — contractive mappings to quantify the speed of convergence of parameter distributions to their asymptotic limits. As a side effect this answers the question of how quickly stochastic gradient descent algorithms reach the asymptotically normal regime.
Subject Area: Optimization

M35 Non-Stochastic Bandit Slate Problems

Satyen Kale skale@yahoo-inc.com
Yahoo!
Lev Reyzin lreyzin@cc.gatech.edu
Georgia Institute of Technology
Robert E Schapire schapire@cs.princeton.edu
Princeton University

We consider bandit problems, motivated by applications in online advertising and news story selection in which the learner must repeatedly select a slate that is a subset of size s from K possible actions and then receives rewards for just the selected actions. The goal is to minimize the regret with respect to total reward of the best slate computed in hindsight. We consider unordered and ordered versions of the problem and give efficient algorithms which have regret O(sqrt(T)) where the constant depends on the specific nature of the problem. We also consider versions of the problem where we have access to a number of policies which make recommendations for slates in every round and give algorithms with O(sqrt(T)) regret for competing with the best such policy as well. We make use of the technique of relative entropy projections combined with the usual multiplicative weight update algorithm to obtain our algorithms.
Subject Area: Theory

M36 Repeated Games against Budgeted Adversaries

Jacob Abernethy jake@cs.berkeley.edu
UC Berkeley
Manfred Warmuth manfred@cse.ucsc.edu
UC santa Cruz

We study repeated zero-sum games against an adversary on a budget. Given that an adversary has some constraint on the sequence of actions that he plays, we consider what ought to be the player's best mixed strategy with knowledge of this budget. We show that, for a general class of normal-form games, the minimax strategy is indeed efficiently computable and relies on a ”random playout” technique. We give three diverse applications of this algorithmic template: a cost-sensitive ”Hedge” setting a particular problem in Metrical Task Systems and the design of combinatorial prediction markets.
Subject Area: Theory

M37 Two-Layer Generalization Analysis for Ranking Using Rademacher Average

Wei Chen chenwei@amss.ac.cn
Zhi-Ming Ma mazm@amt.ac.cn
Chinese Academy of Sciences
Tie-Yan Liu tyliu@microsoft.com
Microsoft Research

This paper is concerned with the generalization analysis on learning to rank for information retrieval (IR). In IR, data are hierarchically organized, i.e., consisting of queries and documents per query. Previous generalization analysis for ranking, however, has not fully considered this structure, and cannot explain how the simultaneous change of query number and document number in the training data will affect the performance of algorithms. In this paper, we propose performing generalization analysis under the assumption of two-layer sampling, i.e., the i.i.d. sampling of queries and the conditional i.i.d sampling of documents per query. Such a sampling can better describe the generation mechanism of real data, and the corresponding generalization analysis can better explain the real behaviors of learning to rank algorithms. However, it is challenging to perform such analysis, because the documents associated with different queries are not identically distributed, and the documents associated with the same query become no longer independent if represented by features extracted from the matching between document and query. To tackle the challenge, we decompose the generalization error according to the two layers, and make use of the new concept of two-layer Rademacher average. The generalization bounds we obtained are quite intuitive and are in accordance with previous empirical studies on the performance of ranking algorithms.
Subject Area: Theory

M38 Empirical Bernstein Inequalities for U-Statistics

Thomas Peel thomas.peel@lif.univ-mrs.fr
Liva Ralaivola liva.ralaivola@lif.univ-mrs.fr
Sandrine Anthoine anthoine@cmi.univ-mrs.fr
LIF Aix-Marseille Universite

We present original empirical Bernstein inequalities for U-statistics with bounded symmetric kernels q. They are expressed with respect to empirical estimates of either the variance of q or the conditional variance that appears in the Bernstein-type inequality for U-statistics derived by Arcones [2]. Our result subsumes other existing empirical Bernstein inequalities, as it reduces to them when U-statistics of order 1 are considered. In addition, it is based on a rather direct argument using two applications of the same (non-empirical) Bernstein inequality for U-statistics. We discuss potential applications of our new inequalities, especially in the realm of learning ranking/scoring functions. In the process we exhibit an efficient procedure to compute the variance estimates for the special case of bipartite ranking that rests on a sorting argument. We also argue that our results may provide test set bounds and particularly interesting empirical racing algorithms for the problem of online learning of scoring functions.
Subject Area: Theory

M39 On the Theory of Learnining with Privileged Information

Dmitry Pechyony pechyony@nec-labs.com
NEC Labs
Vladimir Vapnik vapnik@att.net

In Learning Using Privileged Information (LUPI) paradigm, along with the standard training data in the decision space a teacher supplies a learner with the privileged information in the correcting space. The goal of the learner is to find a classifier with a low generalization error in the decision space. We consider a new version of empirical risk minimization algorithm called Privileged ERM that takes into account the privileged information in order to find a good function in the decision space. We outline the conditions on the correcting space that if satisfied allow Privileged ERM to have much faster learning rate in the decision space than the one of the regular empirical risk minimization.
Subject Area: Theory

M40 Interval Estimation for Reinforcement-Learning Algorithms in Continuous-State Domains

Martha White whitem@ualberta.ca
Adam M White amw8@ualberta.ca
University of Alberta

The reinforcement learning community has explored many approaches to obtain- ing value estimates and models to guide decision making; these approaches, how- ever, do not usually provide a measure of confidence in the estimate. Accurate estimates of an agent's confidence are useful for many applications, such as bi- asing exploration and automatically adjusting parameters to reduce dependence on parameter-tuning. Computing confidence intervals on reinforcement learning value estimates, however, is challenging because data generated by the agent- environment interaction rarely satisfies traditional assumptions. Samples of value- estimates are dependent, likely non-normally distributed and often limited, particularly in early learning when confidence estimates are pivotal. In this work, we investigate how to compute robust confidences for value estimates in continuous Markov decision processes. We illustrate how to use bootstrapping to compute confidence intervals online under a changing policy (previously not possible) and prove validity under a few reasonable assumptions. We demonstrate the applica- bility of our confidence estimation algorithms with experiments on exploration, parameter estimation and tracking.
Subject Area: Control and Reinforcement Learning

M41 Monte-Carlo Planning in Large POMDPs

David Silver davidstarsilver@googlemail.com
MIT
Joel Veness jveness@gmail.com
University of New South Wales

This paper introduces a Monte-Carlo algorithm for online planning in large POMDPs. The algorithm combines a Monte-Carlo update of the agent’s belief state with a Monte-Carlo tree search from the current belief state. The new algorithm, POMCP, has two important properties. First, Monte-Carlo sampling is used to break the curse of dimensionality both during belief state updates and during planning. Second, only a black box simulator of the POMDP is required, rather than explicit probability distributions. These properties enable POMCP to plan effectively in significantly larger POMDPs than has previously been possible. We demonstrate its effectiveness in three large POMDPs. We scale up a well-known benchmark problem, Rocksample, by several orders of magnitude. We also introduce two challenging new POMDPs: 10x10 Battleship and Partially Observable PacMan, with approximately 1018 and 1056 states respectively. Our Monte-Carlo planning algorithm achieved a high level of performance with no prior knowledge, and was also able to exploit simple domain knowledge to achieve better results with less search. POMCP is the first general purpose planner to achieve high performance in such large and unfactored POMDPs.

Subject Area: Control and Reinforcement Learning

M42 Basis Construction from Power Series Expansions of Value Functions

Sridhar Mahadevan mahadeva@cs.umass.edu
Bo Liu boliu@cs.umass.edu
University of Massachusetts Amherst

This paper explores links between basis construction methods in Markov decision processes and power series expansions of value functions. This perspective provides a useful framework to analyze properties of existing bases, as well as provides insight into constructing more effective bases. Krylov and Bellman error bases are based on the Neumann series expansion. These bases incur very large initial Bellman errors, and can converge rather slowly as the discount factor approaches unity. The Laurent series expansion, which relates discounted and average-reward formulations, provides both an explanation for this slow convergence as well as suggests a way to construct more efficient basis representations. The first two terms in the Laurent series represent the scaled average-reward and the average-adjusted sum of rewards, and subsequent terms expand the discounted value function using powers of a generalized inverse called the Drazin (or group inverse) of a singular matrix derived from the transition matrix. Experiments show that Drazin bases converge considerably more quickly than several other bases, particularly for large values of the discount factor. An incremental variant of Drazin bases called Bellman average-reward bases (BARBs) is described, which provides some of the same benefits at lower computational cost.

Subject Area: Control and Reinforcement Learning.

M43 Reward Design via Online Gradient Ascent

Jonathan D Sorg jdsorg@umich.edu
Satinder Singh baveja@umich.edu
Richard L Lewis rickl@umich.edu
University of Michigan

Recent work has demonstrated that when artificial agents are limited in their ability to achieve their goals, the agent designer can benefit by making the agent’s goals different from the designer’s. This gives rise to the optimization problem of designing the artificial agent’s goals—in the RL framework, designing the agent’s reward function. Existing attempts at solving this optimal reward problem do not leverage experience gained online during the agent’s lifetime nor do they take advantage of knowledge about the agent’s structure. In this work, we develop a gradient ascent approach with formal convergence guarantees for approximately solving the optimal reward problem online during an agent’s lifetime. We show that our method generalizes a standard policy gradient approach, and we demonstrate its ability to improve reward functions in agents with various forms of limitations.

Subject Area: Control and Reinforcement Learning

M44 Bootstrapping Apprenticeship Learning

Abdeslam Boularias boularias@gmail.com
MPI for Biological Cybernetics
Brahim Chaib-draa chaib@ift.ulaval.ca

We consider the problem of apprenticeship learning where the examples, demonstrated by an expert, cover only a small part of a large state space. Inverse Reinforcement Learning (IRL) provides an efficient tool for generalizing the demonstration, based on the assumption that the expert is maximizing a utility function that is a linear combination of state-action features. Most IRL algorithms use a simple Monte Carlo estimation to approximate the expected feature counts under the expert’s policy. In this paper, we show that the quality of the learned policies is highly sensitive to the error in estimating the feature counts. To reduce this error, we introduce a novel approach for bootstrapping the demonstration by assuming that: (i), the expert is (near-)optimal, and (ii), the dynamics of the system is known. Empirical results on gridworlds and car racing problems show that our approach is able to learn good policies from a small number of demonstrations.

Subject Area: Control and Reinforcement Learning

M45 PAC-Bayesian Model Selection for Reinforcement Learning

Mahdi Milani Fard mahdi.milanifard@mail.mcgill.ca
Joelle Pineau jpineau@cs.mcgill.ca
McGill University

This paper introduces the first set of PAC-Bayesian bounds for the batch reinforcement learning problem in finite state spaces. These bounds hold regardless of the correctness of the prior distribution. We demonstrate how such bounds can be used for model-selection in control problems where prior information is available either on the dynamics of the environment, or on the value of actions. Our empirical results confirm that PAC-Bayesian model-selection is able to leverage prior distributions when they are informative and, unlike standard Bayesian RL approaches, ignores them when they are misleading.

Subject Area: Control and Reinforcement Learning

M46 An Approximate Inference Approach to Temporal Optimization in Optimal Control

Konrad C Rawlik k.c.rawlik@ed.ac.uk
Sethu Vijayakumar sethu.vijayakumar@ed.ac.uk
University of Edinburgh
Marc Toussaint mtoussai@cs.tu-berlin.de
TU Berlin

Algorithms based on iterative local approximations present a practical approach to optimal control in robotic systems. However, they generally require the temporal parameters (for e.g. the movement duration or the time point of reaching an intermediate goal) to be specified a priori. Here, we present a methodology that is capable of jointly optimising the temporal parameters in addition to the control command profiles. The presented approach is based on a Bayesian canonical time formulation of the optimal control problem, with the temporal mapping from canonical to real time parametrised by an additional control variable. An approximate EM algorithm is derived that efficiently optimises both the movement duration and control commands offering, for the first time, a practical approach to tackling generic via point problems in a systematic way under the optimal control framework. The proposed approach is evaluated on simulations of a redundant robotic plant.

Subject Area: Control and Reinforcement Learning

M47 Nonparametric Bayesian Policy Priors for Reinforcement Learning Finale

Doshi-Velez finale@mit.edu
David Wingate wingated@mit.edu
Nicholas Roy nickroy@mit.edu
Joshua Tenenbaum jbt@mit.edu
Massachusetts Institute of Technology

We consider reinforcement learning in partially observable domains where the agent can query an expert for demonstrations. Our nonparametric Bayesian approach combines model knowledge, inferred from expert information and independent exploration, with policy knowledge inferred from expert trajectories. We introduce priors that bias the agent towards models with both simple representations and simple policies, resulting in improved policy and model learning.

Subject Area: Control and Reinforcement Learning

M48 Predictive State Temporal Difference Learning

Byron Boots beb@cs.cmu.edu
Geoff Gordon ggordon@cs.cmu.edu
Carnegie Mellon University

We propose a new approach to value function approximation which combines linear temporal difference reinforcement learning with subspace identification. In practical applications, reinforcement learning (RL) is complicated by the fact that state is either high-dimensional or partially observable. Therefore, RL methods are designed to work with features of state rather than state itself, and the success or failure of learning is often determined by the suitability of the selected features. By comparison, subspace identification (SSID) methods are designed to select a feature set which preserves as much information as possible about state. In this paper we connect the two approaches, looking at the problem of reinforcement learning with a large set of features, each of which may only be marginally useful for value function approximation. We introduce a new algorithm for this situation, called Predictive State Temporal Difference (PSTD) learning. As in SSID for predictive state representations, PSTD finds a linear compression operator that projects a large set of features down to a small set that preserves the maximum amount of predictive information. As in RL, PSTD then uses a Bellman recursion to estimate a value function. We discuss the connection between PSTD and prior approaches in RL and SSID. We prove that PSTD is statistically consistent, perform several experiments that illustrate its properties, and demonstrate its potential on a difficult optimal stopping problem.

Subject Area: Control and Reinforcement Learning

M49 Double Q-learning

Hado P van Hasselt h.p.van.hasselt@gmail.com
Center for Mathematics and Computer Science

In some stochastic environments the well-known reinforcement learning algorithm Q-learning performs very poorly. This poor performance is caused by large overestimations of action values. These overestimations result from a positive bias that is introduced because Q-learning uses the maximum action value as an approximation for the maximum expected action value. We introduce an alternative way to approximate the maximum expected value for any set of random variables. The obtained double estimator method is shown to sometimes underestimate rather than overestimate the maximum expected value. We apply the double estimator to Q-learning to construct Double Q-learning, a new off-policy reinforcement learning algorithm. We show the new algorithm converges to the optimal policy and that it performs well in some settings in which Q-learning performs poorly due to its overestimation.

Subject Area: Control and Reinforcement Learning

M50 Error Propagation for Approximate Policy and Value Iteration

Amir-massoud Farahmand amirf@ualberta.ca
Csaba Szepesvari szepesva@ualberta.ca
University of Alberta
Remi Munos remi.munos@inria.fr
INRIA Lille - Nord Europe

We address the question of how the approximation error/ Bellman residual at each iteration of the Approximate Policy/Value Iteration algorithms influences the quality of the resulted policy. We quantify the performance loss as the Lp norm of the approximation error/Bellman residual at each iteration. Moreover we show that the performance loss depends on the expectation of the squared Radon-Nikodym derivative of a certain distribution rather than its supremum – as opposed to what has been suggested by the previous results. Also our results indicate that the contribution of the approximation/Bellman error to the performance loss is more prominent in the later iterations of API/AVI and the effect of an error term in the earlier iterations decays exponentially fast.

Subject Area: Control and Reinforcement Learning

M51 Multiparty Differential Privacy via Aggregation of Locally Trained Classifiers

Manas A Pathak manasp@cs.cmu.edu
CMU Shantanu Rane rane@merl.com
Mitsubishi Electric Research Labs
Bhiksha Raj bhiksha@cs.cmu.edu
Carnegie Mellon University

As increasing amounts of sensitive personal information finds its way into data repositories, it is important to develop analysis mechanisms that can derive aggregate information from these repositories without revealing information about individual data instances. Though the differential privacy model provides a framework to analyze such mechanisms for databases belonging to a single party this framework has not yet been considered in a multi-party setting. In this paper we propose a privacy-preserving protocol for composing a differentially private aggregate classifier using classifiers trained locally by separate mutually untrusting parties. The protocol allows these parties to interact with an untrusted curator to construct additive shares of a perturbed aggregate classifier. We also present a detailed theoretical analysis containing a proof of differential privacy of the perturbed aggregate classifier and a bound on the excess risk introduced by the perturbation. We verify the bound with an experimental evaluation on a real dataset.

Subject Area:

M52 A New Probabilistic Model for Rank Aggregation

Tao Qin taoqin@microsoft.com
Xiubo Geng xiubogeng@gmail.com
Chinese Academy of Sciences
Tie-Yan Liu tyliu@microsoft.com
Microsoft Research

This paper is concerned with rank aggregation, which aims to combine multiple input rankings to get a better ranking. A popular approach to rank aggregation is based on probabilistic models on permutations, e.g., the Luce model and the Mallows model. However, these models have their limitations in either poor expressiveness or high computational complexity. To avoid these limitations, in this paper, we propose a new model, which is defined with a coset-permutation distance, and models the generation of a permutation as a stagewise process. We refer to the new model as coset-permutation distance based stagewise (CPS) model. The CPS model has rich expressiveness and can therefore be used in versatile applications, because many different permutation distances can be used to induce the coset-permutation distance. The complexity of the CPS model is low because of the stagewise decomposition of the permutation probability and the efficient computation of most coset-permutation distances. We apply the CPS model to supervised rank aggregation, derive the learning and inference algorithms, and empirically study their effectiveness and efficiency. Experiments on public datasets show that the derived algorithms based on the CPS model can achieve state-of-the-art ranking accuracy, and are much more efficient than previous algorithms.

Subject Area: Probabilistic Models and Methods

M53 The Maximal Causes of Natural Scenes are Edge Filters

Jose G Puertas cidcampeador0@gmail.com
Jorg Bornschein bornschein@fias.uni-frankfurt.de
Jorg Lucke luecke@fias.uni-frankfurt.de
University of Frankfurt

We study the application of a strongly non-linear generative model to image patches. As in standard approaches such as Sparse Coding or Independent Component Analysis, the model assumes a sparse prior with independent hidden variables. However, in the place where standard approaches use the sum to combine basis functions we use the maximum. To derive tractable approximations for parameter estimation we apply a novel approach based on variational Expectation Maximization. The derived learning algorithm can be applied to large-scale problems with hundreds of observed and hidden variables. Furthermore, we can infer all model parameters including observation noise and the degree of sparseness. In applications to image patches we find that Gabor-like basis functions are obtained. Gabor-like functions are thus not a feature exclusive to approaches assuming linear superposition. Quantitatively, the inferred basis functions show a large diversity of shapes with many strongly elongated and many circular symmetric functions. The distribution of basis function shapes reflects properties of simple cell receptive fields that are not reproduced by standard linear approaches. In the study of natural image statistics, the implications of using different superposition assumptions have so far not been investigated systematically because models with strong non-linearities have been found analytically and computationally challenging. The presented algorithm represents the first large-scale application of such an approach.

Subject Area: Probabilistic Models and Methods

M54 Inference with Multivariate Heavy-Tails in Linear Models

Danny Bickson bickson@cs.cmu.edu
Carlos Guestrin guestrin@cs.cmu.edu
Carnegie Mellon University

Heavy-tailed distributions naturally occur in many real life problems. Unfortunately, it is typically not possible to compute inference in closed-form in graphical models which involve such heavy tailed distributions. In this work we propose a novel simple linear graphical model for independent latent random variables called linear characteristic model (LCM) defined in the characteristic function domain. Using stable distributions a heavy-tailed family of distributions which is a generalization of Cauchy L´evy and Gaussian distributions we show for the first time how to compute both exact and approximate inference in such a linear multivariate graphical model. LCMs are not limited to only stable distributions in fact LCMs are always defined for any random variables (discrete continuous or a mixture of both). We provide a realistic problem from the field of computer networks to demonstrate the applicability of our construction. Other potential application is iterative decoding of linear channels with non-Gaussian noise.
Subject Area: Probabilistic Models and Methods

M55 A Bayesian Approach to Concept Drift

Stephen H Bach bach@cs.georgetown.edu
Mark Maloof maloof@cs.georgetown.edu
Georgetown University

To cope with concept drift, we placed a probability distribution over the location of the most-recent drift point. We used Bayesian model comparison to update this distribution from the predictions of models trained on blocks of consecutive observations and pruned potential drift points with low probability. We compare our approach to a non-probabilistic method for drift and a probabilistic method for change-point detection. In our experiments, our approach generally yielded improved accuracy and/or speed over these other methods.
Subject Area: Probabilistic Models and Methods

M56 Auto-Regressive HMM Inference with Incomplete Data for Short-Horizon Wind Forecasting

Chris Barber barberchris01@gmail.com
Joseph Bockhorst joebock@uwm.edu
Paul Roebber roebber@uwm.edu
UW Milwaukee

Accurate short-term wind forecasts (STWFs), with time horizons from 0.5 to 6 hours, are essential for efficient integration of wind power to the electrical power grid. Physical models based on numerical weather predictions are currently not competitive, and research on machine learning approaches is ongoing. Two major challenges confronting these efforts are missing observations and weather-regime induced dependency shifts among wind variables at geographically distributed sites. In this paper we introduce approaches that address both of these challenges. We describe a new regime-aware approach to STWF that use auto-regressive hidden Markov models (AR-HMM), a subclass of conditional linear Gaussian (CLG) models. Although AR-HMMs are a natural representation for weather regimes, as with CLG models in general, exact inference is NP-hard when observations are missing (Lerner and Parr, 2001). Because of this high cost, we introduce a simple approximate inference method for AR-HMMs, which we believe has applications to other sequential and temporal problem domains that involve continuous variables. In an empirical evaluation on publicly available wind data from two geographically distinct regions, our approach makes significantly more accurate predictions than baseline models, and uncovers meteorologically relevant regimes.
Subject Area: Probabilistic Models and Methods

M57 Switching State Space Model for Simultaneously Estimating State Transitions and Nonstationary Firing Rates

Ken Takiyama takiyama@mns.k.u-tokyo.ac.jp
Masato Okada okada@k.u-tokyo.ac.jp
The University of Tokyo

We propose an algorithm for simultaneously estimating state transitions among neural states, the number of neural states, and nonstationary firing rates using a switching state space model (SSSM). This model enables us to detect state transitions based not only on the discontinuous changes of mean firing rates but also on discontinuous changes in temporal profiles of firing rates e.g. temporal correlation. We derive a variational Bayes algorithm for a non-Gaussian SSSM whose non-Gaussian property is caused by binary spike events. Synthetic data analysis reveals the high performance of our algorithm in estimating state transitions the number of neural states and nonstationary firing rates compared to previous methods. We also analyze neural data recorded from the medial temporal area. The statistically detected neural states probably coincide with transient and sustained states which have been detected heuristically. Estimated parameters suggest that our algorithm detects the state transition based on discontinuous change in the temporal correlation of firing rates which transitions previous methods cannot detect. This result suggests the advantage of our algorithm in real-data analysis.
Subject Area: Probabilistic Models and Methods

M58 Computing Marginal Distributions over Continuous Markov Networks for Statistical Relational Learning

Matthias Broecheler matthias@cs.umd.edu
University of Maryland
CP Lise Getoor getoor@cs.umd.edu

Continuous Markov random fields are a general formalism to model joint probability distributions over events with continuous outcomes. We prove that marginal computation for constrained continuous MRFs is #P-hard in general and present a polynomial-time approximation scheme under mild assumptions on the structure of the random field. Moreover, we introduce a sampling algorithm to compute marginal distributions and develop novel techniques to increase its efficiency. Continuous MRFs are a general purpose probabilistic modeling tool and we demonstrate how they can be applied to statistical relational learning. On the problem of collective classification, we evaluate our algorithm and show that the standard deviation of marginals serves as a useful measure of confidence.
Subject Area: Probabilistic Models and Methods

M59 Heavy-Tailed Process Priors for Selective Shrinkage

Fabian L Wauthier flw@cs.berkeley.edu
Michael I Jordan jordan@cs.berkeley.edu
University of California Berkeley

Heavy-tailed distributions are often used to enhance the robustness of regression and classification methods to outliers in output space. Often however we are confronted with "outliers" in input space which are isolated observations in sparsely populated regions. We show that heavy-tailed process priors (which we construct from Gaussian processes via a copula) can be used to improve robustness of regression and classification estimators to such outliers by selectively shrinking them more strongly in sparse regions than in dense regions. We carry out a theoretical analysis to show that selective shrinkage occurs provided the marginals of the heavy-tailed process have sufficiently heavy tails. The analysis is complemented by experiments on biological data which indicate significant improvements of estimates in sparse regions while producing competitive results in dense regions.
Subject Area: Probabilistic Models and Methods

M60 MAP Estimation for Graphical Models by Likelihood Maximization

Akshat Kumar akshat@cs.umass.edu
Shlomo Zilberstein shlomo@cs.umass.edu
UMass Amherst

Computing a maximum a posteriori (MAP) assignment in graphical models is a crucial inference problem for many practical applications. Several provably convergent approaches have been successfully developed using linear programming (LP) relaxation of the MAP problem. We present an alternative approach, which transforms the MAP problem into that of inference in a finite mixture of simple Bayes nets. We then derive the Expectation Maximization (EM) algorithm for this mixture that also monotonically increases a lower bound on the MAP assignment until convergence. The update equations for the EM algorithm are remarkably simple, both conceptually and computationally, and can be implemented using a graph-based message passing paradigm similar to max-product computation. We experiment on the real-world protein design dataset and show that EM's convergence rate is significantly higher than the previous LP relaxation based approach MPLP. EM achieves a solution quality within 95% of optimal for most instances and is often an order-of-magnitude faster than MPLP.
Subject Area: Probabilistic Models and Methods

M61 Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models

Han Liu hanliu@cs.cmu.edu
Kathryn Roeder roeder@stat.cmu.edu
Larry Wasserman larry@stat.cmu.edu
Carnegie Mellon University

A challenging problem in estimating high-dimensional graphical models is to choose the regularization parameter in a data-dependent way. The standard techniques include K-fold cross-validation (K-CV), Akaike information criterion (AIC), and Bayesian information criterion (BIC). Though these methods work well for low-dimensional problems, they are not suitable in high dimensional settings. In this paper, we present StARS: a new stabilitybased method for choosing the regularization parameter in high dimensional inference for undirected graphs. The method has a clear interpretation: we use the least amount of regularization that simultaneously makes a graph sparse and replicable under random sampling. This interpretation requires essentially no conditions. Under mild conditions, we show that StARS is partially sparsistent in terms of graph estimation: i.e. with high probability, all the true edges will be included in the selected model even when the graph size asymptotically increases with the sample size. Empirically, the performance of StARS is compared with the state-of-the-art model selection procedures, including K-CV, AIC, and BIC, on both synthetic data and a real microarray dataset. StARS outperforms all competing procedures.
Subject Area: Probabilistic Models and Methods

M62 Fast Large-scale Mixture Modeling with Component-specific Data Partitions

Bo Thiesson thiesson@microsoft.com
Microsoft Research
Chong Wang chongw@cs.princeton.edu
Princeton University

Remarkably easy implementation and guaranteed convergence has made the EM algorithm one of the most used algorithms for mixture modeling. On the downside, the E-step is linear in both the sample size and the number of mixture components, making it impractical for large-scale data. Based on the variational EM framework, we propose a fast alternative that uses component-specific data partitions to obtain a sub-linear E-step in sample size, while the algorithm still maintains provable convergence. Our approach builds on previous work, but is significantly faster and scales much better in the number of mixture components. We demonstrate this speedup by experiments on large-scale synthetic and real data.
Subject Area: Unsupervised & Semi-supervised Learning

M63 Deterministic Single-Pass Algorithm for LDA

Issei Sato isseis@gmail.com
Hiroshi Nakagawa n3@dl.itc.u-tokyo.ac.jp
Tokyo University
Kenichi Kurihara kenichi.kurihara@gmail.com
Google

We develop a deterministic single-pass algorithm for latent Dirichlet allocation (LDA) in order to process received documents one at a time and then discard them in an excess text stream. Our algorithm does not need to store old statistics for all data. The proposed algorithm is much faster than a batch algorithm and is comparable to the batch algorithm in terms of perplexity in experiments.
Subject Area: Unsupervised & Semi-supervised Learning

M64 Efficient Relational Learning with Hidden Variable Detection

Ni Lao nlao@cs.cmu.edu
Jun Zhu junzhu@cs.cmu.edu
Liu Xinwang liuliu@cs.cmu.edu
Yandong Liu yandongl@cs.cmu.edu
William W Cohen wcohen@cs.cmu.edu
Carnegie Mellon University

Markov networks (MNs) can incorporate arbitrarily complex features in modeling relational data. However, this flexibility comes at a sharp price of training an exponentially complex model. To address this challenge, we propose a novel relational learning approach, which consists of a restricted class of relational MNs (RMNs) called relation tree-based RMN (treeRMN), and an efficient Hidden Variable Detection algorithm called Contrastive Variable Induction (CVI). On one hand, the restricted treeRMN only considers simple (e.g., unary and pairwise) features in relational data and thus achieves computational efficiency; and on the other hand, the CVI algorithm efficiently detects hidden variables which can capture long range dependencies. Therefore, the resultant approach is highly efficient yet does not sacrifice its expressive power. Empirical results on four real datasets show that the proposed relational learning method can achieve similar prediction quality as the state-of-the-art approaches, but is significantly more efficient in training; and the induced hidden variables are semantically meaningful and crucial to improve the training speed and prediction qualities of treeRMNs.
Subject Area: Probabilistic Models and Methods

M65 Fast detection of Multiple Change-points Shared by Many Signals Using Group LARS

Jean-Philippe Vert Jean-Philippe.Vert@Mines-ParisTech.fr
Kevin Bleakley Kevin.Bleakley@mines-paristech.fr
Mines ParisTech

We present a fast algorithm for the detection of multiple change-points when each is frequently shared by members of a set of co-occurring one-dimensional signals. We give conditions on consistency of the method when the number of signals increases, and provide empirical evidence to support the consistency results.
Subject Area: Probabilistic Models and Methods

M66 A VLSI Implementation of the Adaptive Exponential Integrate-and-Fire Neuron Model

Sebastian C Millner sebastian.millner@kip.uni-heidelberg.de
Andreas Grubl agruebl@kip.uni-heidelberg.de
Karlheinz Meier meierk@kip.uni-heidelberg.de
Johannes Schemmel schemmel@kip.uni-heidelberg.de
Marc-Olivier Schwartz marcolivier.schwartz@kip.uni-heidelberg.de
Kirchhoff-Institut for Physics

We describe an accelerated hardware neuron being capable of emulating the adap-tive exponential integrate-and-fire neuron model. Firing patterns of the membrane stimulated by a step current are analyzed in transistor level simulation and in silicon on a prototype chip. The neuron is destined to be the hardware neuron of a highly integrated wafer-scale system reaching out for new computational paradigms and opening new experimentation possibilities. As the neuron is dedicated as a universal device for neuroscientific experiments, the focus lays on parameterizability and reproduction of the analytical model.
Subject Area: Hardware

M67 Structural Epitome: a Way to Summarize One's Visual Experience

Nebojsa Jojic jojic@microsoft.com
Microsoft Research
Alessandro Perina alessandro.perina@gmail.com
Vittorio Murino vittorio.murino@univr.it
University of Verona / Italian Institute of Technology

In order to study the properties of total visual input in humans, a single subject wore a camera for two weeks capturing, on average, an image every 20 seconds (www.research.microsoft.com/ jojic/aihs). The resulting new dataset contains a mix of indoor and outdoor scenes as well as numerous foreground objects. Our first analysis goal is to create a visual summary of the subject's two weeks of life using unsupervised algorithms that would automatically discover recurrent scenes, familiar faces or common actions. Direct application of existing algorithms, such as panoramic stitching (e.g. Photosynth) or appearance-based clustering models (e.g. the epitome), is impractical due to either the large dataset size or the dramatic variation in the lighting conditions. As a remedy to these problems we introduce a novel image representation the "stel epitome" and an associated efficient learning algorithm. In our model each image or image patch is characterized by a hidden mapping T which as in previous epitome models defines a mapping between the image-coordinates and the coordinates in the large "all-I-have-seen" epitome matrix. The limited epitome real-estate forces the mappings of different images to overlap with this overlap indicating image similarity. However in our model the image similarity does not depend on direct pixel-to-pixel intensity/color/feature comparisons as in previous epitome models but on spatial configuration of scene or object parts as the model is based on the palette-invariant stel models. As a result stel epitomes capture structure that is invariant to non-structural changes such as illumination that tend to uniformly affect pixels belonging to a single scene or object part.
Subject Area: Vision

M68 A Unified Model of Short-range and Long-range Motion Perception

Shuang Wu shuangw@stat.ucla.edu
Xuming He hexm@stat.ucla.edu
Hongjing Lu hongjing@ucla.edu
Alan L Yuille yuille@stat.ucla.edu
UCLA

The human vision system is able to effortlessly perceive both short-range and long-range motion patterns in complex dynamic scenes. Previous work has assumed that two different mechanisms are involved in processing these two types of motion. In this paper, we propose a hierarchical model as a unified framework for modeling both short-range and long-range motion perception. Our model consists of two key components: a data likelihood that proposes multiple motion hypotheses using nonlinear matching, and a hierarchical prior that imposes slowness and spatial smoothness constraints on the motion field at multiple scales. We tested our model on two types of stimuli, random dot kinematograms and multiple-aperture stimuli, both commonly used in human vision research. We demonstrate that the hierarchical model adequately accounts for human performance in psychophysical experiments.

M69 Object Bank: A High-Level Image Representation for Scene Classification & Semantic Feature Sparsification

Li-Jia Li lijiali@cs.stanford.edu
Hao Su suhaochina@gmail.com
Eric Xing epxing@cs.cmu.edu
Stanford University
Li Fei-Fei feifeili@cs.stanford.edu
Carnegie Mellon University

Robust low-level image features have been proven to be effective representations for a variety of visual recognition tasks such as object recognition and scene classification; but pixels, or even local image patches, carry little semantic meanings. For high level visual tasks, such low-level image representations are potentially not enough. In this paper, we propose a high-level image representation, called the Object Bank, where an image is represented as a scale invariant response map of a large number of pre-trained generic object detectors, blind to the testing dataset or visual task. Leveraging on the Object Bank representation, superior performances on high level visual recognition tasks can be achieved with simple off-the-shelf classifiers such as logistic regression and linear SVM. Sparsity algorithms make our representation more efficient and scalable for large scene datasets, and reveal semantically meaningful feature patterns.
Subject Area: Vision

M70 Size Matters: Metric Visual Search Constraints from Monocular Metadata

Mario J Fritz mfritz@eecs.berkeley.edu
UC Berkeley & ICSI
Kate Saenko saenko@mit.edu
MIT
Trevor Darrell trevor@eecs.berkeley.edu
UC Berkeley

Metric constraints are known to be highly discriminative for many objects, but if training is limited to data captured from a particular 3-D sensor the quantity of training data may be severly limited. In this paper, we show how a crucial aspect of 3-D information–object and feature absolute size–can be added to models learned from commonly available online imagery, without use of any 3-D sensing or re- construction at training time. Such models can be utilized at test time together with explicit 3-D sensing to perform robust search. Our model uses a “2.1D” local feature, which combines traditional appearance gradient statistics with an estimate of average absolute depth within the local window. We show how category size information can be obtained from online images by exploiting relatively unbiquitous metadata fields specifying camera intrinstics. We develop an efficient metric branch-and-bound algorithm for our search task, imposing 3-D size constraints as part of an optimal search for a set of features which indicate the presence of a category. Experiments on test scenes captured with a traditional stereo rig are shown, exploiting training data from from purely monocular sources with associated EXIF metadata.
Subject Area: Vision

M71 Occlusion Detection and Motion Estimation with Convex Optimization

Alper Ayvaci ayvaci@cs.ucla.edu
Michalis Raptis mraptis@cs.ucla.edu
Stefano Soatto soatto@cs.ucla.edu
UCLA

We tackle the problem of simultaneously detecting occlusions and estimating optical flow. We show that, under standard assumptions of Lambertian reflection and static illumination, the task can be posed as a convex minimization problem. Therefore, the solution, computed using efficient algorithms, is guaranteed to be globally optimal, for any number of independently moving objects, and any number of occlusion layers. We test the proposed algorithm on benchmark datasets, expanded to enable evaluation of occlusion detection performance.
Subject Area: Vision

M72 Functional Form of Motion Priors in Human Motion Perception

Hongjing Lu hongjing@ucla.edu
Tungyou Lin tungyoul@math.ucla.edu
Alan L Lee alanlee@ucla.edu
Luminita Vese lvese@math.ucla.edu
Alan L Yuille yuille@stat.ucla.edu
UCLA

It has been speculated that the human motion system combines noisy measurements with prior expectations in an optimal, or rational, manner. The basic goal of our work is to discover experimentally which prior distribution is used. More specifically, we seek to infer the functional form of the motion prior from the performance of human subjects on motion estimation tasks. We restricted ourselves to priors which combine three terms for motion slowness, first-order smoothness, and second-order smoothness. We focused on two functional forms for prior distributions: L2-norm and L1-norm regularization corresponding to the Gaussian and Laplace distributions respectively. In our first experimental session we estimate the weights of the three terms for each functional form to maximize the fit to human performance. We then measured human performance for motion tasks and found that we obtained better fit for the L1-norm (Laplace) than for the L2-norm (Gaussian). We note that the L1-norm is also a better fit to the statistics of motion in natural environments. In addition, we found large weights for the second-order smoothness term, indicating the importance of high-order smoothness compared to slowness and lower-order smoothness. To validate our results further, we used the best fit models using the L1-norm to predict human performance in a second session with different experimental setups. Our results showed excellent agreement between human performance and model prediction – ranging from 3% to 8% for five human subjects over ten experimental conditions – and give further support that the human visual system uses an L1-norm (Laplace) prior.
Subject Area: Vision

M73 Layered image motion with explicit occlusions, temporal consistency, and depth ordering

Deqing Sun dqsun@cs.brown.edu
Erik Sudderth sudderth@cs.brown.edu
Michael Black black@cs.brown.edu
Brown University

Layered models are a powerful way of describing natural scenes containing smooth surfaces that may overlap and occlude each other. For image motion estimation, such models have a long history but have not achieved the wide use or accuracy of non-layered methods. We present a new probabilistic model of optical flow in layers that addresses many of the shortcomings of previous approaches. In particular, we define a probabilistic graphical model that explicitly captures: 1) occlusions and disocclusions; 2) depth ordering of the layers; 3) temporal consistency of the layer segmentation. Additionally the optical flow in each layer is modeled by a combination of a parametric model and a smooth deviation based on an MRF with a robust spatial prior; the resulting model allows roughness in layers. Finally, a key contribution is the formulation of the layers using an image-dependent hidden field prior based on recent models for static scene segmentation. The method achieves state-of-the-art results on the Middlebury benchmark and produces meaningful scene segmentations as well as detected occlusion regions.

Subject Area: Vision

Spotlight presentation, Monday, 6:30.

M74 Sparse Instrumental Variables (SPIV) for Genome-Wide Studies

Felix V Agakov felixa@aivalley.com
Paul McKeigue paul.mckeigue@ed.ac.uk
Amos J Storkey a.storkey@ed.ac.uk
University of Edinburgh
Jon Krohn jon.krohn@magd.ox.ac.uk
Oxford

This paper describes a probabilistic framework for studying associations between multiple genotypes, biomarkers, and phenotypic traits in the presence of noise and unobserved confounders for large genetic studies. The framework builds on sparse linear methods developed for regression and modified here for inferring causal structures of richer networks with latent variables. The method is motivated by the use of genotypes as “instruments” to infer causal associations between phenotypic biomarkers and outcomes, without making the common restrictive assumptions of instrumental variable methods. The method may be used for an effective screening of potentially interesting genotype phenotype and biomarker-phenotype associations in genome-wide studies, which may have important implications for validating biomarkers as possible proxy endpoints for early stage clinical trials. Where the biomarkers are gene transcripts, the method can be used for fine mapping of quantitative trait loci (QTLs) detected in genetic linkage studies. The method is applied for examining effects of gene transcript levels in the liver on plasma HDL cholesterol levels for a sample of sequenced mice from a heterogeneous stock, with ~ 10^5 genetic instruments and ~ 47×10^3 gene transcripts.

Subject Area: Applications

M75 Predicting Execution Time of Computer Programs Using Sparse Polynomial Regression

Ling Huang ling.huang@intel.com
Byung-Gon Chun byung-gon.chun@intel.com
Petros Maniatis petros.maniatis@intel.com
Mayur Naik Mayur.naik@intel.com Intel
Intel
Jinzhu Jia jzjia@stat.berkeley.edu
Bin Yu binyu@stat.berkeley.edu
UC Berkeley

Predicting the execution time of computer programs is an important but challenging problem in the community of computer systems. Existing methods require experts to perform detailed analysis of program code in order to construct predictors or select important features. We recently developed a new system to automatically extract a large number of features from program execution on sample inputs, on which prediction models can be constructed without expert knowledge. In this paper we study the construction of predictive models for this problem. We propose the SPORE (Sparse POlynomial REgression) methodology to build accurate prediction models of program performance using feature data collected from program execution on sample inputs. Our two SPORE algorithms are able to build relationships between responses (e.g. the execution time of a computer program) and features and select a few from hundreds of the retrieved features to construct an explicitly sparse and non-linear model to predict the response variable. The compact and explicitly polynomial form of the estimated model could reveal important insights into the computer program (e.g. features and their non-linear combinations that dominate the execution time) enabling a better understanding of the program’s behavior. Our evaluation on three widely used computer programs shows that SPORE methods can give accurate prediction with relative error less than 7% by using a moderate number of training data samples. In addition we compare SPORE algorithms to state-of-the-art sparse regression algorithms and show that SPORE methods motivated by real applications outperform the other methods in terms of both interpretability and prediction accuracy.

Subject Area: Applications

M76 Cross Species Expression Analysis using a Dirichlet Process Mixture Model with Latent Matchings

Hai-Son P Le hple@cs.cmu.edu
Ziv Bar-Joseph zivbj@cs.cmu.edu
Carnegie Mellon University

Recent studies compare gene expression data across species to identify core and species specific genes in biological systems. To perform such comparisons researchers need to match genes across species. This is a challenging task since the correct matches (orthologs) are not known for most genes. Previous work in this area used deterministic matchings or reduced multidimensional expression data to binary representation. Here we develop a new method that can utilize soft matches (given as priors) to infer both, unique and similar expression patterns across

species and a matching for the genes in both species. Our method uses a Dirichlet process mixture model which includes a latent data matching variable. We present learning and inference algorithms based on variational methods for this model. Applying our method to immune response data we show that it can accurately identify common and unique response patterns by improving the matchings between human and mouse genes.
Subject Area: Applications

M77 Adaptive Multi-Task Lasso: with Application to eQTL Detection

Seunghak Lee seunghak@cs.cmu.edu
Jun Zhu junzhu@cs.cmu.edu
Eric Xing epxing@cs.cmu.edu
Carnegie Mellon University

To understand the relationship between genomic variations among population and complex diseases, it is essential to detect eQTLs which are associated with phenotypic effects. However, detecting eQTLs remains a challenge due to complex underlying mechanisms and the very large number of genetic loci involved compared to the number of samples. Thus, to address the problem, it is desirable to take advantage of the structure of the data and prior information about genomic locations such as conservation scores and transcription factor binding sites. In this paper we propose a novel regularized regression approach for detecting eQTLs which takes into account related traits simultaneously while incorporating many regulatory features. We first present a Bayesian network for a multi-task learning problem that includes priors on SNPs making it possible to estimate the significance of each covariate adaptively. Then we find the maximum a posteriori (MAP) estimation of regression coefficients and estimate weights of covariates jointly. This optimization procedure is efficient since it can be achieved by using convex optimization and a coordinate descent procedure iteratively. Experimental results on simulated and real yeast datasets confirm that our model outperforms previous methods for finding eQTLs.
Subject Area: Applications

M78 Reverse Multi-Label Learning

James Petterson james.petterson@nicta.com.au
Tiberio Caetano Tiberio.Caetano@nicta.com.au
NICTA and ANU

Multi-label classification is the task of predicting potentially multiple labels for a given instance. This is common in several applications such as image annotation, document classification and gene function prediction. In this paper we present a formulation for this problem based on reverse prediction: we predict sets of instances given the labels. By viewing the problem from this perspective, the most popular quality measures for assessing the performance of multi-label classification admit relaxations that can be efficiently optimised. We optimise these relaxations with standard algorithms and compare our results with several state-of-the-art methods, showing excellent performance.
Subject Area: Applications

M79 Empirical Risk Minimization with Approximations of Probabilistic Grammars

Shay Cohen scohen@cs.cmu.edu
Noah Smith nasmith@cs.cmu.edu
Carnegie Mellon University

Probabilistic grammars are generative statistical models that are useful for compositional and sequential structures. We present a framework reminiscent of structural risk minimization for empirical risk minimization of the parameters of a fixed probabilistic grammar using the log-loss. We derive sample complexity bounds in this framework that apply both to the supervised setting and the unsupervised setting.
Subject Area: Applications

M80 An Alternative to Low-level-Sychrony-Based Methods for Speech Detection

Paul L Ruvolo paul@mplab.ucsd.edu
Javier R Movellan movellan@mplab.ucsd.edu
UC San Diego

Determining whether someone is talking has applications in many areas such as speech recognition speaker diarization social robotics facial expression recognition and human computer interaction. One popular approach to this problem is audio-visual synchrony detection. A candidate speaker is deemed to be talking if the visual signal around that speaker correlates with the auditory signal. Here we show that with the proper visual features (in this case movements of various facial muscle groups) a very accurate detector of speech can be created that does not use the audio signal at all. Further we show that this person independent visual-only detector can be used to train very accurate audiobased person dependent voice models. The voice model has the advantage of being able to identify when a particular person is speaking even when they are not visible to the camera (e.g. in the case of a mobile robot). Moreover we show that a simple sensory fusion scheme between the auditory and visual models improves performance on the task of talking detection. The work here provides dramatic evidence about the efficacy of two very different approaches to multimodal speech detection on a challenging database.
Subject Area: Speech and Signal Processing

M81 Phone Recognition with the Mean-Covariance Restricted Boltzmann Machine

George E Dahl george.dahl@gmail.com
Marc'Aurelio Ranzato ranzato@cs.toronto.edu
Abdel-rahman Mohamed asamir@cs.toronto.edu
Geoffrey Hinton hinton@cs.toronto.edu
University of Toronto

Straightforward application of Deep Belief Nets (DBNs) to acoustic modeling produces a rich distributed representation of speech data that is useful for recognition and yields impressive results on the speaker-independent TIMIT phone recognition task. However, the first-layer Gaussian-Bernoulli Restricted Boltzmann Machine (GRBM) has an important limitation, shared with mixtures of diagonal-covariance Gaussians: GRBMs treat different components of the acoustic input vector as conditionally independent given the hidden state. The mean-covariance restricted Boltzmann machine (mcRBM), first introduced for modeling natural images, is a much more representationally efficient and powerful way of modeling the covariance structure of speech data. Every configuration of the precision units of the mcRBM specifies a different precision matrix for the conditional distribution over the acoustic space. In this work, we use the mcRBM to learn features of speech data that serve as input into a standard DBN. The mcRBM features combined with DBNs allow us to achieve a phone error rate of 20.5%, which is superior to all published results on speaker-independent TIMIT to date.
Subject Area: Speech and Signal Processing

M82 Beyond Actions: Discriminative Models for Contextual Group Activities

Tian Lan taran.lan1986@gmail.com
Yang Wang ywang12@cs.sfu.ca
Weilong Yang weilongyang@gmail.com
Greg Mori mori@cs.sfu.ca
Simon Fraser University

We propose a discriminative model for recognizing group activities. Our model jointly captures the group activity the individual person actions and the interactions among them. Two new types of contextual information group-person interaction and personperson interaction are explored in a latent variable framework. Different from most of the previous latent structured models which assume a predefined structure for the hidden layer e.g. a tree structure we treat the structure of the hidden layer as a latent variable and implicitly infer it during learning and inference. Our experimental results demonstrate that by inferring this contextual information together with adaptive structures the proposed model can significantly improve activity recognition performance.
Subject Area: Vision

M83 Group Sparse Coding with a Laplacian Scale Mixture Prior

Pierre J Garrigues pierre.garrigues@gmail.com
IQ Engines Inc.
Bruno A Olshausen baolshausen@berkeley.edu
UC Berkeley

We propose a class of sparse coding models that utilizes a Laplacian Scale Mixture (LSM) prior to model dependencies among coefficients. Each coefficient is modeled as a Laplacian distribution with a variable scale parameter, with a Gamma distribution prior over the scale parameter. We show that, due to the conjugacy of the Gamma prior, it is possible to derive efficient inference procedures for both the coefficients and the scale parameter. When the scale parameters of a group of coefficients are combined into a single variable, it is possible to describe the dependencies that occur due to common amplitude fluctuations among coefficients, which have been shown to constitute a large fraction of the redundancy in natural images. We show that, as a consequence of this group sparse coding, the resulting inference of the coefficients follows a divisive normalization rule, and that this may be efficiently implemented a network architecture similar to that which has been proposed to occur in primary visual cortex. We also demonstrate improvements in image coding and compressive sensing recovery using the LSM model.
Subject Area: Vision

M84 Regularized estimation of image statistics by Score Matching

Diederik P Kingma dpkingma@gmail.com
Universiteit Utrecht
Yann Le Cun yann@cs.nyu.edu
New York U

Score Matching is a recently-proposed criterion for training high-dimensional density models for which maximum likelihood training is intractable. It has been applied to learning natural image statistics but has so-far been limited to simple models due to the difficulty of differentiating the loss with respect to the model parameters. We show how this differentiation can be automated with an extended version of the double-backpropagation algorithm. In addition we introduce a regularization term for the Score Matching loss that enables its use for a broader range of problem by suppressing instabilities that occur with finite training sample sizes and quantized input values. Results are reported for image denoising and super-resolution.
Subject Area: Vision

M85 Learning from Candidate Labeling Sets

Jie Luo jluo@idiap.ch
Idiap/EPF Lausanne
Francesco Orabona orabona@dsi.unimi.it
University of Milano

In many real world applications we do not have access to fully-labeled training data, but only to a list of possible labels. This is the case, e.g., when learning visual classifiers from images downloaded from the web, using just their text captions or tags as learning oracles. In general, these problems can be very difficult. However most of the time there exist different implicit sources of information, coming from the relations between instances and labels, which are usually dismissed. In this paper, we propose a semi-supervised framework to model this kind of problems. Each training sample is a bag containing multi-instances, associated with a set of candidate labeling vectors. Each labeling vector encodes the possible labels for the instances in the bag, with only one being fully correct. The use of the labeling vectors provides a principled way not to exclude any information. We propose a large margin discriminative formulation, and an efficient algorithm to solve it. Experiments conducted on artificial datasets and a real-world images and captions dataset show that our approach achieves performance comparable to SVM trained with the ground-truth labels, and outperforms other baselines.

Subject Area: Unsupervised & Semi-supervised Learning
Spotlight presentation, Monday, 6:30.

M86 Co-regularization Based Semi-supervised Domain Adaptation

Hal Daume III hal@cs.utah.edu
Abhishek Kumar abhik@cs.utah.edu
Avishek Saha avishek@cs.utah.edu
University of Utah

This paper presents a co-regularization based approach to semi-supervised domain adaptation. Our proposed approach (EA++) builds on the notion of augmented space (introduced in EASYADAPT (EA) [1]) and harnesses unlabeled data in target domain to further enable the transfer of information from source to target. This semi-supervised approach to domain adaptation is extremely simple to implement and can be applied as a pre-processing step to any supervised learner. Our theoretical analysis (in terms of Rademacher complexity) of EA and EA++ show that the hypothesis class of EA++ has lower complexity (compared to EA) and hence results in tighter generalization bounds. Experimental results on sentiment analysis tasks reinforce our theoretical findings and demonstrate the efficacy of the proposed method when compared to EA as well as a few other baseline approaches.

Subject Area: Unsupervised & Semi-supervised Learning

M87 Batch Bayesian Optimization via Simulation Matching

Javad Azimi azimi@eecs.oregonstate.edu
Alan Fern afern@eecs.oregonstate.edu
Xiaoli Fern xfern@eecs.oregonstate.edu
Oregon State University

Bayesian optimization methods are often used to optimize unknown functions that are costly to evaluate. Typically these methods sequentially select inputs to be evaluated one at a time based on a posterior over the unknown function that is updated after each evaluation. There are a number of effective sequential policies for selecting the individual inputs. In many applications however it is desirable to perform multiple evaluations in parallel which requires selecting batches of multiple inputs to evaluate at once. In this paper we propose a novel approach to batch Bayesian optimization providing a policy for selecting batches of inputs with the goal of optimizing the function as efficiently as possible. The key idea is to exploit the availability of high-quality and efficient sequential policies by using Monte-Carlo simulation to select input batches that closely match their expected behavior. To the best of our knowledge this is the first batch selection policy for Bayesian optimization. Our experimental results on six benchmarks show that the proposed approach significantly outperforms two baselines and can lead to large advantages over a top sequential approach in terms of performance per unit time.

Subject Area: Unsupervised & Semi-supervised Learning.

M88 Active Estimation of F-Measures

Christoph Sawade sawade@cs.uni-potsdam.de
Niels Landwehr landwehr@cs.uni-potsdam.de
Tobias Scheffer scheffer@cs.uni-potsdam.de
University of Potsdam

We address the problem of estimating the F-measure of a given model as accurately as possible on a fixed labeling budget. This problem occurs whenever an estimate cannot be obtained from held-out training data; for instance, when data that have been used to train the model are held back for reasons of privacy or do not reflect the test distribution. In this case, new test instances have to be drawn and labeled at a cost. An active estimation procedure selects instances according to an instrumental sampling distribution. An analysis of the sources of estimation error leads to an optimal sampling distribution that minimizes estimator variance. We explore conditions under which active estimates of F-measures are more accurate than estimates based on instances sampled from the test distribution.

Subject Area: Unsupervised & Semi-supervised Learning

M89 Agnostic Active Learning Without Constraints

Alina Beygelzimer beygel@us.ibm.com
IBM Research
Daniel Hsu djhsu@rci.rutgers.edu
Rutgers University and University of Pennsylvania
John Langford jl@yahoo-inc.com
Yahoo Research
Zhang Tong tzhang@stat.rutgers.edu
Rutgers University

We present and analyze an agnostic active learning algorithm that works without keeping a version space. This is unlike all previous approaches where a restricted set of candidate hypotheses is maintained throughout learning, and only hypotheses from this set are ever returned. By avoiding this version space approach, our algorithm sheds the computational burden and brittleness associated with maintaining version spaces, yet still allows for substantial improvements over supervised learning for classification.
Subject Area: Unsupervised & Semi-supervised Learning

M90 CUR from a Sparse Optimization Viewpoint

Jacob Bien jbien@stanford.edu
Ya Xu yax@stanford.edu
Michael Mahoney mmahoney@theory.stanford.edu
Stanford University

The CUR decomposition provides an approximation of a matrix X that has low reconstruction error and that is sparse in the sense that the resulting approximation lies in the span of only a few columns of X. In this regard, it appears to be similar to many sparse PCA methods. However, CUR takes a randomized algorithmic approach whereas most sparse PCA methods are framed as convex optimization problems. In this paper, we try to understand CUR from a sparse optimization viewpoint. In particular, we show that CUR is implicitly optimizing a sparse regression objective and, furthermore, cannot be directly cast as a sparse PCA method. We observe that the sparsity attained by CUR possesses an interesting structure, which leads us to formulate a sparse PCA method that achieves a CUR-like sparsity.
Subject Area: Unsupervised & Semi-supervised Learning

M91 Towards Property-Based Classification of Clustering Paradigms

Margareta Ackerman mackerma@cs.uwaterloo.ca
Shai Ben-David shai@cs.uwaterloo.ca
David R Loker dloker@uwaterloo.ca
University of Waterloo

Clustering is a basic data mining task with a wide variety of applications. Not surprisingly, there exist many clustering algorithms. However, clustering is an ill defined problem - given a data set, it is not clear what a “correct” clustering for that set is. Indeed, different algorithms may yield dramatically different outputs for the same input sets. Faced with a concrete clustering task, a user needs to choose an appropriate clustering algorithm. Currently, such decisions are often made in a very ad hoc, if not completely random, manner. Given the crucial effect of the choice of a clustering algorithm on the resulting clustering, this state of affairs is truly regrettable. In this paper we address the major research challenge of developing tools for helping users make more informed decisions when they come to pick a clustering tool for their data. This is, of course, a very ambitious endeavor, and in this paper, we make some first steps towards this goal. We propose to address this problem by distilling abstract properties of the input-output behavior of different clustering paradigms. In this paper we demonstrate how abstract intuitive properties of clustering functions can be used to taxonomize a set of popular clustering algorithmic paradigms. On top of addressing deterministic clustering algorithms we also propose similar properties for randomized algorithms and use them to highlight functional differences between different common implementations of k-means clustering. We also study relationships between the properties independent of any particular algorithm. In particular we strengthen Kleinbergs famous impossibility result while providing a simpler proof.
Subject Area: Unsupervised & Semi-supervised Learning

M92 Energy Disaggregation via Discriminative Sparse Coding

J. Zico Kolter kolter@csail.mit.edu
MIT
Siddharth Batra sidbatra@cs.stanford.edu
Andrew Ng ang@cs.stanford.edu
Stanford University

Energy disaggregation is the task of taking a whole-home energy signal and separating it into its component appliances. Studies have shown that having device-level energy information can cause users to conserve significant amounts of energy, but current electricity meters only report whole-home data. Thus, developing algorithmic methods for disaggregation presents a key technical challenge in the effort to maximize energy conservation. In this paper, we examine a large scale energy disaggregation task, and apply a novel extension of sparse coding to this problem. In particular, we develop a method, based upon structured prediction, for discriminatively training sparse coding algorithms specifically to maximize disaggregation performance. We show that this significantly improves the performance of sparse coding algorithms on the energy task and illustrate how these disaggregation results can provide useful information about energy usage.
Subject Area: Unsupervised & Semi-supervised Learning

TUESDAY CONFERENCE

TUESDAY, DECEMBER 7TH

8:30–9:40AM - ORAL SESSION 1
Session Chair: Bill Triggs

- ***INVITED TALK:* *How Does the Brain Compute and Compare Values at the Time of Decision-Making?***
  Antonio Rangel, Caltech

- ***Over-complete representations on recurrent neural networks can support persistent percepts***
  Shaul Druckmann and Dmitri B Chklovskii, JFRC

9:40–10:00AM - SPOTLIGHTS SESSION 2
Session Chair: Bill Triggs

- ***Improving Human Judgments by Decontaminating Sequential Dependencies,***
  Michael Mozer, Harold Pashler, Matthew Wilder, Robert Lindsey and Matt Jones, University of Colorado at Boulder; Michael Jones, Indiana University

- ***Learning to Localise Sounds with Spiking Neural Networks***
  Dan F Goodman, Ecole Normale Superieure, and Romain Brette, ENS

- ***SpikeAnts, a spiking neuron network modelling the emergence of organization in a complex system***
  Sylvain Chevallier, INRIA Saclay, Helene Paugam-Moisy, Univ. Lyon 2, and Miche Le Sebag, Laboratoire de Recherche en Informatique CNRS

- ***Attractor Dynamics with Synaptic Depression***
  C. C. Alan Fung, K. Y. Michael Wong, The Hong Kong University of Science and Technology, He Wang, Tsinghua University, and Si Wu, Institute of Neuroscience Chinese Academy of Sciences

10:00–10:20AM ORAL SESSION 2
Session Chair: Wee Sun Lee

- ***A Rational Decision Making Framework for Inhibitory Control***
  Pradeep Shenoy, UCSD, Rajesh Rao, U. Washington, and Angela J Yu, UCSD

10:50–11:10AM ORAL SESSION 3
Session Chair: Ben Taskar

- ***A Theory of Multiclass Boosting***
  Indraneel Mukherjee and Robert E Schapire, Princeton University

11:10–11:30AM SPOTLIGHTS SESSION 3
Session Chair: Ben Taskar

- ***Copula Processes***
  Andrew G Wilson, University of Cambridge, and Zoubin Ghahramani, Cambridge

- ***A Biologically Plausible Network for the Computation of Orientation Dominance***
  Kritika Muralidharan, SVCL, and Nuno Vasconcelos, UC San Diego

- ***Learning Convolutional Feature Hierarchies for Visual Recognition***
  Koray Kavukcuoglu, Pierre Sermanet, Y-Lan Boureau, Karol Gregor, Michael Mathieu, and Yann Le Cun, New York University

- ***LSTD with Random Projections***
  Mohammad Ghavamzadeh, Alessandro Lazaric, Odalric Maillard and Remi Munos, INRIA Lille

11:30–11:50AM ORAL SESSION 4
Session Chair: Ofer Dekel

- ***Online Learning: Random Averages, Combinatorial Parameters, and Learnability***
  Alexander Rakhlin, University of Pennsylvania, Karthik Sridharan, Toyota Technological Institute at Chicago, and Ambuj Tewari, UT Austin

11:50AM–12:10PM SPOTLIGHTS SESSION 4
Session Chair: Ofer Dekel

- ***Kernel Descriptors for Visual Recognition***
  Liefeng Bo, University of Washington, Xiaofeng Ren, Intel, and Dieter Fox

- ***Minimum Average Cost Clustering***
  Kiyohito Nagano, University of Tokyo, Yoshinobu Kawahara, Osaka University, and Satoru Iwata, Kyoto University

- ***Identifying Dendritic Processing***
  Aurel A. Lazar and Yevgeniy B Slutskiy, Columbia University

- ***Probabilistic Inference and Differential Privacy***
  Oliver Williams and Frank McSherry, Microsoft Research

2:00–3:10PM ORAL SESSION 5
Session Chair: Iain Murray

- ***INVITED TALK:* *Machine Learning with Human Intelligence:* Principled Corner Cutting (PC2)**
  Xiao Li Meng

- ***Fast global convergence rates of gradient methods for high-dimensional statistical recovery***
  Alekh Agarwal, Sahand Negahban and Artin Wainwright, UC Berkeley

3:10–3:30PM SPOTLIGHTS SESSION 5
Session Chair: Iain Murray

- ***Trading off Mistakes and Don't-Know Predictions***
  Amin Sayedi and Avrim Blum, Carnegie Mellon University; Morteza Zadimoghaddam, MIT,

- ***Exact learning curves for Gaussian process regression on large random graphs***
  Matthew J Urry and Peter Sollich, Kings College London

- ***Synergies in learning words and their referents***
  Mark Johnson, Katherine Demuth, Macquarie University, Michael Frank, MIT, and Bevan K Jones, University of Edinburgh

- ***Learning concept graphs from text with stick-breaking priors***
  America Chambers, Padhraic Smyth and Mark Steyvers, University of California-Irvine

3:30–3:50PM ORAL SESSION 6
Session Chair: Matthias Seeger

- ***Structured sparsity-inducing norms through submodular functions***
  Francis Bach, Ecole Normale Superieure

4:20–5:00PM ORAL SESSION 7
Session Chair: Raquel Urtasun

- ***Semi-Supervised Learning with Adversarially Missing Label Information***
  Umar Syed and Ben Taskar, University of Pennsylvania

- ***MAP estimation in Binary MRFs via Bipartite Multi-cuts***
  Sashank Jakkam Reddi, Sunita Sarawagi and Sundar Vishwanathan, IIT Bombay

5:00–5:20PM SPOTLIGHTS SESSION 6
Session Chair: Raquel Urtasun

- ***Efficient Minimization of Decomposable Submodular Functions***
  Peter G Stobbe and Andreas Krause, Caltech

- ***Worst-case bounds on the quality of max-product fixed-points***
  Meritxell Vinyals, Jes´us Cerquides and Juan Antonio Rodrıguez-Aguilar, IIIA-CSIC; Alessandro Farinelli, University of Verona

- ***Learning To Count Objects in Images***
  Victor Lempitsky and Andrew Zisserman, Oxford

- ***Learning invariant features using the Transformed Indian Buffet Process***
  Joseph L Austerweil and Tom Griffiths, UC Berkeley

5:20–5:40PM ORAL SESSION 8
Session Chair: Rich Carauna

- ***A Dirty Model for Multi-task Learning***
  Ali Jalali, Pradeep Ravikumar, Sujay Sanghavi and Chao Ruan, University of Texas at Austin

5:40–6:00PM SPOTLIGHTS SESSION 7
Session Chair: Rich Carauna

- ***Variational Inference over Combinatorial Spaces***
  Alexandre Bouchard-Cote and Michael I Jordan, UC Berkeley

- ***Multiple Kernel Learning and the SMO Algorithm***
  S.V.N. Vishwanathan, Zhaonan sun, Nawanol T Ampornpunt, Purdue University, and Manik Varma, Microsoft Research

- ***Inductive Regularized Learning of Kernel Functions***
  Prateek Jain, Microsoft Research India Lab, Brian Kulis, UC Berkeley, and Inderjit Dhillon, University of Texas

- ***Probabilistic Deterministic Infinite Automata***
  David Pfau, Nicholas Bartlett and Frank Wood, Columbia University

7:00–11:59PM - POSTER SESSION

T1 **Inferring Stimulus Selectivity from the Spatial Structure of NeuralNetwork Dynamics**
Kanaka Rajan, Princeton University, L F Abbott,Columbia University, and Haim Sompolinsky, Hebrew University and Harvard University

T2 **Individualized ROI Optimization via Maximization of Group-wise Consistency of Structural and Functional Profiles**, Kaiming Li,NWPU, Lei Guo, Carlos Faraco, Dajiang Zhu, Fan Deng, University of Georgia, TuoZhang, Xi Jiang, Degang Zhang, Hanbo Chen, Northwestern Polytechnical Unv, XintaoHu, Steve Miller, and Tianming Liu

T3 **Categories and Functional Units**: An Infinite Hierarchical Model for Brain Activations, Danial Lashkari, Ramesh Sridharan and Polina Golland, MIT

T4 **Sodium entry efficiency during action potentials: A novelsingle-parameter family of Hodgkin-Huxley models**, AnandSingh, Renaud Jolivet , and Bruno Weber, University of Zurich; Pierre J Magistretti, EPFL

T5 **Identifying Dendritic Processing**, Aurel A. Lazar and Yevgeniy B Slutskiy, Columbia University

T6 **Attractor Dynamics with Synaptic Depression**, C. C. Alan Fung,K. Y. Michael Wong, The Hong Kong University of Science and Technology, He Wang, Tsinghua University, and Si Wu, Institute of Neuroscience Chinese Academy of Sciences

T7 **A rational decision making framework for inhibitory control**, Pradeep Shenoy, UCSD, Rajesh Rao, U. Washington, and Angela JYu, UCSD

T8 **The Neural Costs of Optimal Control**, Samuel Gershman and Robert Wilson, Princeton University

T9 **Over-complete representations on recurrent neural networks can support persistent percepts**, Shaul Druckmann and Dmitri B Chklovskii, JFRC

T10 **Rescaling, thinning or complementing? On goodness-of-fit procedures for point process models and Generalized Linear Models**, Felipe Gerhard and Wulfram Gerstner, Brain Mind Institute

T11 **Divisive Normalization: Justification and Effectiveness as Efficient Coding Transform**, Siwei Lyu, University at Albany SUNY

T12 **Evaluating neuronal codes for inference using Fisher information**, Haefner M Ralf, Max Planck Institute for Biological Cybernetics; Matthias Bethge

T13 **Learning the context of a category**, Dan Navarro, University of Adelaide

T14 **Inference and communication in the game of Password**, Yang Xu and Charles Kemp, Carnegie Mellon University

T15 **Learning Bounds for Importance Weighting**, Corinna Cortes,Google Inc, Yishay Mansour, Tel-Aviv, and Mehryar Mohri

T16 **Smoothness, Low Noise and Fast Rates**, Nathan Srebro, TTI, Karthik Sridharan, Toyota Technological Institute at Chicago, and Ambuj Tewari, UT Austin

T17 **Online Learning: Random Averages, Combinatorial Parameters,and Learn ability**, Alexander Rakhlin, University of Pennsylvania, Karthik Sridharan, Toyota Technological Institute at Chicago, and Ambuj Tewari, UT Austin

T18 **An Analysis on Negative Curvature Induced by Singularity in Multilayerneural-Network Learning**, Eiji Mizutani, Taiwan Univ. Science and Tech; Stuart Dreyfus, UC Berkeley

T19 **Trading off Mistakes and Don’t-Know Predictions**, Amin Sayedi and Avrim Blum, Carnegie Mellon University; Morteza Zadimoghaddam, MIT

T20 **Fast Global Convergence Rates of Gradient Methods for High-Dimensional Statistical Recovery**, Alekh Agarwal, Sahand Negahban and Martin Wainwright, UC Berkeley

T21 **A Dirty Model for Multi-task Learning**, Ali Jalali, Pradeep Ravikumar, Sujay Sanghavi and Chao Ruan, University of Texas at Austin

T22 **Variable Margin Losses for Classifier Design**, Hamed Masnadi-Shirazi and Nuno Vasconcelos, UC San Diego

T23 **Robust PCA via Outlier Pursuit**, Huan Xu, Constantine Caramanis and Sujay Sanghavi, University of Texas

T24 **Network Flow Algorithms for Structured Sparsity**, Julien Mairal, Rodolphe Jenatton, Guillaume R Obozinski, INRIA, and Francis Bach, Ecole Normale Superieure

T25 **Efficient Minimization of Decomposable Submodular Functions**, Peter G Stobbe and Andreas Krause, Caltech

T26 **Minimum Average Cost Clustering**, Kiyohito Nagano, University of Tokyo, Yoshinobu Kawahara, Osaka University, and Satoru Iwata, Kyoto University

T27 **Practical Large-Scale Optimization for Max-norm Regularization**, Jason Lee, Toyota Technological Institute at Chicago, Ben Recht, UW Madison, Ruslan Salakhutdinov, MIT, Nathan Srebro, TTI, and Joel Tropp, CalTech

T28 **Learning Invariant Features using the Transformed Indian Buffet Process**, Joseph L Austerweil and Tom Griffiths,University of California-Berkeley

T29 **Synergies in Learning Words and their Referents**, Mark Johnson,Katherine Demuth, Macquarie University, Michael Frank, MIT, and Bevan K Jones, University of Edinburgh

T30 **Permutation Complexity Bound on Out-Sample Error**, Malik Magdon-Ismail, RPI

T31 **Optimal Learning rates for Kernel Conjugate Gradient Regression**, Gilles Blanchard, Weierstrass Institute Berlin / University of Potsdam, and Nicole Kramer, Weierstrass Institute Berlin

T32 **Nonparametric Density Estimation for Stochastic Optimization with an Observable State Variable**, Lauren Hannah, Warren B Powell and David Blei, Princeton University

T33 **Towards Holistic Scene Understanding: Feedback Enabled Cascaded Classification Models, Congcong Li,** Adarsh P Kowdle, Ashutosh Saxena and Tsuhan Chen, Cornell University

T34 **Learning To Count Objects in Images**, Victor Lempitsky and Andrew Zisserman, University of Oxford

T35 **A biologically Plausible Network for the Computation of Orientation Dominance,** Kritika Muralidharan and Nuno Vasconcelos, UC San Diego

T36 **Pose-Sensitive Embedding by Nonlinear NCA Regression**,Graham Taylor, Rob Fergus, George Williams, Ian Spiro and Christoph Bregler, New York University

T36 **A Biologically Plausible Network for the Computation of Orientation Dominance**, Kritika Muralidharan, SVCL; Nuno Vasconcelos, UC San Diego

T37 **Implicitly Constrained Gaussian Process Regression for Monocular Non-Rigid Pose Estimation**, Mathieu Salzmann and Raquel Urtasun, TTI Chicago

T38 **Large-Scale Matrix Factorization with Missing Data under Additional Constraints**, Kaushik Mitra and Rama Chellappa, University of Maryland College Park; Sameer Sheorey, TTIC, T39 **(RF)² – Random Forest Random Field**, Nadia C Payet and Sinisa Todorovic, Oregon State University

T40 **Kernel Descriptors for Visual Recognition**, Liefeng Bo, University of Washington, Xiaofeng Ren, Intel, and Dieter Fox

T41 **Exploiting weakly-labeled Web images to improve object classification: a domain adaptation approach**, Alessandro Bergamo and Lorenzo Torresani, Dartmouth College

T42 **A Bayesian Framework for Figure-Ground Interpretation**, Vicky Froyen, Rutgers, Jacob Feldman, Rutgers University, and Manish Singh

T43 **Semi-Supervised Learning with Adversarially Missing Label Information**, Umar Syed and Ben Taskar, University of Pennsylvania

T44 **Self-Paced Learning for Latent Variable Models**, M. Pawan Kumar, Stanford University, Benjamin Packer and Daphne Koller, Stanford

T45 **Random Projections for K-means Clustering**, Christos Boutsidis, RPI, Anastasios Zouzias, University of Toronto, and Petros Drineas, Rensselaer Polytechnic Institute

T46 **Discriminative Clustering by Regularized Information Maximization**, Ryan G Gomes, Andreas Krause and Pietro Perona, Caltech

T47 **Transduction with Matrix Completion: Three Birds with One Stone**, Andrew B Goldberg, Xiaojin (Jerry) Zhu, Ben Recht, Junming Xu and Rob Nowak, University of Wisconsin-Madison

T48 **Tiled Convolutional Neural Networks**, Quoc V Le, Jiquan Ngiam, Zhenghao Chen, Daniel Jin hao Chia, Pang Wei Koh and Andrew Ng, Stanford University

T49 **Multi-View Active Learning in the Non-Realizable Case**, Wei Wang and Zhi-Hua Zhou, Nanjing University

T50 **Near-Optimal Bayesian Active Learning with Noisy Observations**, Daniel Golovin, Andreas Krause and Debajyoti Ray, Caltech

T51 **Hashing Hyperplane Queries to Near Points with Applications to Large-Scale Active Learning**, Prateek Jain, Microsoft Research India Lab, Sudheendra Vijayanarasimhan, and Kristen Grauman, University of Texas at Austin

T52 **Unsupervised Kernel Dimension Reduction**, Meihong Wang, Fei Sha, University of Southern California; Michael I Jordan, UC Berkeley

T53 **Large Margin Multi-Task Metric Learning**, Shibin Parameswaran, UCSD, and Kilian Q Weinberger, Washington University in St. Louis

T54 **Deep Coding Network**, Yuanqing Lin, Shenghuo Zhu and Kai Yu NEC Labs; Zhang Tong, Rutgers University

T55 **Inductive Regularized Learning of Kernel Functions**, Prateek Jain, Microsoft Research India Lab, Brian Kulis, University of California Berkeley, and Inderjit Dhillon, University of Texas

T56 **Learning concept graphs from text with stick-breaking priors**, America Chambers, Padhraic Smyth, and Mark Steyvers, University of California-Irvine

T57 **Joint Analysis of Time-Evolving Binary Matrices and Associated Documents**, Eric X Wang, Dehong Liu, Jorge G Silva, David Dunson and Lawrence Carin, Duke University

T58 **Predictive Subspace Learning for Multi-view Data: a Large Margin Approach**, Ning Chen, Tsinghua University, Jun Zhu and Eric Xing, Carnegie Mellon University

T59 **LSTD with Random Projections**, Mohammad Ghavamzadeh, Alessandro Lazaric, Odalric Maillard and Remi Munos, INRIA Lille - Nord Europe

T60 **Constructing Skill Trees for Reinforcement Learning Agents from Demonstration Trajectories**, George D Konidaris, Scott R Kuindersma, Andrew Barto and Roderic A Grupen, UMass Amherst

T61 **Avoiding False Positive in Multi-Instance Learning**, Yanjun Han, Institute of Automation CAS, Qing Tao, and Jue Wang

T62 **A Theory of Multiclass Boosting**, Indraneel Mukherjee and Robert E Schapire, Princeton University

T63 **Joint Cascade Optimization Using A Product Of Boosted Classifiers**, Leonidas Lefakis and Francois Fleuret, Idiap Research Institute

T64 **Multiple Kernel Learning and the SMO Algorithm**, S.V.N.Vishwanathan, Zhaonan sun and Nawanol T Ampornpunt, Purdue University; Manik Varma, Microsoft Research

T65 **Relaxed Clipping: A Global Training Method for Robust Regression and Classification**, Yaoliang Yu, Min Yang, University of Alberta, Linli Xu, University of Science and Technology of China, Martha White and Dale Schuurmans, University of Alberta

T66 **Decomposing Isotonic Regression for Efficiently Solving Large Problems**, Ronny Luss, Saharon Rosset and Moni Shahar, TelAviv University

T67 **Factorized Latent Spaces with Structured Sparsity**, Yangqing Jia, UC Berkeley, Mathieu Salzmann, TTI Chicago, and Trevor Darrell, UC Berkeley

T68 **Evaluation of Rarity of Fingerprints in Forensics**, Chang Su Lee and Sargur N Srihari, Univ. at Buffalo

T69 **Structured sparsity-inducing norms through submodular functions**, Francis Bach, Ecole Normale Superieure

T70 **Learning Convolutional Feature Hierarchies for Visual Recognition**, Koray Kavukcuoglu, Pierre Sermanet, Y-Lan Boureau, KarolGregor, Michael Mathieu and Yann Le Cun, New York University

T71 **Probabilistic Multi-Task Feature Selection**, Yu Zhang, Dit-Yan Yeung and Qian Xu, Hong Kong University of Science and Technology

T72 **Probabilistic Inference and Differential Privacy**, Oliver Williams and Frank McSherry, Microsoft Research

T73 **Inter-time segment information sharing for non-homogeneous dynamic Bayesian networks**, Dirk Husmeier, Frank Dondelinger, Biomathematics & Statistics Scotland (Bioss), and SophieLebre, University of Strasbourg

T74 **Variational Inference over Combinatorial Spaces**, Alexandre Bouchard-Cote, Berkeley, and Michael I Jordan, University of California Berkeley

T75 **Worst-case bounds on the quality of max-product fixed-points**, Merixtell Vinyals, Jes´us Cerquides and Juan Antonio Rodrıguez-Aguilar, IIIA-CSIC; Alessandro Farinelli, University of Verona

T76 **Improving the Asymptotic Performance of Markov Chain Monte-Carlo by Inserting Vortices**, Yi Sun, Faustino Gomez and Juergen Schmidhuber, IDSIA

T77 **Gaussian sampling by local perturbations**, George Papandreou and Alan L Yuille, UCLA

T78 **Approximate Inference by Compilation to Arithmetic Circuits**, Daniel Lowd, University of Oregon; Pedro Domingos, University of Washington.

T79 **MAP estimation in Binary MRFs via Bipartite Multi-cuts**, Sashank Jakkam Reddi, Sunita Sarawagi and Sundar Vishwanathan, IIT Bombay

T80 **Improvements to the Sequence Memoizer**, Jan Gasthaus and Yee Whye Teh, Gatsby Unit UCL

T81 **Probabilistic Deterministic Infinite Automata**, David Pfau, Nicholas Bartlett and Frank Wood, Columbia

T82 **Copula Processes**, Andrew G Wilson and Zoubin Ghahramani, Cambridge

T83 **Learning sparse dynamic linear systems using stable spline kernels and exponential hyperpriors**, Alessandro Chiuso and Gianluigi Pillonetto, University of Padova

T84 **Exact learning curves for Gaussian process regression on large random graphs**, Matthew J Urry and Peter Sollich, Kings College London

T85 **Subgraph Detection Using Eigen vector L1 Norms**, Benjamin AMiller, Nadya T Bliss, MIT Lincoln Laboratory, and Patrick Wolfe, Harvard University

T86 **Gaussian Process Preference Elicitation**, Edwin V Bonilla, Shengbo Guo, NICTA, and Scott Sanner, Nicta

T87 **Implicit Differentiation by Perturbation**, Justin Domke, Rochester Institute of Technology

T88 **A Primal-Dual Message-Passing Algorithm for Approximated Large Scale Structured Prediction**, Tamir Hazan and Raquel Urtasun, TTI Chicago

T89 **Extended Bayesian Information Criteria for Gaussian Graphical Models**, Foygel Rina and Mathias Drton, University of Chicago

T90 **Causal discovery in multiple models from different experiments**, Tom Claassen and Tom Heskes, Radboud University Nijmegen

T91 **Lifted Inference Seen from the Other Side : The Tractable Features**, Abhay Jha, Vibhav G Gogate,Univ. of Washington, Alexandra Meliou and Dan Suciu

T92 **Movement extraction by detecting dynamics switches and repetitions**, Silvia Chiappa, Statistical Laboratory Cambridge University UK, and Jan Peters, MPI for biological cybernetics

T93 **Active Learning Applied to Patient-Adaptive Heartbeat Classification**, Jenna Wiens and John Guttag, MIT

T94 **Static Analysis of Binary Executables Using Structural SVMs**, Nikos Karampatziakis, Cornell Univ.

T95 **Latent Variable Models for Predicting File Dependencies in Large-Scale Software Development**, Diane Hu, Laurensvan der Maaten, Youngmin Cho, Lawrence Saul, and SorinLerner, UC San Diego

T96 **Link Discovery using Graph Feature Tracking**, Emile Richard, Nicolas Baskiotis and Nicolas Vayatis, CMLA/ENS Cachan, TheodorosEvgeniou, INSEAD

T97 **Global seismic monitoring as probabilistic inference**, Nimar S Arora, Stuart Russell, UC Berkeley, Paul Kidwell, Lawrence Livermore National Lab, and Erik Sudderth, Brown University

T98 **Improving Human Judgments by Decontaminating Sequential Dependencies**, Michael Mozer, Harold Pashler, Matthew Wilder, Robert Lindsey, Matt Jones, Univ. of Colorado at Boulder, and Michael Jones, Indiana University

T99 **SpikeAnts, a spiking neuron network modelling the emergence of organization in a complex system**, Sylvain Chevallier, INRIA Saclay, Helene Paugam-Moisy, Univ. Lyon 2, and Miche Le Sebag, Laboratoire de Recherche en Informatique CNRS

T100 **Learning to localise sounds with spiking neural networks**, Dan F Goodman and Romain Brette, Ecole Normale Superieure

11:30–11:59PM DEMONSTRATIONS

D1 **2-D Cursor Movement using EEG**, Chris Laver, University of Guelph

D2 **A framework for Evaluating and Designing "Attention Mechanism" Implementations Based on Tracking of Human Eye Movements**, Dmitry Chichkov

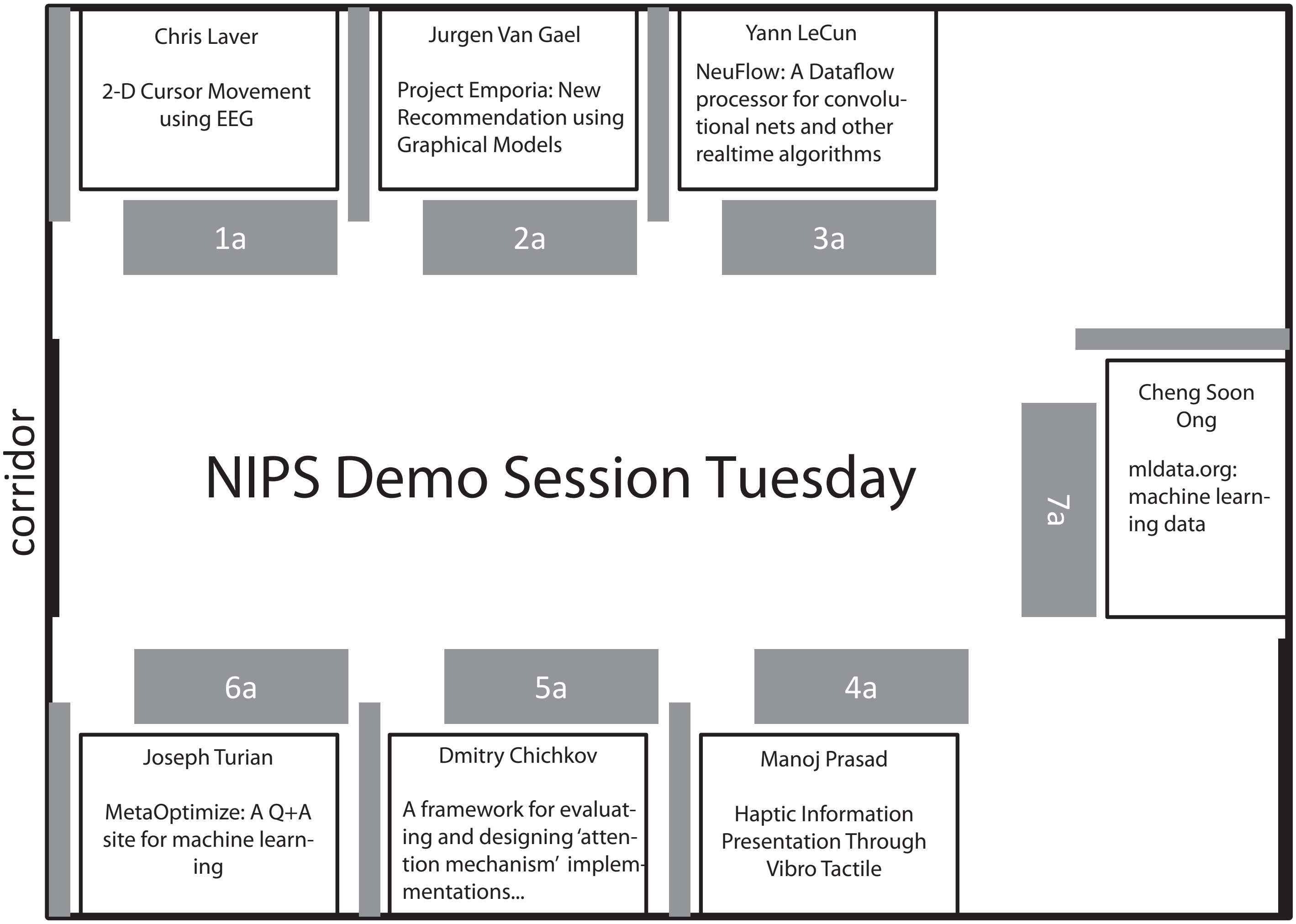
D3 **Haptic Information Presentation Through Vibro Tactile**, Manoj Prasad, Texas A&M University

D4 **MetaOptimize: A Q+A site for machine learning**, Joseph Turian

D5 **mldata.org - machine learning data and benchmark**, Cheng Soon Ong, ETH Zurich

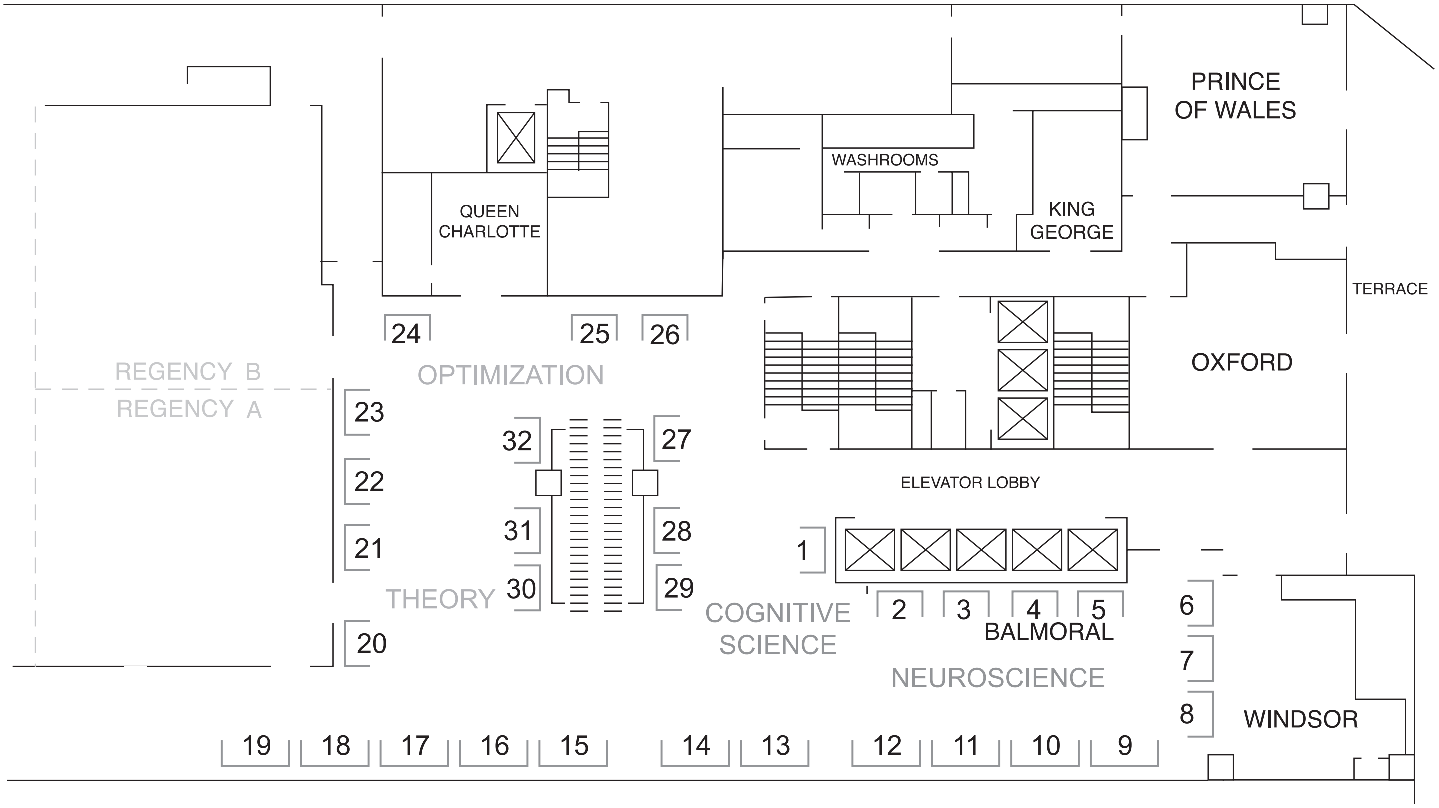
D6 **NeuFlow: a dataflow processor for convolutional nets and other real-time algorithms**, Yann Le Cun, New York University

D7 **Project Emporia: News Recommendation using Graphical Models**, Jurgen Van Gael, Microsoft

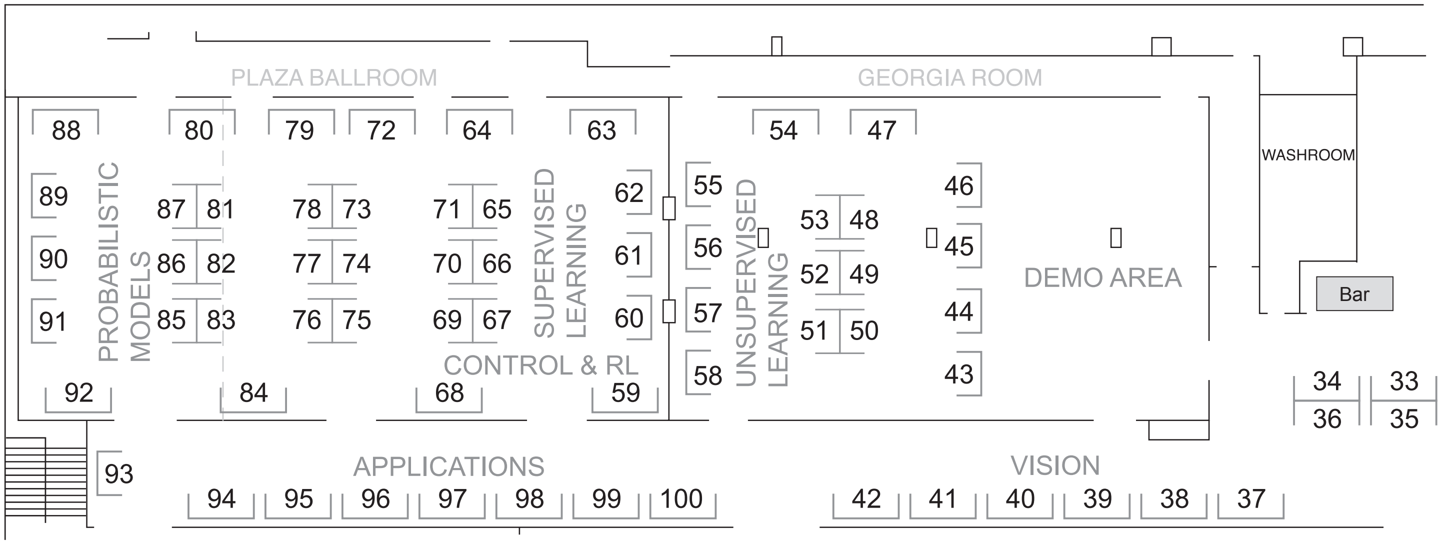

Chris Laver
2-D Cursor Movement using EEG
Jurgen Van Gael
Project Emporia: New Recommendation using Graphical Models
Yann LeCun
NeuFlow: A Dataflow processor for convolutional nets and other realtime algorithms
1a
2a
3a
corridor
NIPS Demo Session Tuesday
7a
Cheng Soon Ong
mldata.org: machine learning data
6a
5a
4a
Joseph Turian
MetaOptimize: A Q+A site for machine learning
Dmitry Chichkov
A framework for evaluating and designing 'attention mechanism' implementations...
Manoj Prasad
Haptic Information Presentation Through Vibro Tactile

NIPS POSTER BOARDS - TUESDAY, DECEMBER 7TH

CONVENTION LEVEL (3RD FLOOR)



PLAZA LEVEL (2ND FLOOR)



ORAL SESSION 1 (8:30–9:40AM):

INVITED TALK: How Does the Brain Compute and Compare Values at the Time of Decision-Making?

Antonio Rangel rangel@hss.caltech.edu
Caltech

Most organisms facing a choice between multiple stimuli will look repeatedly at them, presumably implementing a comparison process between the items‘; values. Little is known about the exact nature of the comparison process in value-based decision-making, or about the role that the visual fixations play in this process. We propose a computational model in which fixations guide the comparison process in simple binary value-based choice and test it using eye-tracking. We present results from an eye-tracking choice experiment showing that the model is able to quantitatively explain complex relationships between fixation patterns and choices, as well as several fixation-driven decision biases. We also present results from several fMRI choice experiments showing that the key processes at work in the model are implemented in the ventromedial and dorsomedial prefrontal cortices.

Over-complete representations on recurrent neural networks can support persistent percepts

Shaul Druckmann druckmanns@janelia.hhmi.org
Dmitri B Chklovskii chklovskiid@janelia.hhmi.org,
JFRC

A striking aspect of cortical neural networks is the divergence of a relatively small number of input channels from the peripheral sensory apparatus into a large number of cortical neurons, an over-complete representation strategy. Cortical neurons are then connected by a sparse network of lateral synapses. Here we propose that such architecture may increase the persistence of the representation of an incoming stimulus, or a percept. We demonstrate that for a family of networks in which the receptive field of each neuron is re-expressed by its outgoing connections, a represented percept can remain constant despite changing activity. We term this choice of connectivity REceptive FIeld REcombination (REFIRE) networks. The sparse REFIRE network may serve as a high-dimensional integrator and a biologically plausible model of the local cortical circuit.
Subject Area: Neuroscience

SPOTLIGHTS SESSION 2 (9:40–10:00AM)

- ***Improving Human Judgments by Decontaminating Sequential Dependencies***
  Michael Mozer, Harold Pashler, Matthew Wilder, Robert Lindsey and Matt Jones, University of Colorado at Boulder; Michael Jones, Indiana Univ.
  Subject Area: Applications
  See abstract, page 72

- ***Learning to Localise Sounds with Spiking Neural Networks,***
  See abstract, page.
  Dan F Goodman and Romain Brette, Ecole Normale Superieure
  Subject Area: Speech and Signal Processing
  See abstract, page 73

- ***SpikeAnts, a spiking neuron network modelling the emergence of organization in a complex system***
  Sylvain Chevallier, INRIA-Saclay, Helene Paugam-Moisy, Univ. Lyon 2, and MicheLe Sebag, Laboratoire de Recherche en Informatique CNRS.
  Subject Area: Applications See abstract, page 73

- ***Attractor Dynamics with Synaptic Depression***
  C. C. Alan Fung, K. Y. Michael Wong, The Hong Kong University of Science and Technology, He Wang, Tsinghua University, and Si Wu, Institute of Neuroscience Chinese Academy of Sciences.
  Subject Area: Neuroscience
  See abstract, page 52

ORAL SESSION 2 (10:00–10:20AM):

A Rational Decision Making Framework for Inhibitory Control

Pradeep Shenoy pshenoy@ucsd.edu
Angela J Yu ajyu@ucsd.edu
UCSD
Rajesh Rao rao@cs.washington.edu
U. Washington

Intelligent agents are often faced with the need to choose actions with uncertain consequences, and to modify those actions according to ongoing sensory processing and changing task demands. The requisite ability to dynamically modify or cancel planned actions is known as inhibitory control in psychology. We formalize inhibitory control as a rational decision-making problem, and apply to it to the classical stop-signal task. Using Bayesian inference and stochastic control tools, we show that the optimal policy systematically depends on various parameters of the problem, such as the relative costs of different action choices, the noise level of sensory inputs, and the dynamics of changing environmental demands. Our normative model accounts for a range of behavioral data in humans and animals in the stop-signal task, suggesting that the brain implements statistically optimal, dynamically adaptive, and reward-sensitive decision-making in the context of inhibitory control problems.
Subject Area: Neuroscience

ORAL SESSION 3 (10:50–11:10AM):

A Theory of Multiclass Boosting

Indraneel Mukherjee imukherj@princeton.edu
Robert E Schapire schapire@cs.princeton.edu
Princeton University

Boosting combines weak classifiers to form highly accurate predictors. Although the case of binary classification is well understood in the multiclass setting the “correct” requirements on the weak classifier or the notion of the most efficient boosting algorithms are missing. In this paper we create a broad and general framework within which we make precise and identify the optimal requirements on the weak-classifier as well as design the most effective in a certain sense boosting algorithms that assume such requirements.
Subject Area: Supervised Learning

SPOTLIGHTS SESSION 3 (11:10–11:30AM)

- ***Copula Processes***
  Andrew G Wilson and Zoubin Ghahramani, Cambridge.
  Subject Area: Probabilistic Models and Methods
  See abstract, page 69

- ***A biologically plausible network for the computation of orientation dominance***
  Kritika Muralidharan, SVCL; Nuno Vasconcelos, UC San Diego.
  Subject Area: Vision
  See abstract, page 59

- ***Learning Convolutional Feature Hierarchies for Visual Recognition***
  Koray Kavukcuoglu, Pierre Sermanet, Y-Lan Boureau, Karol Gregor, Michael Mathieu and Yann Le Cun, New York University.
  Subject Area: Supervised Learning
  See abstract, page 66

- ***LSTD with Random Projections***
  Mohammad Ghavamzadeh, Alessandro Lazaric, Odalric Maillard and Remi Munos, INRIA Lille
  Subject Area: Control and Reinforcement Learning
  See abstract, page 64

ORAL SESSION 4 (11:30–11:50AM):

Online Learning: Random Averages, Combinatorial Parameters, and Learnability

Alexander Rakhlin rakhlin@gmail.com
University of Pennsylvania
Karthik Sridharan karthik@ttic.edu
Toyota Technological Institute at Chicago
Ambuj Tewari ambuj@cs.utexas.edu
UT Austin

We develop a theory of online learning by defining several complexity measures. Among them are analogues of Rademacher complexity, covering numbers and fat-shattering dimension from statistical learning theory. Relationship among these complexity measures, their connection to online learning, and tools for bounding them are provided. We apply these results to various learning problems. We provide a complete characterization of online learnability in the supervised setting.
Subject Area: Theory

SPOTLIGHTS SESSION 4 (11:50AM–12:10PM)

- ***Kernel Descriptors for Visual Recognition***
  Liefeng Bo, University of Washington, Xiaofeng Ren, Intel, and Dieter Fox.
  Subject Area: Vision
  See abstract, page 60

- ***Minimum Average Cost Clustering***
  Kiyohito Nagano, University of Tokyo; Yoshinobu Kawahara, Osaka Univ; Satoru Iwata, Kyoto University.
  Subject Area: Optimization
  See abstract, page 56

- ***Identifying Dendritic Processing***
  Aurel A. Lazar and Yevgeniy B Slutskiy, Columbia Univ.
  Subject Area: Neuroscience
  See abstract, page 52

- ***Probabilistic Inference and Differential Privacy***
  Oliver Williams and Frank McSherry, Microsoft Research.
  Subject Area: Probabilistic Models and Methods
  See abstract, page 67

ORAL SESSION 5 (2:00–3:10PM):

INVITED TALK: Machine Learning with Human Intelligence: Principled Corner Cutting (PC2)

Xiao-Li Meng meng@stat.harvard.edu
Harvard University

With the ever increasing availability of quantitative information, especially data with complex spatial and/or temporal structures, two closely related fields are undergoing substantial evolution: Machine learning and Statistics. On a grand scale, both have the same goal: separating signal from noise. In terms of methodological choices, however, it is not uncommon to hear machine learners complain about statisticians‘; excessive worrying over modeling and inferential principles to a degree of being willing to produce nothing, and to hear statisticians express discomfort with machine learners‘; tendency to let ease of practical implementation trump principled justifications, to a point of being willing to deliver anything. To take advantage of the strengths of both fields, we need to train substantially more principled corner cutters. That is, we must train researchers who are at ease in formulating the solution from the soundest principles available, and equally at ease in cutting corners, guided by these principles, to retain as much statistical efficiency as feasible while maintaining algorithmic efficiency under time and resource constraints. This thinking process is demonstrated by applying the self-consistency principle (Efron, 1967; Lee, Li and Meng, 2010) to handling incomplete and/or irregularly spaced data with non-parametric and semi-parametric models, including signal processing via wavelets and sparsity estimation via the LASSO and related penalties.

Xiao-Li Meng is the Whipple V. N. Jones Professor of Statistics and Chair of the Department of Statistics at Harvard University. He was the recipient of the 2001 COPSS (Committee of Presidents of Statistical Societies) Award for "The outstanding statistician under the age of forty", of the 2003 Distinguished Achievement Award and of the 2008 Distinguished Service Award from the International Chinese Statistics Association, and of the 1997-1998 University of Chicago Faculty Award for Excellence in Graduate Teaching. His degrees include B.S. (Fudan Mathematics, 1982), M.A. (Harvard Statistics, 1987), and Ph.D. (Harvard Statistics, 1990). He has served on editorial boards of The Annals of Statistics, Biometrika, Journal of The American Statistical Association, Bayesian Analysis and Bernoulli, as well as the co-editor of Statistica Sinica. He is an elected fellow of ASA and of IMS. His research interests include: Statistical inference with partially observed data and simulated data; Quantifying statistical information and efficiency; Statistical principles and foundational issues, such as multi-party inferences, the theory of ignorance, and the interplay between Bayesian and frequentist perspectives; Effective deterministic and stochastic algorithms for Bayesian and likelihood computation; Markov chain Monte Carlo; Multi-resolution modelling for signal and image data; Statistical issues in astronomy and astrophysics; Modelling and imputation in health and medical studies; Elegant mathematical statistics.

Fast global convergence rates of gradient methods for high-dimensional statistical recovery

Alekh Agarwal alekhagarwal@gmail.com
Sahand Negahban sahand_n@eecs.Berkeley.edu
Martin Wainwright wainwrig@eecs.berkeley.edu
UC Berkeley

Many statistical M-estimators are based on convex optimization problems formed by the weighted sum of a loss function with a norm-based regularizer. We analyze the convergence rates of first-order gradient methods for solving such problems within a high-dimensional framework that allows the data dimension d to grow with (and possibly exceed) the sample size n. This high-dimensional structure precludes the usual global assumptions—namely strong convexity and smoothness conditions—that underlie classical optimization analysis. We define appropriately restricted versions of these conditions and show that they are satisfied with high probability for various statistical models. Under these conditions our theory guarantees that Nesterov's first-order method [Nesterov07] has a globally geometric rate of convergence up to the statistical precision of the model meaning the typical Euclidean distance between the true unknown parameter and the optimal solution b. This globally linear rate is substantially faster than previous analyses of global convergence for specific methods that yielded only sublinear rates. Our analysis applies to a wide range of M-estimators and statistical models including sparse linear regression using Lasso (`1-regularized regression) group Lasso block sparsity and low-rank matrix recovery using nuclear norm regularization. Overall this result reveals an interesting connection between statistical precision and computational efficiency in high-dimensional estimation.
Subject Area: Theory

SPOTLIGHTS SESSION 5 (3:10–3:30PM)

- ***Trading off Mistakes and Don't-Know Predictions***
  Amin Sayedi, Morteza and Avrim Blum, Carnegie Mellon University; Zadimoghaddam, MIT.
  Subject Area: Theory
  See abstract, page 55

- ***Exact learning curves for Gaussian process regression on* large random graphs**
  Matthew J Urry, and Peter Sollich, Kings College London.
  Subject Area: Probabilistic Models and Methods
  See abstract, page 69

- ***Synergies in learning words and their referents***
  Mark Johnson, Katherine Demuth, Macquarie University, Michael Frank, MIT, and Bevan K Jones, University of Edinburgh.
  Subject Area: Cognitive Science
  See abstract, page 57

- ***Learning concept graphs from text with stick-breaking priors***
  America Chambers, Padhraic Smyth and Mark Steyvers, University of California-Irvine.
  Subject Area: Unsupervised & Semi-supervised Learning
  See abstract, page 54

ORAL SESSION 6 (3:30–3:50PM):

Structured sparsity-inducing norms through submodular functions

Francis Bach francis.bach@ens.fr
Ecole Normale Superieure

Sparse methods for supervised learning aim at finding good linear predictors from as few variables as possible, i.e., with small cardinality of their supports. This combinatorial selection problem is often turned into a convex optimization problem by replacing the cardinality function by its convex envelope (tightest convex lower bound), in this case the L1-norm. In this paper, we investigate more general set-functions than the cardinality, that may incorporate prior knowledge or structural constraints which are common in many applications: namely, we show that for nondecreasing submodular set-functions, the corresponding convex envelope can be obtained from its Lovasz extension, a common tool in submodular analysis. This defines a family of polyhedral norms, for which we provide generic algorithmic tools (subgradients and proximal operators) and theoretical results (conditions for support recovery or high-dimensional inference). By selecting specific submodular functions, we can give a new interpretation to known norms, such as those based on rank-statistics or grouped norms with potentially overlapping groups; we also define new norms, in particular ones that can be used as non-factorial priors for supervised learning. Subject Area: Supervised Learning

ORAL SESSION 7 (4:20–5:00PM):

Semi-Supervised Learning with Adversarially Missing Label Information

Umar Syed usyed@cis.upenn.edu
Ben Taskar taskar@cis.upenn.edu
University of Pennsylvania

We address the problem of semi-supervised learning in an adversarial setting. Instead of assuming that labels are missing at random, we analyze a less favorable scenario where the label information can be missing partially and arbitrarily, which is motivated by several practical examples. We present nearly matching upper and lower generalization bounds for learning in this setting under reasonable assumptions about available label information. Motivated by the analysis, we formulate a convex optimization problem for parameter estimation, derive an efficient algorithm, and analyze its convergence. We provide experimental results on several standard data sets showing the robustness of our algorithm to the pattern of missing label information, outperforming several strong baselines.
Subject Area: Unsupervised & Semi-supervised Learning

MAP estimation in Binary MRFs via Bipartite Multi-cuts

Sashank Jakkam Reddi sashank@cse.iitb.ac.in
Sunita Sarawagi sunita.sarawagi@gmail.com
Sundar Vishwanathan sundar@cse.iitb.ac.in IIT
Bombay

We propose a new LP relaxation for obtaining the MAP assignment of a binary MRF with pairwise potentials. Our relaxation is derived from reducing the MAP assignment problem to an instance of a recently proposed Bipartite Multi-cut problem where the LP relaxation is guaranteed to provide an O(log *k*) approximation where *k* is the number of vertices adjacent to non-submodular edges in the MRF. We then propose a combinatorial algorithm to efficiently solve the LP and also provide a lower bound by concurrently solving its dual to within an $\in$ approximation. The algorithm is up to an order of magnitude faster and provides better MAP scores and bounds than the state of the art message passing algorithm that tightens the local marginal polytope with third-order marginal constraints. Subject Area: Probabilistic Models and Methods

SPOTLIGHTS SESSION 6 (5:00–5:20PM)

- ***Efficient Minimization of Decomposable Submodular Functions***
  Peter G Stobbe and Andreas Krause, Caltech.
  Subject Area: Optimization
  See abstract, page 56

- ***Worst-case bounds on the quality of max-product fixed-points***
  Meritxell Vinyals, Jes´us Cerquides, IIIA-CSIC, Alessandro Farinelli, University of Verona, and Juan Antonio Rodr´ıguez-Aguilar, IIIA-CSIC.
  Subject Area: Probabilistic Models and Methods
  See abstract, page 68

- ***Learning To Count Objects in Images***
  Victor Lempitsky and Andrew Zisserman, Oxford.
  Subject Area: Vision
  See abstract, page 58

- ***Learning invariant features using the Transformed Indian Buffet Process***
  Joseph L Austerweil, UC Berkeley, and Tom Griffiths, University of California-Berkeley.
  Subject Area: Cognitive Science
  See abstract, page 57

ORAL SESSION 8 (5:20–5:40PM):

A Dirty Model for Multi-task Learning

Ali Jalali alij@mail.utexas.edu
Pradeep Ravikumar pradeepr@cs.utexas.edu
Sujay Sanghavi sanghavi@mail.utexas.edu
Chao Ruan ruan@cs.utexas.edu
University of Texas at Austin

We consider the multiple linear regression problem, in a setting where some of the set of relevant features could be shared across the tasks. A lot of recent research has studied the use of ℓ_1/ℓ_q norm block-regularizations with q > 1 for such (possibly) block-structured problems, establishing strong guarantees on recovery even under high-dimensional scaling where the number of features scale with the number of observations. However, these papers also caution that the performance of such block-regularized methods are very dependent on the extent to which the features are shared across tasks. Indeed they show that if the extent of overlap is less than a threshold, or even if parameter values in the shared features are highly uneven, then block ℓ_1/ℓ_q regularization could actually perform worse than simple separate elementwise ℓ_1 regularization. We are far away from a realistic multitask setting: not only do the set of relevant features have to be exactly the same across tasks, but their values have to as well. Here we ask the question: can we leverage support and parameter overlap when it exists but not pay a penalty when it does not? Indeed this falls under a more general question of whether we can model such dirty data which may not fall into a single neat structural bracket (all block-sparse or all low-rank and so on). Here we take a first step focusing on developing a dirty model for the multiple regression problem. Our method uses a very simple idea: we decompose the parameters into two components and regularize these differently. We show both theoretically and empirically our method strictly and noticeably outperforms both ℓ_1 and ℓ_1/ℓ_q methods over the entire range of possible overlaps. We also provide theoretical guarantees that the method performs well under high-dimensional scaling. Subject Area: Theory

SPOTLIGHTS SESSION 7 (5:40–6:00PM)

- ***Variational Inference over Combinatorial Spaces***
  Alexandre Bouchard-Cote and Michael I Jordan, University of California Berkeley.
  Subject Area: Probabilistic Models and Methods
  See abstract, page 67

- ***Multiple Kernel Learning and the SMO Algorithm***
  S.V.N. Vishwanathan, Zhaonan Sun and Nawanol T Ampornpunt, Purdue University; Manik Varma, Microsoft Research.
  Subject Area: Supervised Learning
  See abstract, page 65

- ***Inductive Regularized Learning of Kernel Functions***
  Prateek Jain, Microsoft Research India Lab, Brian Kulis, University of California Berkeley, and Inderjit Dhillon, University of Texas.
  Subject Area: Unsupervised and Semi-supervised Learning
  See abstract, page 63

- ***Probabilistic Deterministic Infinite Automata***
  David Pfau, Nicholas Bartlett and Frank Wood, Columbia University.
  Subject Area: Probabilistic Models and Methods
  See abstract, page 69

T1 Inferring Stimulus Selectivity from the Spatial Structure of Neural Network Dynamics

Kanaka Rajan krajan@princeton.edu
Princeton University
L F Abbott lfa2103@columbia.edu
Columbia University
Haim Sompolinsky haim@fiz.huji.ac.il
Hebrew University and Harvard University

How are the spatial patterns of spontaneous and evoked population responses related? We study the impact of connectivity on the spatial pattern of fluctuations in the inputgenerated response of a neural network, by comparing the distribution of evoked and intrinsically generated activity across the different units. We develop a complementary approach to principal component analysis in which separate high-variance directions are typically derived for each input condition. We analyze subspace angles to compute the difference between the shapes of trajectories corresponding to different network states, and the orientation of the low-dimensional subspaces that driven trajectories occupy within the full space of neuronal activity. In addition to revealing how the spatiotemporal structure of spontaneous activity affects input-evoked responses, these methods can be used to infer input selectivity induced by network dynamics from experimentally accessible measures of spontaneous activity (e.g. from voltage- or calcium-sensitive optical imaging experiments). We conclude that the absence of a detailed spatial map of afferent inputs and cortical connectivity does not limit our ability to design spatially extended stimuli that evoke strong responses.
Subject Area: Neuroscience

T2 Individualized ROI Optimization via Maximization of Group-wise Consistency of Structural and Functional Profiles

Kaiming Li likaiming@gmail.com
NWPU
Lei Guo guolei.npu@gmail.com
Carlos Faraco cfaraco@uga.edu
Dajiang Zhu dajiang.zhu@gmail.com
Fan Deng enetoremail@gmail.com
University of Georgia
Tuo Zhang zhangtuo.npu@gmail.com
Xi Jiang superjx2318@hotmail.com
Degang Zhang lczhdgm@gmail.com
Hanbo Chen cojoc@hotmail.com
Northwestern Polytechnical Unv
Xintao Hu xintao.hu@gmail.com
Steve Miller lsmiller@uga.edu
Tianming Liu tianming.liu@gmail.com

Functional segregation and integration are fundamental characteristics of the human brain. Studying the connectivity among segregated regions and the dynamics of integrated brain networks has drawn increasing interest. A very controversial, yet fundamental issue in these studies is how to determine the best functional brain regions or ROIs (regions of interests) for individuals. Essentially, the computed connectivity patterns and dynamics of brain networks are very sensitive to the locations, sizes, and shapes of the ROIs. This paper presents a novel methodology to optimize the locations of an individual's ROIs in the working memory system. Our strategy is to formulate the individual ROI optimization as a group variance minimization problem, in which group-wise functional and structural connectivity patterns, and anatomic profiles are defined as optimization constraints. The optimization problem is solved via the simulated annealing approach. Our experimental results show that the optimized ROIs have significantly improved consistency in structural and functional profiles across subjects, and have more reasonable localizations and more consistent morphological and anatomic profiles.
Subject Area: Neuroscience

T3 Categories and Functional Units: An Infinite Hierarchical Model for Brain Activations

Danial Lashkari danial@csail.mit.edu
Ramesh Sridharan rameshvs@csail.mit.edu
Polina Golland polina@csail.mit.edu
MIT

We present a model that describes the structure in the responses of different brain areas to a set of stimuli in terms of ”stimulus categories” (clusters of stimuli) and ”functional units” (clusters of voxels). We assume that voxels within a unit respond similarly to all stimuli from the same category, and design a nonparametric hierarchical model to capture intersubject variability among the units. The model explicitly captures the relationship between brain activations and fMRI time courses. A variational inference algorithm derived based on the model can learn categories, units, and a set of unit-category activation probabilities from data. When applied to data from an fMRI study of object recognition the method finds meaningful and consistent clusterings of stimuli into categories and voxels into units.
Subject Area: Neuroscience

T4 Sodium entry efficiency during action potentials: A novel single-parameter family of Hodgkin-Huxley models

Anand Singh anands@pharma.uzh.ch
Renaud Jolivet renaud.jolivet@a3.epfl.ch
Bruno Weber bweber@pharma.uzh.ch
University of Zurich
Pierre J Magistretti pierre.magistretti@epfl.ch
EPFL

Sodium entry during an action potential determines the energy efficiency of a neuron. The classic Hodgkin-Huxley model of action potential generation is notoriously inefficient in that regard with about 4 times more charges flowing through the membrane than the theoretical minimum required to achieve the observed depolarization. Yet, recent experimental results show that mammalian neurons are close to the optimal metabolic efficiency and that the dynamics of their voltage-gated channels is significantly different than the one exhibited by the classic Hodgkin-Huxley model during the action potential. Nevertheless, the original Hodgkin-Huxley model is still widely used and rarely to model the squid giant axon from which it was extracted. Here, we introduce a novel family of Hodgkin-Huxley models that correctly account for sodium entry, action potential width and whose voltage-gated channels display a dynamics very similar to the most recent experimental observations in mammalian neurons. We speak here about a family of models because the model is parameterized by a unique parameter the variations of which allow to reproduce the entire range of experimental observations from cortical pyramidal neurons to Purkinje cells, yielding a very economical framework to model a wide range of different central neurons. The present paper demonstrates the performances and discuss the properties of this new family of models.

Subject Area: Neuroscience

T5 Identifying Dendritic Processing

Aurel A. Lazar aurel@ee.columbia.edu
Yevgeniy B Slutskiy ys2146@columbia.edu
Columbia University

In system identification both the input and the output of a system are available to an observer and an algorithm is sought to identify parameters of a hypothesized model of that system. Here we present a novel formal methodology for identifying dendritic processing in a neural circuit consisting of a linear dendritic processing filter in cascade with a spiking neuron model. The input to the circuit is an analog signal that belongs to the space of bandlimited functions. The output is a time sequence associated with the spike train. We derive an algorithm for identification of the dendritic processing filter and reconstruct its kernel with arbitrary precision.

Subject Area: Neuroscience
Spotlight presentation, Tuesday, 12:00.

T6 Attractor Dynamics with Synaptic Depression

C. C. Alan Fung alanfung@ust.hk
K. Y. Michael Wong phkywong@ust.hk
The Hong Kong University of Science and Technology
He Wang wanghe07@mails.tsinghua.edu.cn
Tsinghua University
Si Wu siwu@ion.ac.cn

Institute of Neuroscience Chinese Academy of Sciences

Neuronal connection weights exhibit short-term depression (STD). The present study investigates the impact of STD on the dynamics of a continuous attractor neural network (CANN) and its potential roles in neural information processing. We find that the network with STD can generate both static and traveling bumps, and STD enhances the performance of the network in tracking external inputs. In particular, we find that STD endows the network with slow-decaying plateau behaviors, namely, the network being initially stimulated to an active state will decay to silence very slowly in the time scale of STD rather than that of neural signaling. We argue that this provides a mechanism for neural systems to hold short-term memory easily and shut off persistent activities naturally.

Subject Area: Neuroscience
Spotlight presentation, Tuesday, 9:55.

T7 A rational Decision Making Framework for Inhibitory Control

Pradeep Shenoy and Angela J Yu, UCSD; Rajesh Rao, U. Washington
UCSD. Subject Area: Neuroscience

See abstract, page 47

T8 The Neural Costs of Optimal Control

Samuel Gershman sjgershm@princeton.edu
Robert Wilson rcw2@princeton.edu
Princeton University

Optimal control entails combining probabilities and utilities. However, for most practical problems probability densities can be represented only approximately. Choosing an approximation requires balancing the benefits of an accurate approximation against the costs of computing it. We propose a variational framework for achieving this balance and apply it to the problem of how a population code should optimally represent a distribution under resource constraints. The essence of our analysis is the conjecture that population codes are organized to maximize a lower bound on the log expected utility. This theory can account for a plethora of experimental data, including the reward-modulation of sensory receptive fields.

Subject Area: Neuroscience

T9 Over-Complete Representations on Recurrent Neural Networks can Support Persistent Percepts

Shaul Druckmann and Dmitri B Chklovskii, JFRC.

Subject Area: Neuroscience
Oral presentation, Tuesday, 9:20am.
See abstract, page 47

T10 Rescaling, Thinning or Complementing? On Goodness-of-fit Procedures for Point Process Models and Generalized Linear Models

Felipe Gerhard felipe.gerhard@epfl.ch
Wulfram Gerstner wulfram.gerstner@epfl.ch
Brain Mind Institute EPFL

Generalized Linear Models (GLMs) are an increasingly popular framework for modeling neural spike trains. They have been linked to the theory of stochastic point processes and researchers have used this relation to assess goodness-of-fit using methods from pointprocess theory, e.g. the time-rescaling theorem. However, high neural firing rates or coarse discretization lead to a breakdown of the assumptions necessary for this connection. Here, we show how goodness-of-fit tests from point-process theory can still be applied to GLMs by constructing equivalent surrogate point processes out of time-series observations. Furthermore, two additional tests based on thinning and complementing point processes are introduced. They augment the instruments available for checking model adequacy of point processes as well as discretized models.

Subject Area: Neuroscience

T11 Divisive Normalization: Justification and Effectiveness as Efficient Coding Transform

Siwei Lyu lsw@cs.albany.edu
University at Albany SUNY

Divisive normalization (DN) has been advocated as an effective nonlinear efficient coding transform for natural sensory signals with applications in biology and engineering. In this work, we aim to establish a connection between the DN transform and the statistical properties of natural sensory signals. Our analysis is based on the use of multivariate t model to capture some important statistical properties of natural sensory signals. The multivariate t model justifies DN as an approximation to the transform that completely eliminates its statistical dependency. Furthermore, using the multivariate t model and measuring statistical dependency with multi-information, we can precisely quantify the statistical dependency that is reduced by the DN transform. We compare this with the actual performance of the DN transform in reducing statistical dependencies of natural sensory signals. Our theoretical analysis and quantitative evaluations confirm DN as an effective efficient coding transform for natural sensory signals. On the other hand, we also observe a previously unreported phenomenon that DN may increase statistical dependencies when the size of pooling is small.

Subject Area: Neuroscience

T12 Evaluating neuronal codes for inference using Fisher information

Haefner M Ralf ralf.haefner@tuebingen.mpg.de
Matthias Bethge matthias.bethge@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics

Many studies have explored the impact of response variability on the quality of sensory codes. The source of this variability is almost always assumed to be intrinsic to the brain. However, when inferring a particular stimulus property, variability associated with other stimulus attributes also effectively act as noise. Here we study the impact of such stimulus-induced response variability for the case of binocular disparity inference. We characterize the response distribution for the binocular energy model in response to random dot stereograms and find it to be very different from the Poisson-like noise usually assumed. We then compute the Fisher information with respect to binocular disparity, present in the monocular inputs to the standard model of early binocular processing, and thereby obtain an upper bound on how much information a model could theoretically extract from them. Then we analyze the information loss incurred by the different ways of combining those inputs to produce a scalar single-neuron response. We find that in the case of depth inference, monocular stimulus variability places a greater limit on the extractable information than intrinsic neuronal noise for typical spike counts. Furthermore, the largest loss of information is incurred by the standard model for position disparity neurons (tuned-excitatory), that are the most ubiquitous in monkey primary visual cortex, while more information from the inputs is preserved in phase-disparity neurons (tuned-near or tuned-far) primarily found in higher cortical regions.

Subject Area: Neuroscience

T13 Learning the context of a category

Dan Navarro daniel.navarro@adelaide.edu.au
University of Adelaide

This paper outlines a hierarchical Bayesian model for human category learning that learns both the organization of objects into categories, and the context in which this knowledge should be applied. The model is fit to multiple data sets, and provides a parsimonious method for describing how humans learn context specific conceptual representations. Subject Area: Cognitive Science T34 Learning To Count Objects in Images Victor Lempitsky victorlempitsky@gmail.com U. Oxford Andrew Zisserman az@robots.ox.ac.uk University of Oxford We propose a new supervised learning framework for visual object counting tasks, such as estimating the number of cells in a microscopic image or the number of humans in surveillance video frames. We focus on the practically-attractive case when the training images are annotated with dots (one dot per object). Our goal is to accurately estimate the count. However we evade the hard task of learning to detect and localize individual object instances. Instead we cast the problem as that of estimating an image density whose integral over any image region gives the count of objects within that region. Learning to infer such density can be formulated as a minimization of a regularized risk quadratic cost function. We introduce a new loss function which is well-suited for such learning and at the same time can be computed efficiently via a maximum subarray algorithm. The learning can then be posed as a convex quadratic program solvable with cutting-plane optimization. The proposed framework is very flexible as it can accept any domain-specific visual features. Once trained our system provides accurate object counts and requires a very small time overhead over the feature extraction step making it a good candidate for applications involving real-time processing or dealing with huge amount of visual data.
Subject Area: Vision

T14 Inference and communication in the game of Password

Yang Xu yang_xu_ch@yahoo.com
Charles Kemp ckemp@cmu.edu
Carnegie Mellon University

Communication between a speaker and hearer will be most efficient when both parties make accurate inferences about the other. We study inference and communication in a television game called Password, where speakers must convey secret words to hearers by providing one-word clues. Our working hypothesis is that human communication is relatively efficient, and we use game show data to examine three predictions. First, we predict that speakers and hearers are both considerate, and that both take the other's perspective into account. Second, we predict that speakers and hearers are calibrated, and that both make accurate assumptions about the strategy used by the other. Finally, we predict that speakers and hearers are collaborative, and that they tend to share the cognitive burden of communication equally. We find evidence in support of all three predictions, and demonstrate in addition that efficient communication tends to break down when speakers and hearers are placed under time pressure.
Subject Area: Cognitive Science

T15 Learning Bounds for Importance Weighting

Corinna Cortes corinna@google.com
Google Inc
Yishay Mansour mansour.yishay@gmail.com
Tel-Aviv
Mehryar Mohri mohri@cims.nyu.edu

This paper presents an analysis of importance weighting for learning from finite samples and gives a series of theoretical and algorithmic results. We point out simple cases where importance weighting can fail, which suggests the need for an analysis of the properties of this technique. We then give both upper and lower bounds for generalization with bounded importance weights and, more significantly, give learning guarantees for the more common case of unbounded importance weights under the weak assumption that the second moment is bounded, a condition related to the Renyi divergence of the training and test distributions. These results are based on a series of novel and general bounds we derive for unbounded loss functions, which are of independent interest. We use these bounds to guide the definition of an alternative reweighting algorithm and report the results of experiments demonstrating its benefits. Finally, we analyze the properties of normalized importance weights which are also commonly used.
Subject Area: Theory

T16 Smoothness, Low Noise and Fast Rates

Nathan Srebro nati@ttic.edu
Karthik Sridharan karthik@ttic.edu
Toyota Technological Institute at Chicago
Ambuj Tewari ambujtewari@gmail.com
UT Austin

We establish an excess risk bound of $O(HR^2_n + sqrtH\, L^*R_n)$ for ERM with an H-smooth loss function and a hypothesis class with Rademacher complexity R_n, where L* is the best risk achievable by the hypothesis class. For typical hypothesis classes where $R_n = \sqrt{}R/n$, this translates to a learning rate of $O(RH/n)$ in the separable $(L^* = 0)$ case and $O(RH/n + sqrt\ L^*\ RH/n)$ more generally. We also provide similar guarantees for online and stochastic convex optimization of a smooth non-negative objective.
Subject Area: Theory

T17 Online Learning: Random Averages, Combinatorial Parameters, and Learnability

Alexander Rakhlin,
University of Pennsylvania,
Karthik Sridharan,
Toyota Technological Institute at Chicago,
Ambuj Tewari,
UT Austin.

Subject Area: Theory
Oral presentation, Tuesday, 11:30am.
See abstract, page 54

T18 An analysis on negative curvature induced by singularity in multi-layer neural-network learning

Eiji Mizutani eiji@mail.ntust.edu.tw
Taiwan Univ. Science and Tech.
Stuart Dreyfus dreyfus@ieor.berkeley.edu
UC Berkeley

In the neural-network parameter space, an attractive field is likely to be induced by singularities. In such a singularity region first-order gradient learning typically causes a long plateau with very little change in the objective function value E (hence a flat region). Therefore it may be confused with “attractive” local minima. Our analysis shows that the Hessian matrix of E tends to be indefinite in the vicinity of (perturbed) singular points suggesting a promising strategy that exploits negative curvature so as to escape from the singularity plateaus. For numerical evidence we limit the scope to small examples (some of which are found in journal papers) that allow us to confirm singularities and the eigenvalues of the Hessian matrix and for which computation using a descent direction of negative curvature encounters no plateau. Even for those small problems no efficient methods have been previously developed that avoided plateaus.
Subject Area: Theory

T19 Trading off Mistakes and Don’t-Know Predictions

Amin Sayedi ssayedir@cmu.edu
Avrim Blum avrim@cs.cmu.edu
Carnegie Mellon University
Morteza Zadimoghaddam morteza@mit.edu
MIT

We discuss an online learning framework in which the agent is allowed to say “I don’t know” as well as making incorrect predictions on given examples. We analyze the trade off between saying “I don’t know” and making mistakes. If the number of don’t know predictions is forced to be zero, the model reduces to the well-known mistake-bound model introduced by Littlestone [Lit88]. On the other hand, if no mistakes are allowed, the model reduces to KWIK framework introduced by Li et. al. [LLW08]. We propose a general, though inefficient, algorithm for general finite concept classes that minimizes the number of don’t-know predictions if a certain number of mistakes are allowed. We then present specific polynomial-time algorithms for the concept classes of monotone disjunctions and linear separators.
Subject Area: Theory
Spotlight presentation, Tuesday, 3:10.

T20 Fast global convergence rates of gradient methods for high-dimensional statistical recovery

Alekh Agarwal, Sahand Negahban and Martin Wainwright, UC Berkeley.
Subject Area: Theory Oral presentation, Tuesday,
See abstract, page 49

T21 A Dirty Model for Multi-task Learning

Ali Jalali, Pradeep Ravikumar, Sujay Sanghavi and Chao Ruan, University of Texas at Austin.
Subject Area: Theory
See abstract, page 50

T22 Variable margin losses for classifier design

Hamed Masnadi-Shirazi hmasnadi@ucsd.edu
Nuno Vasconcelos nuno@ucsd.edu
UC San Diego

The problem of controlling the margin of a classifier is studied. A detailed analytical study is presented on how properties of the classification risk, such as its optimal link and minimum risk functions, are related to the shape of the loss, and its margin enforcing properties. It is shown that for a class of risks, denoted canonical risks, asymptotic Bayes consistency is compatible with simple analytical relationships between these functions. These enable a precise characterization of the loss for a popular class of link functions. It is shown that, when the risk is in canonical form and the link is inverse sigmoidal, the margin properties of the loss are determined by a single parameter. Novel families of Bayes consistent loss functions, of variable margin, are derived. These families are then used to design boosting style algorithms with explicit control of the classification margin. The new algorithms generalize well established approaches such as LogitBoost. Experimental results show that the proposed variable margin losses outperform the fixed margin counterparts used by existing algorithms. Finally it is shown that best performance can be achieved by cross-validating the margin parameter.
Subject Area: Theory

T23 Robust PCA via Outlier Pursuit

Huan Xu huan.xu@mail.utexas.edu
Constantine Caramanis caramanis@mail.utexas.edu
Sujay Sanghavi sanghavi@mail.utexas.edu
The University of Texas

Singular Value Decomposition (and Principal Component Analysis) is one of the most widely used techniques for dimensionality reduction: successful and efficiently computable, it is nevertheless plagued by a well-known, well-documented sensitivity to outliers. Recent work has considered the setting where each point has a few arbitrarily corrupted components. Yet, in applications of SVD or PCA such as robust collaborative filtering or bioinformatics, malicious agents, defective genes, or simply corrupted or contaminated experiments may effectively yield entire points that are completely corrupted. We present an efficient convex optimization-based algorithm we call Outlier Pursuit that under some mild assumptions on the uncorrupted points (satisfied e.g. by the standard generative assumption in PCA problems) recovers the *exact* optimal low-dimensional subspace and identifies the corrupted points. Such identification of corrupted points that do not conform to the low-dimensional approximation is of paramount interest in bioinformatics and financial applications and beyond. Our techniques involve matrix decomposition using nuclear norm minimization however our results setup and approach necessarily differ considerably from the existing line of work in matrix completion and matrix decomposition since we develop an approach to recover the correct *column space* of the uncorrupted matrix rather than the exact matrix itself.
Subject Area: Optimization

T24 Network Flow Algorithms for Structured Sparsity

Julien Mairal julien.mairal@m4x.org
Rodolphe Jenatton rodolphe.jenatton@inria.fr
Guillaume R Obozinski guillaume.obozinski@inria.fr
INRIA
Francis Bach francis.bach@ens.fr
Ecole Normale Superieure

We consider a class of learning problems that involve a structured sparsity-inducing norm defined as the sum of ℓ_∞-norms over groups of variables. Whereas a lot of effort has been put in developing fast optimization methods when the groups are disjoint or embedded in a specific hierarchical structure, we address here the case of general overlapping groups. To this end, we show that the corresponding optimization problem is related to network flow optimization. More precisely, the proximal problem associated with the norm we consider is dual to a quadratic min-cost flow problem. We propose an efficient procedure which computes its solution exactly in polynomial time. Our algorithm scales up to millions of groups and variables, and opens up a whole new range of applications for structured sparse models. We present several experiments on image and video data, demonstrating the applicability and scalability of our approach for various problems.
Subject Area: Optimization

T25 Efficient Minimization of Decomposable Submodular Functions

Peter G Stobbe stobbe@acm.caltech.edu
Andreas Krause krausea@caltech.edu
Caltech

Many combinatorial problems arising in machine learning can be reduced to the problem of minimizing a submodular function. Submodular functions are a natural discrete analog of convex functions, and can be minimized in strongly polynomial time. Unfortunately, state-of-the-art algorithms for general submodular minimization are intractable for practical problems. In this paper, we introduce a novel subclass of submodular minimization problems that we call decomposable. Decomposable submodular functions are those that can be represented as sums of concave functions applied to linear functions. We develop an algorithm, SLG, that can efficiently minimize decomposable submodular functions with tens of thousands of variables. Our algorithm exploits recent results in smoothed convex minimization. We apply SLG to synthetic benchmarks and a joint classification-and-segmentation task, and show that it outperforms the state-of-the-art general purpose submodular minimization algorithms by several orders of magnitude.
Subject Area: Optimization
Spotlight presentation, Tuesday, 5:00.

T26 Minimum Average Cost Clustering

Kiyohito Nagano nagano@sat.t.u-tokyo.ac.jp
University of Tokyo
Yoshinobu Kawahara kawahara@ar.sanken.osaka-u.ac.jp
Osaka University
Satoru Iwata iwata@kurims.kyoto-u.ac.jp
Kyoto University

A number of objective functions in clustering problems can be described with submodular functions. In this paper, we introduce the minimum average cost criterion, and show that the theory of intersecting submodular functions can be used for clustering with submodular objective functions. The proposed algorithm does not require the number of clusters in advance, and it will be determined by the property of a given set of data points. The minimum average cost clustering problem is parameterized with a real variable, and surprisingly, we show that all information about optimal clusterings for all parameters can be computed in polynomial time in total. Additionally, we evaluate the performance of the proposed algorithm through computational experiments.
Subject Area: Optimization
Spotlight presentation, Monday, 11:55.

T27 Practical Large-Scale Optimization for Max-norm Regularization

Jason Lee jl115@yahoo.com
Nathan Srebro nati@ttic.edu
Toyota Technological Institute at Chicago
Ben Recht brecht@cs.wisc.edu
UW-Madison
Ruslan Salakhutdinov rsalakhu@mit.edu
MIT
Joel Tropp jtropp@acm.caltech.edu
CalTech

The max-norm was proposed as a convex matrix regularizer by Srebro et al (2004) and was shown to be empirically superior to the trace-norm for collaborative filtering problems. Although the max-norm can be computed in polynomial time, there are currently no practical algorithms for solving large-scale optimization problems that incorporate the max-norm. The present work uses a factorization technique of Burer and Monteiro (2003) to devise scalable first-order algorithms for convex programs involving the max-norm. These algorithms are applied to solve huge collaborative filtering, graph cut, and clustering problems. Empirically, the new methods outperform mature techniques from all three areas.
Subject Area: Optimization

T28 Learning invariant features using the Transformed Indian Buffet Process

Joseph L Austerweil joseph.austerweil@gmail.com
Tom Griffiths tom_griffiths@berkeley.edu
University of California-Berkeley

Identifying the features of objects becomes a challenge when those features can change in their appearance. We introduce the Transformed Indian Buffet Process (tIBP), and use it to define a nonparametric Bayesian model that infers features that can transform across instantiations. We show that this model can identify features that are location invariant by modeling a previous experiment on human feature learning. However, allowing features to transform adds new kinds of ambiguity: Are two parts of an object the same feature with different transformations or two unique features? What transformations can features undergo? We present two new experiments in which we explore how people resolve these questions, showing that the tIBP model demonstrates a similar sensitivity to context to that shown by human learners when determining the invariant aspects of features.
Subject Area: Cognitive Science
Spotlight presentation, Monday, 5:15.

T29 Synergies in learning words and their referents

Mark Johnson mark.johnson@mq.edu.au
Katherine Demuth Katherine.Demuth@mq.edu.au
Macquarie University
Michael Frank mcfrank@mit.edu
MIT
Bevan K Jones B.K.Jones@sms.ed.ac.uk
University of Edinburgh

This paper presents Bayesian non-parametric models that simultaneously learn to segment words from phoneme strings and learn the referents of some of those words, and shows that there is a synergistic interaction in the acquisition of these two kinds of linguistic information. The models themselves are novel kinds of Adaptor Grammars that are an extension of an embedding of topic models into PCFGs. These models simultaneously segment phoneme sequences into words and learn the relationship between non-linguistic objects to the words that refer to them. We show (i) that modelling inter-word dependencies not only improves the accuracy of the word segmentation but also of word-object relationships, and (ii) that a model that simultaneously learns word-object relationships and word segmentation segments more accurately than one that just learns word segmentation on its own. We argue that these results support an interactive view of language acquisition that can take advantage of synergies such as these.
Subject Area: Cognitive Science
Spotlight presentation, Monday, 3:20.

T30 Permutation Complexity Bound on Out-Sample Error

Malik Magdon-Ismail magdon@cs.rpi.edu
RPI

We define a data dependent permutation complexity for a hypothesis set H which is similar to a Rademacher complexity or maximum discrepancy. The permutation complexity is based like the maximum discrepancy on (dependent) sampling. We prove a uniform bound on the generalization error as well as a concentration result which means that the permutation estimate can be efficiently estimated.
Subject Area: Theory

T31 Optimal learning rates for Kernel Conjugate Gradient regression

Gilles Blanchard gilles.blanchard@gmail.com
Weierstrass Institute Berlin / University of Potsdam
Nicole Kramer nicole.kraemer@wias-berlin.de
Weierstrass Institute Berlin

We prove rates of convergence in the statistical sense for kernel-based least squares regression using a conjugate gradient algorithm, where regularization against overfitting is obtained by early stopping. This method is directly related to Kernel Partial Least Squares a regression method that combines supervised dimensionality reduction with least squares projection. The rates depend on two key quantities: first on the regularity of the target regression function and second on the effective dimensionality of the data mapped into the kernel space. Lower bounds on attainable rates depending on these two quantities were established in earlier literature and we obtain upper bounds for the considered method that match these lower bounds (up to a log factor) if the true regression function belongs to the reproducing kernel Hilbert space. If the latter assumption is not fulfilled we obtain similar convergence rates provided additional unlabeled data are available. The order of the learning rates in these two situations match state-of-the-art results that were recently obtained for the least squares support vector machine and for linear regularization operators.
Subject Area: Theory

T32 Nonparametric Density Estimation for Stochastic Optimization with an Observable State Variable

Lauren Hannah lauren.hannah@duke.edu
Warren B Powell powell@princeton.edu
David Blei blei@cs.princeton.edu
Princeton University

We study convex stochastic optimization problems where a noisy objective function value is observed after a decision is made. There are many stochastic optimization problems whose behavior depends on an exogenous state variable which affects the shape of the objective function. Currently, there is no general purpose algorithm to solve this class of problems. We use nonparametric density estimation for the joint distribution of stateoutcome pairs to create weights for previous observations. The weights effectively group similar states. Those similar to the current state are used to create a convex, deterministic approximation of the objective function. We propose two solution methods that depend on the problem characteristics: function-based and gradient-based optimization. We offer two weighting schemes, kernel based weights and Dirichlet process based weights, for use with the solution methods. The weights and solution methods are tested on a synthetic multiproduct newsvendor problem and the hour ahead wind commitment problem. Our results show Dirichlet process weights can offer substantial benefits over kernel based weights and, more generally, that nonparametric estimation methods provide good solutions to otherwise intractable problems.
Subject Area: Optimization

T33 Towards Holistic Scene Understanding: Feedback Enabled Cascaded Classification Models

Congcong Li cl758@cornell.edu
Adarsh P Kowdle apk64@cornell.edu
Ashutosh Saxena asaxena@cs.cornell.edu
Tsuhan Chen tsuhan@ece.cornell.edu
Cornell University

In many machine learning domains (such as scene understanding), several related sub-tasks (such as scene categorization, depth estimation, object detection) operate on the same raw data and provide correlated outputs. Each of these tasks is often notoriously hard, and state-of-the-art classifiers already exist for many sub-tasks. It is desirable to have an algorithm that can capture such correlation without requiring to make any changes to the inner workings of any classifier. We propose Feedback Enabled Cascaded Classification Models (FE-CCM) that maximizes the joint likelihood of the sub-tasks while requiring only a 'black-box' interface to the original classifier for each sub-task. We use a two-layer cascade of classifiers which are repeated instantiations of the original ones with the output of the first layer fed into the second layer as input. Our training method involves a feedback step that allows later classifiers to provide earlier classifiers information about what error modes to focus on. We show that our method significantly improves performance in all the sub-tasks in two different domains: (i) scene understanding where we consider depth estimation scene categorization event categorization object detection geometric labeling and saliency detection and (ii) robotic grasping where we consider grasp point detection and object classification.
Subject Area: Vision

T34 Learning To Count Objects in Images

Victor Lempitsky victorlempitsky@gmail.com
Andrew Zisserman az@robots.ox.ac.uk
University of Oxford

We propose a new supervised learning framework for visual object counting tasks, such as estimating the number of cells in a microscopic image or the number of humans in surveillance video frames. We focus on the practically-attractive case when the training images are annotated with dots (one dot per object). Our goal is to accurately estimate the count. However we evade the hard task of learning to detect and localize individual object instances. Instead we cast the problem as that of estimating an image density whose integral over any image region gives the count of objects within that region. Learning to infer such density can be formulated as a minimization of a regularized risk quadratic cost function. We introduce a new loss function which is well-suited for such learning and at the same time can be computed efficiently via a maximum subarray algorithm. The learning can then be posed as a convex quadratic program solvable with cutting-plane optimization. The proposed framework is very flexible as it can accept any domain-specific visual features. Once trained our system provides accurate object counts and requires a very small time overhead over the feature extraction step making it a good candidate for applications involving real-time processing or dealing with huge amount of visual data.
Subject Area: Vision
Spotlight presentation, Tuesday, 5:10.

T35 A biologically Plausible Network for the Computation of Orientation Dominance

Kritika Muralidharan krmurali@ucsd.edu
Nuno Vasconcelos nuno@ece.ucsd.edu
UC San Diego

The determination of dominant orientation at a given image location is formulated as a decision-theoretic question. This leads to a novel measure for the dominance of a given orientation , which is similar to that used by SIFT. It is then shown that the new measure can be computed with a network that implements the sequence of operations of the standard neurophysiological model of V1. The measure can thus be seen as a biologically plausible version of SIFT, and is denoted as bioSIFT. The network units are shown to exhibit trademark properties of V1 neurons, such as cross-orientation suppression, sparseness and independence. The connection between SIFT and biological vision provides a justification for the success of SIFT-like features and reinforces the importance of contrast normalization in computer vision. We illustrate this by replacing the Gabor units of an HMAX network with the new bioSIFT units. This is shown to lead to significant gains for classification tasks, leading to state-of-the-art performance among biologically inspired network models and performance competitive with the best non-biological object recognition systems.
Subject Area: Vision
Spotlight presentation, Monday, 11:15.

T36 Pose-Sensitive Embedding by Nonlinear NCA Regression

Graham Taylor gwtaylor@cs.nyu.edu
Rob Fergus fergus@cs.nyu.edu
George Williams gwilliam@cs.nyu.edu
Ian Spiro spiro@cs.nyu.edu
Christoph Bregler chris.bregler@cs.nyu.edu
New York University

This paper tackles the complex problem of visually matching people in similar pose but with different clothes, background, and other appearance changes. We achieve this with a novel method for learning a nonlinear embedding based on several extensions to the Neighborhood Component Analysis (NCA) framework. Our method is convolutional, enabling it to scale to realistically-sized images. By cheaply labeling the head and hands in large video databases through Amazon Mechanical Turk (a crowd-sourcing service), we can use the task of localizing the head and hands as a proxy for determining body pose. We apply our method to challenging real-world data and show that it can generalize beyond hand localization to infer a more general notion of body pose. We evaluate our method quantitatively against other embedding methods. We also demonstrate that real-world performance can be improved through the use of synthetic data.
Subject Area: Vision

T37 Implicitly Constrained Gaussian Process Regression for Monocular Non-Rigid Pose Estimation

Mathieu Salzmann salzmann@ttic.edu
Raquel Urtasun rurtasun@ttic.edu
TTI Chicago

Estimating 3D pose from monocular images is a highly ambiguous problem. Physical constraints can be exploited to restrict the space of feasible configurations. In this paper we propose an approach to constraining the prediction of a discriminative predictor. We first show that the mean prediction of a Gaussian process implicitly satisfies linear constraints if those constraints are satisfied by the training examples. We then show how, by performing a change of variables, a GP can be forced to satisfy quadratic constraints. As evidenced by the experiments, our method outperforms state-of-the-art approaches on the tasks of rigid and non-rigid pose estimation.
Subject Area: Vision

T38 Large-Scale Matrix Factorization with Missing Data under Additional Constraints

Kaushik Mitra kmitra@umiacs.umd.edu
Rama Chellappa Rama@umiacs.umd.edu
University of Maryland College Park
Sameer Sheorey ssameer@ttic.edu
TTIC

Matrix factorization in the presence of missing data is at the core of many computer vision problems such as structure from motion (SfM), non-rigid SfM and photometric stereo. We formulate the problem of matrix factorization with missing data as a low-rank semidefinite program (LRSDP) with the advantage that: 1) an efficient quasi-Newton implementation of the LRSDP enables us to solve large-scale factorization problems, and 2) additional constraints such as ortho-normality, required in orthographic SfM, can be directly incorporated in the new formulation. Our empirical evaluations suggest that, under the conditions of matrix completion theory, the proposed algorithm finds the optimal solution, and also requires fewer observations compared to the current state-of-the-art algorithms. We further demonstrate the effectiveness of the proposed algorithm in solving the affine SfM problem, non-rigid SfM and photometric stereo problems.
Subject Area: Vision

T39 (RF)^2 – Random Forest Random Field

Nadia C Payet payetn@onid.orst.edu
Sinisa Todorovic sinisa@eecs.oregonstate.edu
Oregon State University

We combine random forest (RF) and conditional random field (CRF) into a new computational framework, called random forest random field (RF)^2. Inference of (RF)^2 uses the Swendsen-Wang cut algorithm, characterized by Metropolis-Hastings jumps. A jump from one state to another depends on the ratio of the proposal distributions, and on the ratio of the posterior distributions of the two states. Prior work typically resorts to a parametric estimation of these four distributions, and then computes their ratio. Our key idea is to instead directly estimate these ratios using RF. RF collects in leaf nodes of each decision tree the class histograms of training examples. We use these class histograms for a non-parametric estimation of the distribution ratios. We derive the theoretical error bounds of a two-class (RF)^2. (RF)^2 is applied to a challenging task of multiclass object recognition and segmentation over a random field of input image regions. In our empirical evaluation, we use only the visual information provided by image regions (e.g., color, texture, spatial layout), whereas the competing methods additionally use higher-level cues about the horizon location and 3D layout of surfaces in the scene. Nevertheless, (RF)^2 outperforms the state of the art on benchmark datasets, in terms of accuracy and computation time.
Subject Area: Vision

T40 Kernel Descriptors for Visual Recognition

Liefeng Bo lfb@cs.washington.edu
Dieter Fox fox@cs.washington.edu
University of Washington
Xiaofeng Ren xiaofeng.ren@intel.com
Intel

The design of low-level image features is critical for computer vision algorithms. Orientation histograms, such as those in SIFT and HOG, are the most successful and popular features for visual object and scene recognition. We highlight the kernel view of orientation histograms, and show that they are equivalent to a certain type of match kernels over image patches. This novel view allows us to design a family of kernel descriptors which provide a unified and principled framework to turn pixel attributes (gradient, color, local binary pattern, etc.) into compact patch-level features. In particular, we introduce three types of match kernels to measure similarities between image patches, and construct compact low-dimensional kernel descriptors from these match kernels using kernel principal component analysis (KPCA). Kernel descriptors are easy to design and can turn any type of pixel attribute into patch-level features. They outperform carefully tuned and sophisticated features including SIFT and deep belief networks. We report superior performance on standard image classification benchmarks: Scene-15, Caltech-101, CIFAR10 and CIFAR10-ImageNet.
Subject Area: Vision
Spotlight presentation, Tuesday, 11:50.

T41 Exploiting Weakly-labeled Web Images to Improve Object Classification: A Domain Adaptation Approach

Alessandro Bergamo aleb@cs.dartmouth.edu
Lorenzo Torresani lorenzo@cs.dartmouth.edu
Dartmouth College

Most current image categorization methods require large collections of manually annotated training examples to learn accurate visual recognition models. The time-consuming human labeling effort effectively limits these approaches to recognition problems involving a small number of different object classes. In order to address this shortcoming, in recent years several authors have proposed to learn object classifiers from weakly-labeled Internet images, such as photos retrieved by keyword-based image search engines. While this strategy eliminates the need for human supervision, the recognition accuracies of these methods are considerably lower than those obtained with fully-supervised approaches, because of the noisy nature of the labels associated to Web data. In this paper we investigate and compare methods that learn image classifiers by combining very few manually annotated examples (e.g. 1-10 images per class) and a large number of weakly-labeled Web photos retrieved using keyword-based image search. We cast this as a domain adaptation problem: given a few strongly-labeled examples in a target domain (the manually annotated examples) and many source domain examples (the weakly-labeled Web photos) learn classifiers yielding small generalization error on the target domain. Our experiments demonstrate that for the same number of strongly-labeled examples our domain adaptation approach produces significant recognition rate improvements over the best published results (e.g. 65% better when using 5 labeled training examples per class) and that our classifiers are one order of magnitude faster to learn and to evaluate than the best competing method despite our use of large weakly-labeled data sets.
Subject Area: Vision

T42 A Bayesian Framework for Figure-Ground Interpretation

Vicky Froyen vickyf@eden.rutgers.edu
Jacob Feldman jacob@ruccs.rutgers.edu
Manish Singh manish@ruccs.rutgers.edu
Rutgers University

Figure/ground assignment, in which the visual image is divided into nearer (figural) and farther (ground) surfaces, is an essential step in visual processing, but its underlying computational mechanisms are poorly understood. Figural assignment (often referred to as border ownership) can vary along a contour, suggesting a spatially distributed process whereby local and global cues are combined to yield local estimates of border ownership. In this paper we model figure/ground estimation in a Bayesian belief network, attempting to capture the propagation of border ownership across the image as local cues (contour curvature and T-junctions) interact with more global cues to yield a figure/ground assignment. Our network includes as a nonlocal factor skeletal (medial axis) structure, under the hypothesis that medial structure “draws” border

ownership so that borders are owned by their interiors. We also briefly present a psychophysical experiment in which we measured local border ownership along a contour at various distances from an inducing cue (a T-junction). Both the human subjects and the network show similar patterns of performance, converging rapidly to a similar pattern of spatial variation in border ownership along contours.
Subject Area: Vision

T43 Semi-Supervised Learning with Adversarially Missing Label Information

Umar Syed and Ben Taskar, University of Pennsylvania.
Subject Area: Unsupervised & Semi-supervised Learning
Oral presentation, Tuesday, 4:20pm.
See abstract, page 49

T44 Self-Paced Learning for Latent Variable Models

M. Pawan Kumar pawan@cs.stanford.edu
Benjamin Packer bpacker@cs.stanford.edu
Daphne Koller koller@cs.stanford.edu
Stanford University

Latent variable models are a powerful tool for addressing several tasks in machine learning. However the algorithms for learning the parameters of latent variable models are prone to getting stuck in a bad local optimum. To alleviate this problem we build on the intuition that rather than considering all samples simultaneously the algorithm should be presented with the training data in a meaningful order that facilitates learning. The order of the samples is determined by how easy they are. The main challenge is that often we are not provided with a readily computable measure of the easiness of samples. We address this issue by proposing a novel iterative self-paced learning algorithm where each iteration simultaneously selects easy samples and learns a new parameter vector. The number of samples selected is governed by a weight that is annealed until the entire training data has been considered. We empirically demonstrate that the self-paced learning algorithm outperforms the state of the art method for learning a latent structural SVM on four applications: object localization noun phrase coreference motif finding and handwritten digit recognition.
Subject Area: Unsupervised & Semi-supervised Learning

T45 Random Projections for k-means Clustering

Christos Boutsidis boutsc@cs.rpi.edu
Petros Drineas drinep@cs.rpi.edu
Rensselaer Polytechnic Institute
Anastasios Zouzias zouzias@cs.toronto.edu
University of Toronto

This paper discusses the topic of dimensionality reduction for k-means clustering. We prove that any set of n points in d dimensions (rows in a matrix $A \in R^{n\times d}$) can be projected into $t = \Omega(k/\in^2)$ dimensions for any $\varepsilon \in (01/3)$ in $O(nd[\varepsilon^{-2}k/\log(d)])$ time such that with constant probability the optimal k-partition of the point set is preserved within a factor of 2 + ε. The projection is done by post-multiplying A with a $d \times t$ random matrix R having entries $+1/\sqrt{t}$ or $-1/\sqrt{t}$ with equal probability. A numerical implementation of our technique and experiments on a large face images dataset verify the speed and the accuracy of our theoretical results.
Subject Area: Unsupervised & Semi-supervised Learning

T46 Discriminative Clustering by Regularized Information Maximization

Ryan G Gomes gomes@caltech.edu
Andreas Krause krausea@caltech.edu
Pietro Perona perona@caltech.edu
Caltech

Is there a principled way to learn a probabilistic discriminative classifier from an unlabeled data set? We present a framework that simultaneously clusters the data and trains a discriminative classifier. We call it Regularized Information Maximization (RIM). RIM optimizes an intuitive information-theoretic objective function which balances class separation, class balance and classifier complexity. The approach can flexibly incorporate different likelihood functions, express prior assumptions about the relative size of different classes and incorporate partial labels for semi-supervised learning. In particular, we instantiate the framework to unsupervised, multi-class kernelized logistic regression. Our empirical evaluation indicates that RIM outperforms existing methods on several real data sets, and demonstrates that RIM is an effective model selection method.
Subject Area: Unsupervised & Semi-supervised Learning

T47 Transduction with Matrix Completion: Three Birds with One Stone

Andrew B Goldberg goldberg@cs.wisc.edu
Xiaojin (Jerry) Zhu jerryzhu@cs.wisc.edu
Ben Recht brecht@cs.wisc.edu
Junming Xu xujm@cs.wisc.edu
Rob Nowak nowak@ece.wisc.edu
University of Wisconsin-Madison

We pose transductive classification as a matrix completion problem. By assuming the underlying matrix has a low rank, our formulation is able to handle three problems simultaneously: i) multi-label learning, where each item has more than one label, ii) transduction, where most of these labels are unspecified, and iii) missing data, where a large number of features are missing. We obtained satisfactory results on several real-world tasks, suggesting that the low rank assumption may not be as restrictive as it seems. Our method allows for different loss functions to apply on the feature and label entries of the matrix. The resulting nuclear norm minimization problem is solved with a modified fixed-point continuation method that is guaranteed to find the global optimum.
Subject Area: Unsupervised & Semi-supervised Learning

T48 Tiled convolutional neural networks

Quoc V Le quocle@stanford.edu
Jiquan Ngiam jngiam@cs.stanford.edu
Zhenghao Chen zhenghao@stanford.edu
Daniel Jin hao Chia danchia@stanford.edu
Pang Wei Koh pangwei@stanford.edu
Andrew Ng ang@cs.stanford.edu
Stanford University

Convolutional neural networks (CNNs) have been successfully applied to many tasks such as digit and object recognition. Using convolutional (tied) weights significantly reduces the number of parameters that have to be learned, and also allows translational invariance to be hard-coded into the architecture. In this paper, we consider the problem of learning invariances, rather than relying on hard-coding. We propose tiled convolution neural networks (Tiled CNNs), which use a regular "tiled" pattern of tied weights that does not require that adjacent hidden units share identical weights, but instead requires only that hidden units k steps away from each other to have tied weights. By pooling over neighboring units, this architecture is able to learn complex invariances (such as scale and rotational invariance) beyond translational invariance. Further, it also enjoys much of CNNs' advantage of having a relatively small number of learned parameters (such as ease of learning and greater scalability). We provide an efficient learning algorithm for Tiled CNNs based on Topographic ICA and show that learning complex invariant features allows us to achieve highly competitive results for both the NORB and CIFAR-10 datasets.
Subject Area: Unsupervised & Semi-supervised Learning

T49 Multi-View Active Learning in the Non-Realizable Case

Wei Wang wangw@lamda.nju.edu.cn
Zhi-Hua Zhou zhouzh@lamda.nju.edu.cn
Nanjing University

The sample complexity of active learning under the realizability assumption has been well-studied. The realizability assumption, however, rarely holds in practice. In this paper, we theoretically characterize the sample complexity of active learning in the nonrealizable case under multi-view setting. We prove that, with unbounded Tsybakov noise, the sample complexity of multi-view active learning can be $\tilde{O}(\log 1/\varepsilon)$ contrasting to singleview setting where the polynomial improvement is the best possible achievement. We also prove that in general multi-view setting the sample complexity of active learning with unbounded Tsybakov noise is $\tilde{O}(\log 1/\varepsilon)$ where the order of $1/\varepsilon$ is independent of the parameter in Tsybakov noise contrasting to previous polynomial bounds where the order of $1/\varepsilon$ is related to the parameter in Tsybakov noise.
Subject Area: Unsupervised & Semi-supervised Learning

T50 Near-Optimal Bayesian Active Learning with Noisy Observations

Daniel Golovin dgolovin@caltech.edu
Andreas Krause krausea@caltech.edu
Debajyoti Ray dray@caltech.edu
Caltech

We tackle the fundamental problem of Bayesian active learning with noise, where we need to adaptively select from a number of expensive tests in order to identify an unknown hypothesis sampled from a known prior distribution. In the case of noise-free observations, a greedy algorithm called generalized binary search (GBS) is known to perform near-optimally. We show that if the observations are noisy, perhaps surprisingly, GBS can perform very poorly. We develop EC2, a novel, greedy active learning algorithm and prove that it is competitive with the optimal policy, thus obtaining the first competitiveness guarantees for Bayesian active learning with noisy observations. Our bounds rely on a recently discovered diminishing returns property called adaptive submodularity, generalizing the classical notion of submodular set functions to adaptive policies. Our results hold even if the tests have non–uniform cost and their noise is correlated. We also propose EffECXtive, a particularly fast approximation of EC2, and evaluate it on a Bayesian experimental design problem involving human subjects, intended to tease apart competing economic theories of how people make decisions under uncertainty.
Subject Area: Unsupervised and Semi-supervised Learning

T51 Hashing Hyperplane Queries to Near Points with Applications to Large-Scale Active Learning

Prateek Jain pjain9@gmail.com
Microsoft Research India Lab
Sudheendra Vijayanarasimhan svnaras@cs.utexas.edu
Kristen Grauman grauman@cs.utexas.edu
University of Texas at Austin

We consider the problem of retrieving the database points nearest to a given hyperplane query without exhaustively scanning the database. We propose two hashing-based solutions. Our first approach maps the data to two-bit binary keys that are locality-sensitive for the angle between the hyperplane normal and a database point. Our second approach embeds the data into a vector space where the Euclidean norm reflects the desired distance between the original points and hyperplane query. Both use hashing to retrieve near points in sub-linear time. Our first method's preprocessing stage is more efficient, while the second has stronger accuracy guarantees. We apply both to pool-based active learning: taking the current hyperplane classifier as a query, our algorithm identifies those points (approximately) satisfying the well-known minimal distance-to-hyperplane selection criterion. We empirically demonstrate our methods' tradeoffs, and show that they make it practical to perform active selection with millions of unlabeled points.
Subject Area: Unsupervised & Semi-supervised Learning

T52 Unsupervised Kernel Dimension Reduction

Meihong Wang	meihongw@usc.edu
Fei Sha	feisha@usc.edu
University of Southern California	
Michael I Jordan	jordan@cs.berkeley.edu
UC Berkeley	

We apply the framework of kernel dimension reduction, originally designed for supervised problems, to unsupervised dimensionality reduction. In this framework, kernel-based measures of independence are used to derive low-dimensional representations that maximally capture information in covariates in order to predict responses. We extend this idea and develop similarly motivated measures for unsupervised problems where covariates and responses are the same. Our empirical studies show that the resulting compact representation yields meaningful and appealing visualization and clustering of data. Furthermore, when used in conjunction with supervised learners for classification, our methods lead to lower classification errors than state-of-the-art methods, especially when embedding data in spaces of very few dimensions.

Subject Area: Unsupervised & Semi-supervised Learning

T53 Large Margin Multi-Task Metric Learning

Shibin Parameswaran	sparames@ucsd.edu
UCSD	
Kilian Q Weinberger	kilian@wustl.edu
Washington University in St. Louis	

Multi-task learning (MTL) improves the prediction performance on multiple, different but related, learning problems through shared parameters or representations. One of the most prominent multi-task learning algorithms is an extension to svms by Evgeniou et al. Although very elegant, multi-task svm is inherently restricted by the fact that support vector machines require each class to be addressed explicitly with its own weight vector which, in a multi-task setting, requires the different learning tasks to share the same set of classes. This paper proposes an alternative formulation for multi-task learning by extending the recently published large margin nearest neighbor (lmnn) algorithm to the MTL paradigm. Instead of relying on separating hyperplanes, its decision function is based on the nearest neighbor rule which inherently extends to many classes and becomes a natural fit for multitask learning. We evaluate the resulting multi-task lmnn on realworld insurance data and speech classification problems and show that it consistently outperforms single-task kNN under several metrics and state-of-the-art MTL classifiers.

Subject Area: Supervised Learning

T54 Deep Coding Network

Yuanqing Lin	ylin@sv.nec-labs.com
Kai Yu	kyu@sv.nec-labs.com
Shenghuo Zhu	zsh@sv.nec-labs.com
NEC Laboratories America	
Zhang Tong	tzhang@stat.rutgers.edu
Rutgers University	

This paper proposes a principled extension of the traditional single-layer flat sparse coding scheme, where a two-layer coding scheme is derived based on theoretical analysis of nonlinear functional approximation that extends recent results for local coordinate coding. The two-layer approach can be easily generalized to deeper structures in a hierarchical multiple-layer manner. Empirically it is shown that the deep coding approach yields improved performance in benchmark datasets.

Subject Area: Unsupervised & Semi-supervised Learning

T55 Inductive Regularized Learning of Kernel Functions

Prateek Jain	pjain9@gmail.com
Microsoft Research India Lab	
Brian Kulis	brian.kulis@gmail.com
UC Berkeley	
Inderjit Dhillon	inderjit@cs.utexas.edu
University of Texas	

In this paper we consider the fundamental problem of semi-supervised kernel function learning. We propose a general regularized framework for learning a kernel matrix, and then demonstrate an equivalence between our proposed kernel matrix learning framework and a general linear transformation learning problem. Our result shows that the learned kernel matrices parameterize a linear transformation kernel function and can be applied inductively to new data points. Furthermore, our result gives a constructive method for kernelizing most existing Mahalanobis metric learning formulations. To make our results practical for large-scale data, we modify our framework to limit the number of parameters in the optimization process. We also consider the problem of kernelized inductive dimensionality reduction in the semi-supervised setting. We introduce a novel method for this problem by considering a special case of our general kernel learning framework where we select the trace norm function as the regularizer. We empirically demonstrate that our framework learns useful kernel functions, improving the k-NN classification accuracy significantly in a variety of domains. Furthermore, our kernelized dimensionality reduction technique significantly reduces the dimensionality of the feature space while achieving competitive classification accuracies.

Subject Area: Unsupervised & Semi-supervised Learning
Spotlight presentation, Tuesday, 5:50.

T56 Learning concept graphs from text with stick-breaking priors

America Chambers	ahollowa@ics.uci.edu
Padhraic Smyth	smyth@ics.uci.edu
Mark Steyvers	mark.steyvers@uci.edu

University of California-Irvine

We present a generative probabilistic model for learning general graph structures, which we term concept graphs, from text. Concept graphs provide a visual summary of the thematic content of a collection of documents-a task that is difficult to accomplish using only keyword search. The proposed model can learn different types of concept graph structures and is capable of utilizing partial prior knowledge about graph structure as well as labeled documents. We describe a generative model that is based on a stick-breaking process for graphs, and a Markov Chain Monte Carlo inference procedure. Experiments on simulated data show that the model can recover known graph structure when learning in both unsupervised and semi-supervised modes. We also show that the proposed model is competitive in terms of empirical log likelihood with existing structure-based topic models (such as hPAM and hLDA) on real-world text data sets. Finally, we illustrate the application of the model to the problem of updating Wikipedia category graphs.
Subject Area: Unsupervised & Semi-supervised Learning
Spotlight presentation, Tuesday, 3:25.

T57 Joint Analysis of Time-Evolving Binary Matrices and Associated Documents

Eric X Wang	ew28@duke.edu
Dehong Liu	liudh97@gmail.com
Jorge G Silva j	g.silva@duke.edu
David Dunson	dunson@stat.duke.edu
Lawrence Carin	lcarin@ee.duke.edu

Duke University

We consider problems for which one has incomplete binary matrices that evolve with time (e.g., the votes of legislators on particular legislation, with each year characterized by a different such matrix). An objective of such analysis is to infer structure and interrelationships underlying the matrices, here defined by latent features associated with each axis of the matrix. In addition, it is assumed that documents are available for the entities associated with at least one of the matrix axes. By jointly analyzing the matrices and documents, one may be used to inform the other within the analysis, and the model offers the opportunity to predict matrix values (e.g., votes) based only on an associated document (e.g., legislation). The research presented here merges two areas of machinelearning that have previously been investigated separately: incomplete-matrix analysis and topic modeling. The analysis is performed from a Bayesian perspective, with efficient inference constituted via Gibbs sampling. The framework is demonstrated by considering all voting data and available documents (legislation) during the 220-year lifetime of the United States Senate and House of Representatives.
Subject Area: Unsupervised & Semi-supervised Learning

T58 Predictive Subspace Learning for Multi-view Data: a Large Margin Approach

Ning Chen	ningchen@cs.cmu.edu
Jun Zhu	junzhu@cs.cmu.edu
Eric Xing	epxing@cs.cmu.edu

Carnegie Mellon University

Learning from multi-view data is important in many applications, such as image classification and annotation. In this paper, we present a large-margin learning framework to discover a predictive latent subspace representation shared by multiple views. Our approach is based on an undirected latent space Markov network that fulfills a weak conditional independence assumption that multi-view observations and response variables are independent given a set of latent variables. We provide efficient inference and parameter estimation methods for the latent subspace model. Finally we demonstrate the advantages of large-margin learning on real video and web image data for discovering predictive latent representations and improving the performance on image classification annotation and retrieval.
Subject Area: Unsupervised & Semi-supervised Learning

T59 LSTD with Random Projections

Mohammad Ghavamzadeh	mohammad.ghavamzadeh@inria.fr
Alessandro Lazaric	alessandro.lazaric@inria.fr
Lille Odalric Maillard	odalric.maillard@inria.fr
Remi Munos	remi.munos@inria.fr

INRIA Lille

We consider the problem of reinforcement learning in high-dimensional spaces when the number of features is bigger than the number of samples. In particular, we study the leastsquares temporal difference (LSTD) learning algorithm when a space of low dimension is generated with a random projection from a high-dimensional space. We provide a thorough theoretical analysis of the LSTD with random projections and derive performance bounds for the resulting algorithm. We also show how the error of LSTD with random projections is propagated through the iterations of a policy iteration algorithm and provide a performance bound for the resulting least-squares policy iteration (LSPI) algorithm.
Subject Area: Control and Reinforcement Learning
Spotlight presentation, Tuesday, 11:25.

T60 Constructing Skill Trees for Reinforcement Learning Agents from Demonstration Trajectories

George D Konidaris	gdk@cs.umass.edu
Scott R Kuindersma	scottk@cs.umass.edu
Andrew Barto	barto@cs.umass.edu
Roderic A Grupen	grupen@cs.umass.edu

University of Massachusetts Amherst

We introduce CST, an algorithm for constructing skill trees from demonstration trajectories in continuous reinforcement learning domains. CST uses a changepoint detection method to segment each trajectory into a skill chain by detecting a change of appropriate abstraction, or

that a segment is too complex to model as a single skill. The skill chains from each trajectory are then merged to form a skill tree. We demonstrate that CST constructs an appropriate skill tree that can be further refined through learning in a challenging continuous domain, and that it can be used to segment demonstration trajectories on a mobile manipulator into chains of skills where each skill is assigned an appropriate abstraction.
Subject Area: Control and Reinforcement Learning

T61 Avoiding False Positive in Multi-Instance Learning

Yanjun Han yanjun.han@ia.ac.cn
Qing Tao qing.tao@ia.ac.cn
Jue Wang jue.wang@ia.ac.cn
Institute of Automation CAS

In multi-instance learning, there are two kinds of prediction failure, i.e., false negative and false positive. Current research mainly focus on avoding the former. We attempt to utilize the geometric distribution of instances inside positive bags to avoid both the former and the latter. Based on kernel principal component analysis we define a projection constraint for each positive bag to classify its constituent instances far away from the separating hyperplane while place positive instances and negative instances at opposite sides. We apply the Constrained Concave-Convex Procedure to solve the resulted problem. Empirical results demonstrate that our approach offers improved generalization performance.
Subject Area: Supervised Learning

T62 A Theory of Multiclass Boosting

Indraneel Mukherjee and Robert E Schapire,
Princeton University.
Subject Area: Supervised Learning
See abstract, page 47

T63 Joint Cascade Optimization Using A Product Of Boosted Classifiers

Leonidas Lefakis llefakis@idiap.ch
Francois Fleuret francois.fleuret@idiap.ch
IDIAP Research Institute

The standard strategy for efficient object detection consists of building a cascade composed of several binary classifiers. The detection process takes the form of a lazy evaluation of the conjunction of the responses of these classifiers, and concentrates the computation on difficult parts of the image which can not be trivially rejected. We introduce a novel algorithm to construct jointly the classifiers of such a cascade. We interpret the response of a classifier as a probability of a positive prediction and the overall response of the cascade as the probability that all the predictions are positive. From this noisy-AND model we derive a consistent loss and a Boosting procedure to optimize that global probability on the training set. Such a joint learning allows the individual predictors to focus on a more restricted modeling problem and improves the performance compared to a standard cascade. We demonstrate the efficiency of this approach on face and pedestrian detection with standard data-sets and comparisons with reference baselines.
Subject Area: Supervised Learning

T64 Multiple Kernel Learning and the SMO Algorithm S.V.N.

Vishwanathan vishy@stat.purdue.edu
Zhaonan sun sunz@stat.purdue.edu
Nawanol T Ampornpunt ntheeraa@cs.purdue.edu
Purdue University
Manik Varma manik@microsoft.com
Microsoft Research

Our objective is to train p-norm Multiple Kernel Learning (MKL) and, more generally linear MKL regularised by the Bregman divergence using the Sequential Minimal Optimization (SMO) algorithm. The SMO algorithm is simple easy to implement and adapt and efficiently scales to large problems. As a result it has gained widespread acceptance and SVMs are routinely trained using SMO in diverse real world applications. Training using SMO has been a long standing goal in MKL for the very same reasons. Unfortunately the standard MKL dual is not differentiable and therefore can not be optimised using SMO style co-ordinate ascent. In this paper we demonstrate that linear MKL regularised with the p-norm squared or with certain Bregman divergences can indeed be trained using SMO. The resulting algorithm retains both simplicity and efficiency and is significantly faster than the state-of-the-art specialised p-norm MKL solvers. We show that we can train on a hundred thousand kernels in approximately seven minutes and on fifty thousand points in less than half an hour on a single core.
Subject Area: Supervised Learning
Spotlight presentation, Tuesday, 5:45.

T65 Relaxed Clipping: A Global Training Method for Robust Regression and Classification

Yaoliang Yu yaoliang@cs.ualberta.ca
Min Yang myang2@cs.ualberta.ca
University of Alberta
Linli Xu linlixu@ustc.edu.cn
University of Science and Technology of China
Martha White whitem@cs.ualberta.ca
Dale Schuurmans dale@cs.ualberta.ca
University of Alberta

Robust regression and classification are often thought to require non-convex loss functions that prevent scalable, global training. However, such a view neglects the possibility of reformulated training methods that can yield practically solvable alternatives. A natural way to make a loss function more robust to outliers is to truncate loss values that exceed a maximum threshold. We demonstrate that a relaxation of this form of “loss clipping” can be made globally solvable and applicable to any standard loss while guaranteeing robustness against outliers. We present a generic procedure that can be applied to standard loss functions and demonstrate improved robustness in regression and classification problems.
Subject Area: Supervised Learning

T66 Decomposing Isotonic Regression for Efficiently Solving Large Problems

Ronny Luss ronnyluss@gmail.com
Saharon Rosset saharon@post.tau.ac.il
Moni Shahar moni@eng.tau.ac.il
Tel Aviv University

A new algorithm for isotonic regression is presented based on recursively partitioning the solution space. We develop efficient methods for each partitioning subproblem through an equivalent representation as a network flow problem, and prove that this sequence of partitions converges to the global solution. These network flow problems can further be decomposed in order to solve very large problems. Success of isotonic regression in prediction and our algorithm's favorable computational properties are demonstrated through simulated examples as large as $2x10^5$ variables and 10^7 constraints.

Subject Area: Supervised Learning

T67 Factorized Latent Spaces with Structured Sparsity

Yangqing Jia jiayq@eecs.berkeley.edu
Trevor Darrell trevor@eecs.berkeley.edu
UC Berkeley
Mathieu Salzmann salzmann@ttic.edu
TTI Chicago

Recent approaches to multi-view learning have shown that factorizing the information into parts that are shared across all views and parts that are private to each view could effectively account for the dependencies and independencies between the different input modalities. Unfortunately, these approaches involve minimizing non-convex objective functions. In this paper, we propose an approach to learning such factorized representations inspired by sparse coding techniques. In particular, we show that structured sparsity allows us to address the multi-view learning problem by alternately solving two convex optimization problems. Furthermore, the resulting factorized latent spaces generalize over existing approaches in that they allow :having latent dimensions shared between any subset of the views instead of between all the views only. We show that our approach outperforms state-of-the-art methods on the task of human pose estimation.

Subject Area: Supervised Learning

T68 Evaluation of Rarity of Fingerprints in Forensics

Chang Su LEE changsu@buffalo.edu
Sargur N Srihari srihari@buffalo.edu
University at Buffalo

A method for computing the rarity of latent fingerprints represented by minutiae is given. It allows determining the probability of finding a match for an evidence print in a database of n known prints. The probability of random correspondence between evidence and database is determined in three procedural steps. In the registration step the latent print is aligned by finding its core point; which is done using a procedure based on a machine learning approach based on Gaussian processes. In the evidence probability evaluation step a generative model based on Bayesian networks is used to determine the probability of the evidence; it takes into account both the dependency of each minutia on nearby minutiae and the confidence of their presence in the evidence. In the specific probability of random correspondence step the evidence probability is used to determine the probability of match among n for a given tolerance; the last evaluation is similar to the birthday correspondence probability for a specific birthday. The generative model is validated using a goodness-of-fit test evaluated with a standard database of fingerprints. The probability of random correspondence for several latent fingerprints are evaluated for varying numbers of minutiae.

Subject Area: Probabilistic Models and Methods

T69 Structured Sparsity-Inducing Norms Through Submodular Functions

Francis Bach,
Ecole Normale Superieure.

Subject Area: Supervised Learning
See abstract, page 49

T70 Learning Convolutional Feature Hierarchies for Visual Recognition

Koray Kavukcuoglu koray@cs.nyu.edu
Pierre Sermanet sermanet@cs.nyu.edu
Y-Lan Boureau ylan@cs.nyu.edu
Karol Gregor kgregor@cs.nyu.edu
Michael Mathieu mmathieu@clipper.ens.fr
Yann Le Cun yann@cs.nyu.edu
New York University

We propose an unsupervised method for learning multi-stage hierarchies of sparse convolutional features. While sparse coding has become an increasingly popular method for learning visual features it is most often trained at the patch level. Applying the resulting filters convolutionally results in highly redundant codes because overlapping patches are encoded in isolation. By training convolutionally over large image windows our method reduces the redudancy between feature vectors at neighboring locations and improves the efficiency of the overall representation. In addition to a linear decoder that reconstructs the image from sparse features our method trains an efficient feed-forward encoder that predicts quasi-sparse features from the input. While patch-based training rarely produces anything but oriented edge detectors we show that convolutional training produces highly diverse filters including center-surround filters corner detectors cross detectors and oriented grating detectors. We show that using these filters in multi-stage convolutional network architecture improves performance on a number of visual recognition and detection tasks.

Subject Area: Supervised Learning
Spotlight presentation, Tuesday, 11:20.

T71 Probabilistic Multi-Task Feature Selection

Yu Zhang zhangyu@cse.ust.hk
Dit-Yan Yeung dyyeung@cse.ust.hk
Qian Xu fleurxq@ust.hk
Hong Kong University of Science and Technology

Recently, some variants of the l_1 norm, particularly matrix norms such as the $l_{1,2}$ and $l_{1,\infty}$ norms, have been widely used in multi-task learning, compressed sensing and other related areas to enforce sparsity via joint regularization. In this paper, we unify the $l_{1,2}$ and $l_{1,\infty}$ norms by considering a family of $l_{1\infty}$ norms for $1 < q \leq \infty$ and study the problem of determining the most appropriate sparsity enforcing norm to use in the context of multi-task feature selection. Using the generalized normal distribution, we provide a probabilistic interpretation of the general multi-task feature selection problem using the $l_{1,\infty}$ norm. Based on this probabilistic interpretation, we develop a probabilistic model using the noninformative Jeffreys prior. We also extend the model to learn and exploit more general types of pairwise relationships between tasks. For both versions of the model, we devise expectation-maximization (EM) algorithms to learn all model parameters, including q, automatically. Experiments have been conducted on two cancer classification applications using microarray gene expression data.
Subject Area: Supervised Learning

T72 Probabilistic Inference and Differential Privacy

Oliver Williams olliew@microsoft.com
Frank McSherry mcsherry@microsoft.com
Microsoft Research

We identify and investigate a strong connection between probabilistic inference and differential privacy the latter being a recent privacy definition that permits only indirect observation of data through noisy measurement. Previous research on differential privacy has focused on designing measurement processes whose output is likely to be useful on its own. We consider the potential of applying probabilistic inference to the measurements and measurement process to derive posterior distributions over the data sets and model parameters thereof. We find that probabilistic inference can improve accuracy integrate multiple observations measure uncertainty and even provide posterior distributions over quantities that were not directly measured.
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Tuesday, 12:05.

T73 Inter-time segment information sharing for non-homogeneous dynamic Bayesian networks

Dirk Husmeier dirk@bioss.ac.uk
Frank Dondelinger frankd@bioss.ac.uk
Biomathematics & Statistics Scotland (Bioss)
Sophie Lebre sophie.lebre@lsiit-cnrs.unistra.fr
University of Strasbourg

Conventional dynamic Bayesian networks (DBNs) are based on the homogeneous Markov assumption which is too restrictive in many practical applications. Various approaches to relax the homogeneity assumption have therefore been proposed in the last few years. The present paper aims to improve the flexibility of two recent versions of non-homogeneous DBNs which either (i) suffer from the need for data discretization or (ii) assume a timeinvariant network structure. Allowing the network structure to be fully flexible leads to the risk of overfitting and inflated inference uncertainty though especially in the highly topical field of systems biology where independent measurements tend to be sparse. In the present paper we investigate three conceptually different regularization schemes based on inter-segment information sharing. We assess the performance in a comparative evaluation study based on simulated data. We compare the predicted segmentation of gene expression time series obtained during embryogenesis in Drosophila melanogaster with other state-ofthe- art techniques. We conclude our evaluation with an application to synthetic biology where the objective is to predict a known regulatory network of five genes in Saccharomyces cerevisiae.
Subject Area: Probabilistic Models and Methods

T74 Variational Inference over Combinatorial Spaces

Alexandre Bouchard-Côté bouchard@cs.berkeley.edu
Michael I Jordan jordan@cs.berkeley.edu
University of California Berkeley

Since the discovery of sophisticated fully polynomial randomized algorithms for a range of #P problems (Karzanov et al., 1991; Jerrum et al., 2001; Wilson, 2004), theoretical work on approximate inference in combinatorial spaces has focused on Markov chain Monte Carlo methods. Despite their strong theoretical guarantees, the slow running time of many of these randomized algorithms and the restrictive assumptions on the potentials have hindered the applicability of these algorithms to machine learning. Because of this, in applications to combinatorial spaces simple exact models are often preferred to more complex models that require approximate inference (Siepel et al., 2004). Variational inference would appear to provide an appealing alternative, given the success of variational methods for graphical models (Wainwright et al., 2008); unfortunately, however, it is not obvious how to develop variational approximations for combinatorial objects such as matchings, partial orders, plane partitions and sequence alignments. We propose a new framework that extends variational inference to a wide range of combinatorial spaces. Our method is based on a simple assumption: the existence of a tractable measure factorization, which we show holds in many examples. Simulations on a range of matching models show that the algorithm is more general and empirically faster than a popular fully polynomial randomized algorithm. We also apply the framework to the problem of multiple alignment of protein sequences obtaining state-of-the-art results on the BAliBASE dataset (Thompson et al. 1999).
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Tuesday, 5:40.

T75 Worst-case bounds on the quality of max-product fixed-points

Meritxell Vinyals meritxell@iiia.csic.es
Jes´us Cerquides cerquide@iiia.csic.es
Juan Antonio Rodrıguez-Aguilar jar@iiia.csic.es
IIIA-CSIC
Alessandro Farinelli alessandro.farinelli@univr.it
University of Verona

We study worst-case bounds on the quality of any fixed point assignment of the max-product algorithm for Markov Random Fields (MRF). We start proving a bound independent of the MRF structure and parameters. Afterwards we show how this bound can be improved for MRFs with particular structures such as bipartite graphs or grids. Our results provide interesting insight into the behavior of max-product. For example we prove that max-product provides very good results (at least 90% of the optimal) on MRFs with large variable-disjoint cycles (MRFs in which all cycles are variable-disjoint namely that they do not share any edge and in which each cycle contains at least 20 variables).
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Tuesday, 5:05

T76 Improving the Asymptotic Performance of Markov Chain Monte-Carlo by Inserting Vortices

Yi Sun yi@idsia.ch
Faustino Gomez tino@idsia.ch
Juergen Schmidhuber juergen@idsia.ch
IDSIA

We present a new way of converting a reversible finite Markov chain into a nonreversible one, with a theoretical guarantee that the asymptotic variance of the MCMC estimator based on the non-reversible chain is reduced. The method is applicable to any reversible chain whose states are not connected through a tree, and can be interpreted graphically as inserting vortices into the state transition graph. Our result confirms that non-reversible chains are fundamentally better than reversible ones in terms of asymptotic performance, and suggests interesting directions for further improving MCMC.
Subject Area: Probabilistic Models and Methods

T77 Gaussian sampling by local perturbations

George Papandreou gpapan@stat.ucla.edu
Alan L Yuille yuille@stat.ucla.edu
UCLA

We present a technique for exact simulation of Gaussian Markov random fields (GMRFs), which can be interpreted as locally injecting noise to each Gaussian factor independently, followed by computing the mean/mode of the perturbed GMRF. Coupled with standard iterative techniques for the solution of symmetric positive definite systems, this yields a very efficient sampling algorithm with essentially linear complexity in terms of speed and memory requirements, well suited to extremely large scale probabilistic models. Apart from synthesizing data under a Gaussian model, the proposed technique directly leads to an efficient unbiased estimator of marginal variances. Beyond Gaussian models, the proposed algorithm is also very useful for handling highly non-Gaussian continuously-valued MRFs such as those arising in statistical image modeling or in the first layer of deep belief networks describing real-valued data, where the non-quadratic potentials coupling different sites can be represented as finite or infinite mixtures of Gaussians with the help of local or distributed latent mixture assignment variables. The Bayesian treatment of such models most naturally involves a block Gibbs sampler which alternately draws samples of the conditionally independent latent mixture assignments and the conditionally multivariate Gaussian continuous vector and we show that it can directly benefit from the proposed methods.
Subject Area: Probabilistic Models and Methods

T78 Approximate Inference by Compilation to Arithmetic Circuits

Daniel Lowd lowd@cs.uoregon.edu
University of Oregon
Pedro Domingos pedrod@cs.washington.edu
University of Washington

Arithmetic circuits (ACs) exploit context-specific independence and determinism to allow exact inference even in networks with high treewidth. In this paper, we introduce the first ever approximate inference methods using ACs, for domains where exact inference remains intractable. We propose and evaluate a variety of techniques based on exact compilation, forward sampling, AC structure learning, Markov network parameter learning, variational inference, and Gibbs sampling. In experiments on eight challenging real-world domains, we find that the methods based on sampling and learning work best: one such method (AC2-F) is faster and usually more accurate than loopy belief propagation, mean field, and Gibbs sampling; another (AC2-G) has a running time similar to Gibbs sampling but is consistently more accurate than all baselines.
Subject Area: Probabilistic Models and Methods

T79 MAP estimation in Binary MRFs via Bipartite Multi-cuts

Sashank Jakkam Reddi, Sunita Sarawagi and Sundar Vishwanathan, IIT Bombay.
Subject Area: Probabilistic Models and Methods
Oral presentation, Tuesday, 4:40pm.
See abstract, page 50

T80 Improvements to the Sequence Memoizer

Jan Gasthaus j.gasthaus@gatsby.ucl.ac.uk
Yee Whye Teh ywteh@gatsby.ucl.ac.uk
Gatsby Unit UCL

The sequence memoizer is a model for sequence data with state-of-the-art performance on language modeling and compression. We propose a number of improvements to the model and inference algorithm, including an enlarged range of hyperparameters, a memoryefficient representation, and inference algorithms operating on the new representation. Our derivations are based on precise definitions of the various processes that will also allow us to provide an elementary proof of the ”mysterious” coagulation and fragmentation properties used in the original paper on the sequence memoizer by Wood et al. (2009). We present some experimental results supporting our improvements. Subject Area: Probabilistic Models and Methods

T81 Probabilistic Deterministic Infinite Automata

David Pfau pfau@neurotheory.columbia.edu
Nicholas Bartlett bartlett@stat.columbia.edu
Frank Wood fwood@stat.columbia.edu
Columbia University

We propose a novel Bayesian nonparametric approach to learning with probabilistic deterministic finite automata (PDFA). We define and develop and sampler for a PDFA with an infinite number of states which we call the probabilistic deterministic infinite automata (PDIA). Posterior predictive inference in this model, given a finite training sequence, can be interpreted as averaging over multiple PDFAs of varying structure, where each PDFA is biased towards having few states. We suggest that our method for averaging over PDFAs is a novel approach to predictive distribution smoothing. We test PDIA inference both on PDFA structure learning and on both natural language and DNA data prediction tasks. The results suggest that the PDIA presents an attractive compromise between the computational cost of hidden Markov models and the storage requirements of hierarchically smoothed Markov models.
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Tuesday, 5:55.

T82 Copula Processes

Andrew G Wilson agw38@cam.ac.uk
University of Cambridge
Zoubin Ghahramani zoubin@eng.cam.ac.uk
Cambridge

We define a copula process which describes the dependencies between arbitrarily many random variables independently of their marginal distributions. As an example, we develop a stochastic volatility model, Gaussian Copula Process Volatility (GCPV), to predict the latent standard deviations of a sequence of random variables. To make predictions we use Bayesian inference, with the Laplace approximation, and with Markov chain Monte Carlo as an alternative. We find our model can outperform GARCH on simulated and financial data. And unlike GARCH, GCPV can easily handle missing data, incorporate covariates other than time, and model a rich class of covariance structures.
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Tuesday, 11:10.

T83 Learning sparse dynamic linear systems using stable spline kernels and exponential hyperpriors

Alessandro Chiuso chiuso@dei.unipd.it
Gianluigi Pillonetto giapi@dei.unipd.it
University of Padova

We introduce a new Bayesian nonparametric approach to identification of sparse dynamic linear systems. The impulse responses are modeled as Gaussian processes whose autocovariances encode the BIBO stability constraint, as defined by the recently introduced “Stable Spline kernel”. Sparse solutions are obtained by placing exponential hyperpriors on the scale factors of such kernels. Numerical experiments regarding estimation of ARMAX models show that this technique provides a definite advantage over a group LAR algorithm and state-of-the-art parametric identification techniques based on prediction error minimization.
Subject Area: Probabilistic Models and Methods

T84 Exact learning curves for Gaussian process regression on large random graphs

Matthew J Urry matthew.urry@kcl.ac.uk
Peter Sollich peter.sollich@kcl.ac.uk
Kings College London

We study learning curves for Gaussian process regression which characterise performance in terms of the Bayes error averaged over datasets of a given size. Whilst learning curves are in general very difficult to calculate we show that for discrete input domains, where similarity between input points is characterised in terms of a graph, accurate predictions can be obtained. These should in fact become exact for large graphs drawn from a broad range of random graph ensembles with arbitrary degree distributions where each input (node) is connected only to a finite number of others. The method is based on translating the appropriate belief propagation equations to the graph ensemble. We demonstrate the accuracy of the predictions for Poisson (Erdos-Renyi) and regular random graphs, and discuss when and why previous approximations to the learning curve fail.
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Tuesday, 3:15.

T85 Subgraph Detection Using Eigenvector L1 Norms

Benjamin A Miller bamiller@ll.mit.edu
Nadya T Bliss nt@ll.mit.edu MIT
Lincoln Laboratory
Patrick Wolfe wolfe@stat.harvard.edu
Harvard University

When working with network datasets, the theoretical framework of detection theory for Euclidean vector spaces no longer applies. Nevertheless, it is desirable to determine the detectability of small, anomalous graphs embedded into background networks with known statistical properties. Casting the problem of subgraph detection in a signal processing context, this article provides a framework and empirical results that elucidate a ”detection theory” for graph-valued data. Its focus is the detection of anomalies in unweighted, undirected graphs through L1 properties of the eigenvectors of the graph’s so-called modularity matrix. This metric is observed to have relatively low variance for certain categories of randomly-generated graphs and to reveal the presence of an anomalous subgraph with reasonable reliability when the anomaly is not well-correlated with stronger portions of the background graph. An analysis of subgraphs in real network datasets confirms the efficacy of this approach.
Subject Area: Probabilistic Models and Methods

T86 Gaussian Process Preference Elicitation

Edwin V Bonilla edwin.bonilla@nicta.com.au
Shengbo Guo Shengbo.Guo@nicta.com.au
Scott Sanner Scott.Sanner@nicta.com.au
NICTA

Bayesian approaches to preference elicitation (PE) are particularly attractive due to their ability to explicitly model uncertainty in users’ latent utility functions. However, previous approaches to Bayesian PE have ignored the important problem of generalizing from previous users to an unseen user in order to reduce the elicitation burden on new users. In this paper, we address this deficiency by introducing a Gaussian Process (GP) prior over users’ latent utility functions on the joint space of user and item features. We learn the hyper-parameters of this GP on a set of preferences of previous users and use it to aid in the elicitation process for a new user. This approach provides a flexible model of a multi-user utility function, facilitates an efficient value of information (VOI) heuristic query selection strategy, and provides a principled way to incorporate the elicitations of multiple users back into the model. We show the effectiveness of our method in comparison to previous work on a real dataset of user preferences over sushi types.
Subject Area: Probabilistic Models and Methods

T87 Implicit Differentiation by Perturbation

Justin Domke justin.domke@rit.edu
Rochester Institute of Technology

This paper proposes a simple and efficient finite difference method for implicit differentiation of marginal inference results in discrete graphical models. Given an arbitrary loss function, defined on marginals, we show that the derivatives of this loss with respect to model parameters can be obtained by running the inference procedure twice, on slightly perturbed model parameters. This method can be used with approximate inference, with a loss function over approximate marginals. Convenient choices of loss functions make it practical to fit graphical models with hidden variables, high treewidth and/or model misspecification.
Subject Area: Probabilistic Models and Methods

T88 A Primal-Dual Message-Passing Algorithm for Approximated Large Scale Structured Prediction

Tamir Hazan tamir@ttic.edu
Raquel Urtasun rurtasun@ttic.edu
TTI Chicago

In this paper we propose an approximated learning framework for large scale graphical models and derive message passing algorithms for learning their parameters efficiently. We first relate CRFs and structured SVMs and show that in the CRF’s primal a variant of the log-partition function known as soft-max smoothly approximates the hinge loss function of structured SVMs. We then propose an intuitive approximation for structured prediction problems using Fenchel duality based on a local entropy approximation that computes the exact gradients of the approximated problem and is guaranteed to converge. Unlike existing approaches this allow us to learn graphical models with cycles and very large number of parameters efficiently. We demonstrate the effectiveness of our approach in an image denoising task. This task was previously solved by sharing parameters across cliques. In contrast our algorithm is able to efficiently learn large number of parameters resulting in orders of magnitude better prediction.
Subject Area: Probabilistic Models and Methods

T89 Extended Bayesian Information Criteria for Gaussian Graphical Models

Foygel Rina rina@uchicago.edu
Mathias Drton drton@uchicago.edu
University of Chicago

Gaussian graphical models with sparsity in the inverse covariance matrix are of significant interest in many modern applications. For the problem of recovering the graphical structure, information criteria provide useful optimization objectives for algorithms searching through sets of graphs or for selection of tuning parameters of other methods such as the graphical lasso, which is a likelihood penalization technique. In this paper we establish the asymptotic consistency of an extended Bayesian information criterion for Gaussian graphical models in a scenario where both the number of variables p and the sample size n grow. Compared to earlier work on the regression case, our treatment allows for growth in the number of non-zero parameters in the true model, which is necessary in order to cover connected graphs. We demonstrate the performance of this criterion on simulated data when used in conjuction with the graphical lasso, and verify that the criterion indeed

performs better than either cross-validation or the ordinary Bayesian information criterion when p and the number of non-zero parameters q both scale with n.
Subject Area: Probabilistic Models and Methods

T90 Causal discovery in multiple models from different experiments

Tom Claassen tomc@cs.ru.nl
Tom Heskes tomh@cs.ru.nl
Radboud University Nijmegen

A long-standing open research problem is how to use information from different experiments, including background knowledge, to infer causal relations. Recent developments have shown ways to use multiple data sets provided they originate from identical experiments. We present the MCI-algorithm as the first method that can infer provably valid causal relations in the large sample limit from different experiments. It is fast reliable and produces very clear and easily interpretable output. It is based on a result that shows that constraint-based causal discovery is decomposable into a candidate pair identification and subsequent elimination step that can be applied separately from different models. We test the algorithm on a variety of synthetic input model sets to assess its behavior and the quality of the output. The method shows promising signs that it can be adapted to suit causal discovery in real-world application areas as well including large databases.
Subject Area: Probabilistic Models and Methods

T91 Lifted Inference Seen from the Other Side : The Tractable Features

Abhay Jha abhaykj@cs.washington.edu
Vibhav G Gogate vgogate@cs.washington.edu
Alexandra Meliou ameli@cs.washington.edu
Dan Suciu suciu@cs.washington.edu
University of Washington

Lifted inference algorithms for representations that combine first-order logic and probabilistic graphical models have been the focus of much recent research. All lifted algorithms developed to date are based on the same underlying idea: take a standard probabilistic inference algorithm (e.g., variable elimination, belief propagation etc.) and improve its efficiency by exploiting repeated structure in the first-order model. In this paper, we propose an approach from the other side in that we use techniques from logic for probabilistic inference. In particular, we define a set of rules that look only at the logical representation to identify models for which exact efficient inference is possible. We show that our rules yield several new tractable classes that cannot be solved efficiently by any of the existing techniques.
Subject Area: Probabilistic Models and Methods

T92 Movement extraction by detecting dynamics switches and repetitions

Silvia Chiappa silvia@statslab.cam.ac.uk
Cambridge University
Jan Peters jan.peters@tuebingen.mpg.de
MPI for biological cybernetics

Many time-series such as human movement data consist of a sequence of basic actions, e.g., forehands and backhands in tennis. Automatically extracting and characterizing such actions is an important problem for a variety of different applications. In this paper, we present a probabilistic segmentation approach in which an observed time-series is modeled as a concatenation of segments corresponding to different basic actions. Each segment is generated through a noisy transformation of one of a few hidden trajectories representing different types of movement, with possible time re-scaling. We analyze three different approximation methods for dealing with model intractability, and demonstrate how the proposed approach can successfully segment table tennis movements recorded using a robot arm as haptic input device.
Subject Area: Probabilistic Models and Methods

T93 Active Learning Applied to Patient-Adaptive Heartbeat Classification

Jenna Wiens jwiens@mit.edu
John Guttag guttag@csail.mit.edu
MIT

While clinicians can accurately identify different types of heartbeats in electrocardiograms (ECGs) from different patients, researchers have had limited success in applying supervised machine learning to the same task. The problem is made challenging by the variety of tasks, inter- and intra-patient differences, an often severe class imbalance, and the high cost of getting cardiologists to label data for individual patients. We address these difficulties using active learning to perform patient-adaptive and task-adaptive heartbeat classification. When tested on a benchmark database of cardiologist annotated ECG recordings, our method had considerably better performance than other recently proposed methods on the two primary classification tasks recommended by the Association for the Advancement of Medical Instrumentation. Additionally, our method required over 90% less patient-specific training data than the methods to which we compared it.
Subject Area: Applications

T94 Static Analysis of Binary Executables Using Structural SVMs

Nikos Karampatziakis nk@cs.cornell.edu
Cornell University

We cast the problem of identifying basic blocks of code in a binary executable as learning a mapping from a byte sequence to a segmentation of the sequence. In general, inference in segmentation models, such as semi-CRFs, can be cubic in the length of the sequence. By taking advantage of the structure of our problem, we derive a linear-time inference algorithm which makes our approach practical, given that even small programs are tens or hundreds of thousands bytes long. Furthermore, we introduce two loss functions which are appropriate for our problem and show how to use structural SVMs to optimize the learned mapping for these losses. Finally, we present experimental results that demonstrate the advantages of our method against a strong baseline.
Subject Area: Applications

T95 Latent Variable Models for Predicting File Dependencies in Large-Scale Software Development

Diane Hu dhu@cs.ucsd.edu
Laurens van der Maaten lvdmaaten@gmail.com
Youngmin Cho yoc002@cs.ucsd.edu
Lawrence Saul saul@cs.ucsd.edu
Sorin Lerner lerner@cs.ucsd.edu
University of California San Diego

When software developers modify one or more files in a large code base, they must also identify and update other related files. Many file dependencies can be detected by mining the development history of the code base: in essence, groups of related files are revealed by the logs of previous workflows. From data of this form, we show how to detect dependent files by solving a problem in binary matrix completion. We explore different latent variable models (LVMs) for this problem, including Bernoulli mixture models, exponential family PCA, restricted Boltzmann machines, and fully Bayesian approaches. We evaluate these models on the development histories of three large, open-source software systems: Mozilla Firefox, Eclipse Subversive, and Gimp. In all of these applications, we find that LVMs improve the performance of related file prediction over current leading methods.
Subject Area: Applications

T96 Link Discovery using Graph Feature Tracking

Emile Richard e1000richard@gmail.com
Nicolas Baskiotis nicolas.Baskiotis@gmail.com
Nicolas Vayatis nicolas.vayatis@cmla.ens-cachan.fr
CMLA/ENS Cachan
Theodoros Evgeniou theodoros.evgeniou@insead.edu
INSEAD

We consider the problem of discovering links of an evolving undirected graph given a series of past snapshots of that graph. The graph is observed through the time sequence of its adjacency matrix and only the presence of edges is observed. The absence of an edge on a certain snapshot cannot be distinguished from a missing entry in the adjacency matrix. Additional information can be provided by examining the dynamics of the graph through a set of topological features such as the degrees of the vertices. We develop a novel methodology by building on both static matrix completion methods and the estimation of the future state of relevant graph features. Our procedure relies on the formulation of an optimization problem which can be approximately solved by a fast alternating linearized algorithm whose properties are examined. We show experiments with both simulated and real data which reveal the interest of our methodology.
Subject Area: Applications

T97 Global seismic monitoring as probabilistic inference

Nimar S Arora nimar.arora@gmail.com
Stuart Russell russell@cs.berkeley.edu
UC Berkeley
Paul Kidwell kidwell1@llnl.gov
Lawrence Livermore National Lab
Erik Sudderth sudderth@cs.brown.edu
Brown University

The International Monitoring System (IMS) is a global network of sensors whose purpose is to identify potential violations of the Comprehensive Nuclear-Test-Ban Treaty (CTBT), primarily through detection and localization of seismic events. We report on the first stage of a project to improve on the current automated software system with a Bayesian inference system that computes the most likely global event history given the record of local sensor data. The new system, VISA (Vertically Integrated Seismological Analysis), is based on empirically calibrated, generative models of event occurrence, signal propagation, and signal detection. VISA exhibits significantly improved precision and recall compared to the current operational system and is able to detect events that are missed even by the human analysts who post-process the IMS output.
Subject Area: Applications

T98 Improving Human Judgments by Decontaminating Sequential Dependencies

Michael Mozer mozer@cs.colorado.edu
Harold Pashler hpashler@ucsd.edu
Matthew Wilder mattwilder.cu@gmail.com
Robert Lindsey robert.lindsey@colorado.edu
Matt Jones mcj@colorado.edu
University of Colorado at Boulder
Michael Jones jonesmn@indiana.edu
Indiana University

For over half a century, psychologists have been struck by how poor people are at expressing their internal sensations, impressions, and evaluations via rating scales. When individuals make judgments, they are incapable of using an absolute rating scale, and instead rely on reference points from recent experience. This relativity of judgment limits the usefulness of responses provided by individuals to surveys, questionnaires, and evaluation forms. Fortunately, the cognitive processes that transform internal states to responses are not simply noisy, but rather are influenced by recent experience in a lawful manner. We explore techniques to remove sequential dependencies, and thereby decontaminate a series of ratings to obtain more meaningful human judgments. In our formulation, decontamination is fundamentally a problem of inferring latent states (internal sensations) which, because of the relativity of judgment, have temporal dependencies. We propose a decontamination solution using a conditional random field with constraints motivated by psychological theories of relative judgment. Our exploration of decontamination models is supported by two experiments we conducted to obtain ground-truth rating data on a simple length estimation task. Our decontamination techniques yield an over 20% reduction in the error of human judgments.
Subject Area: Applications
Spotlight presentation, Tuesday, 9:40pm

T99 SpikeAnts, a spiking neuron network modelling the emergence of organization in a complex system

Sylvain Chevallier sylvain.chevallier@lri.fr
MicheLe Sebag sebag@lri.fr
Laboratoire de Recherche en Informatique CNRS
H´el`ene Paugam-Moisy hpaugam@lri.fr
Univ. Lyon 2

Many complex systems, ranging from neural cell assemblies to insect societies, involve and rely on some division of labor. How to enforce such a division in a decentralized and distributed way, is tackled in this paper, using a spiking neuron network architecture. Specifically, a spatio-temporal model called SpikeAnts is shown to enforce the emergence of synchronized activities in an ant colony. Each ant is modelled from two spiking neurons; the ant colony is a sparsely connected spiking neuron network. Each ant makes its decision (among foraging, sleeping and self-grooming) from the competition between its two neurons, after the signals received from its neighbor ants. Interestingly, three types of temporal patterns emerge in the ant colony: asynchronous, synchronous, and synchronous periodic foraging activities - similar to the actual behavior of some living ant colonies. A phase diagram of the emergent activity patterns with respect to two control parameters respectively accounting for ant sociability and receptivity is presented and discussed.
Subject Area: Applications
Spotlight presentation, Tuesday, 9:50pm

T100 Learning to Localise Sounds with Spiking Neural Networks

Dan F Goodman dan.goodman@ens.fr
Romain Brette romain.brette@ens.fr
Ecole Normale Superieure

To localise the source of a sound, we use location-specific properties of the signals received at the two ears caused by the asymmetric filtering of the original sound by our head and pinnae, the head-related transfer functions (HRTFs). These HRTFs change throughout an organism's lifetime, during development for example, and so the required neural circuitry cannot be entirely hardwired. Since HRTFs are not directly accessible from perceptual experience, they can only be inferred from filtered sounds. We present a spiking neural network model of sound localisation based on extracting location-specific synchrony patterns, and a simple supervised algorithm to learn the mapping between synchrony patterns and locations from a set of example sounds, with no previous knowledge of HRTFs. After learning, our model was able to accurately localise new sounds in both azimuth and elevation, including the difficult task of distinguishing sounds coming from the front and back.
Subject Area: Speech and Signal Processing
Spotlight presentation, Tuesday, 9:45.

DEMONSTRATIONS

7:30–11:59PM

D1 2-D Cursor Movement using EEG

Chris Laver calaver@gmail.com
University of Guelph

The demonstration will show a non-invasive brain-computer interface device used to control continuous cursor movements in two dimensions. The decision of which direction to move the cursor is made by analyzing ERP components in the time-series EEG signal, using the continuous wavelet transform for time-scale analysis. These ERP components are observed as the result of a stimulus (the cursor moving), and are recorded for 300-1000 milliseconds after the stimulus. Details 1) The moving cursor on the screen is continually observed by the subject, and the results are recorded in 300-1000 ms windows, with a new window being generated every ~100 ms. 2) The EEG trace has filters applied. 3) Important features of the EEG signal are extracted using the continuous wavelet transform for time-scale analysis. 4) The important features are used to determine the new direction for the cursor to proceed. This approach is novel in the area of 2-D EEG movement control because: - The training time for a user is minimal, no more than a few minutes. Other techniques for 2-D movement involve the subject learning to evoke potentials in the EEG trace over a period of days or weeks. - Physical movement analogues, such as imagined movements of the feet or hands, are not employed. - The system will work with low-cost, low-channel EEG hardware. The system will be shown with the Emotiv EPOC hardware, a 14-channel EEG headset, costing approximately $300 USD.

D2 A framework for Evaluating and Designing "Attention Mechanism" Implementations Based on Tracking of Human Eye Movements

Dmitry Chichkov dchichkov@gmail.com
Texas A&M University

We present a framework for collecting large scale multimodal data sets useful for studying human attention mechanism through an eye tracking data analysis. Two approaches of data collection are presented. First - we have designed a low cost, acceptable in the public setting wearable hardware for recording point of view scene, audio, location and human eye movements. Second, we present a web based framework for crowdsourcing a human visual attention corpus that can be completed by annotating online video data with an eye tracking information. The power of our framework is that it is based entirely on low cost, commodity hardware, it is open source and does not require any software to be installed on the user side, thus allowing to use inexpensive labor pools available through crowdsourcing Internet marketplace and scale up the data set.

DEMONSTRATIONS

D3 Haptic Information Presentation Through Vibro Tactile

Manoj A Prasad manoj.prasad@neo.tamu.edu

Our spatial awareness of the environment is usually attributed to our ability to process the visual and auditory cues from the environment. However, this is also complemented by tactile feedback from the hands. Computer human interfaces can take advantage of the complementary nature of modality interaction. The use of tactile feedback in computer human interfaces reduces the cognitive overload on visual and audio channels and acts as an additional source of information. In this paper, we have explored the possibility of encoding and communicating complex information through vibro-tactile stimulation. Our goal here is to estimate the sensitivity of the hand to vibrations and the ability of subjects to recognize simple shapes presented by a matrix of vibro-tactile stimulators.

D4 MetaOptimize: A Q+A site for machine learning

Joseph Turian turian@gmail.com

MetaOptimize Q+A is a site for ML scientists to share knowledge and techniques, to document our ideas in an informal online setting, and to discuss details that don't always make it into publications. Why should you sign up and post a question or answer? * Communicate with experts outside of your lab * Crosspolinate information with researchers in adjacent fields (statistics, vision, data mining, etc.) * Answer a question once publicly, instead of potentially many times over email * Share knowledge to create additional impact beyond publication * Find new collaborators

D5 mldata.org - machine learning data and benchmark

Cheng Soon Ong chengsoon.ong@inf.ethz.ch
ETH Zurich

We introduce mldata.org, an on-going effort to build a data and benchmark repository for the machine learning community (http://mldata.org/about/motivation/). Compared with existing data repositories such as the UCI machine learning repository, the aims of mldata.org are to provide (a) a dynamic site which is frequently updated by a selfmoderating user community (b) machine learning benchmarks and challenges in addition to a store of datasets (c) an easy-to-use API to facilitate and automate interaction with the repository from popular scientific computing environments (such as python, matlab, and R)

D6 NeuFlow: a dataflow processor for convolutional nets and other real-time algorithms

Yann LeCun yann@cs.nyu.edu
New York University

NeuFlow is a new concept of ”dataflow” architecture, which is particularly well-suited for algorithms that perform a fixed set of operations on a stream of data (e.g. an image). NeuFlow is particularly efficient for such vision algorithms as Convolutional Networks. The NeuFlow architecture is currently instantiated on an FPGA board built around a Xilinx Virtex-6, which communicate with a laptop computer through a gigabit ethernet connection. It is capable of a sustained performance of 100 billion multiply-accumulate operations per second while consuming less than 15Watts of power: about 100 times faster than on a conventional processor, and considerably faster than GPUs for a fraction of the power consumtion and a fraction of the volume. The system runs a number of real-time vision demos, such as a face detector, a general object recognition systems (trainable online), a pedestrian detector, and a vision system for off-road mobile robots that can classify obstacles from traversable areas. The system can also be trained on-line to recognize just about anything.

D7 Project Emporia: News Recommendation using Graphical Models

Jurgen Van Gael jvangael@microsoft.com
Microsoft

Project Emporia is a recommendation engine for news. Based on the Matchbox technology (http://research.microsoft.com/apps/pubs/default.aspx?id=79460) it uses a Bayesian probabilistic model to learn the preferences of users for recent news stories. When a person visits Project Emporia he can up or down vote each link according to her taste. The Matchbox model is then updated in real time so it can instantly improve its link recommendation. The news stories themselves are mined by crawling various RSS feeds and Twitter. In this way, Project Emporia performs Bayesian inference on more than 100,000,000 data points every day. Another feature of Project Emporia is the automatic classification of links into categories. The classification is based on a recently published classifier (http://research.microsoft.com/apps/pubs/default.aspx?id=122779). More interestingly though, we have developed a pipeline which uses active learning to automatically discover links that cannot reliably be classified. These links are then automatically sent to Amazon Mechanical Turk for labelling, after which we spam filter the results and update the classification model.

WEDNESDAY CONFERENCE

WEDNESDAY, DECEMBER 8TH

8:30–9:40AM - ORAL SESSION 9
Session Chair: Pascal Poupart

INVITED TALK: The Interplay of Machine Learning and Mechanism Design,
David Parkes, Harvard University

In the economic theory of mechanism design, the goal is to elicit private information from each of multiple agents in order to select a desirable system wide outcome, and despite agent self-interest in promoting individually beneficial outcomes. Auctions provide a canonical example, with information elicited in the form of bids, and an allocation of resources and payments defining an outcome. Indeed, one aspect of the emerging interplay between machine learning (ML) and mechanism design (MD) arises by interpreting auctions as a method for learning agent valuation functions. In addition to seeking sufficient accuracy to support optimal resource allocation, we require for incentive compatibility that prices are insensitive to the inputs of any individual agent and find an interesting connection with regularization in statistical ML. More broadly, ML can be used for de novo design, in learning payment rules with suitable incentive properties. Ideas from MD are also flowing into ML. One example considers the use of mechanisms to elicit private state, reward and transition models, in enabling coordinated exploration and exploitation in multi-agent systems despite self-interest. Another application is to supervised learning, where labeled training data is elicited from self-interested agents, each with its own objective criterion on the hypothesis learned by the mechanism. Looking ahead, a tantalizing challenge problem is to adopt incentive mechanisms for the design of robust agent architectures, for example in assigning internal rewards that promote modular intelligent systems.

David C. Parkes is Gordon McKay Professor of Computer Science in the School of Engineering and Applied Sciences at Harvard University. He was the recipient of the NSF Career Award, the Alfred P. Sloan Fellowship, the Thouron Scholarship and the Harvard University Roslyn Abramson Award for Teaching. Parkes received his Ph.D. degree in Computer and Information Science from the University of Pennsylvania in 2001, and an M.Eng. (First class) in Engineering and Computing Science from Oxford University in 1995. At Harvard, Parkes leads the EconCS group and teaches classes in artificial intelligence, optimization, and topics at the intersection between computer science and economics. Parkes has served as Program Chair of ACM EC'07 and AAMAS'08 and General Chair of ACM EC'10, served on the editorial board of Journal of Artificial Intelligence Research, and currently serves as Editor of Games and Economic Behavior and on the boards of Journal of Autonomous Agents and Multi-agent Systems and INFORMS Journal of Computing. His research interests include computational mechanism design, electronic commerce, stochastic optimization, preference elicitation, market design, bounded rationality, computational social choice, networks and incentives, multi-agent systems, crowd-sourcing and social computing.

- ***Linear Complementarity for Regularized Policy Evaluation and Improvement,*** Jeffrey T Johns, Ronald Parr and Christopher Painter-Wakefield, Duke University

  Recent work in reinforcement learning has emphasized the power of L1 regularization to perform feature selection and prevent overfitting. We propose formulating the L1 regularized linear fixed point problem as a linear complementarity problem (LCP). This formulation offers several advantages over the LARS-inspired formulation, LARS-TD. The LCP formulation allows the use of efficient off-the-shelf solvers, leads to a new uniqueness result, and can be initialized with starting points from similar problems (warm starts). We demonstrate that warm starts, as well as the efficiency of LCP solvers, can speed up policy iteration. Moreover, warm starts permit a form of modified policy iteration that can be used to approximate a ”greedy” homotopy path, a generalization of the LARS-TD homotopy path that combines policy evaluation and optimization.
  Subject Area: Control and Reinforcement Learning

9:40–10:00AM - SPOTLIGHTS SESSION 8
Session Chair: Pascal Poupart

- ***Optimal Bayesian Recommendation Sets and Myopically Optimal Choice Query Sets***
  Paolo Viappiani and Craig Boutilier, University of Toronto
  See abstract, page 93

- ***Online Classification with Specificity Constraints***
  Andrey Bernstein, Shie Mannor and Nahum Shimkin, Technion
  See abstract, page 105

- ***Exact inference and learning for cumulative distribution functions on loopy graphs***
  Jim C Huang, Nebojsa Jojic and Chris Meek, Microsoft Research
  See abstract, page 103

- ***Global Analytic Solution for Variational Bayesian Matrix Factorization***
  Shinichi Nakajima, Nikon Corporation, Masashi Sugiyama, Tokyo Institute of Technology, and Ryota Tomioka, University of Tokyo
  See abstract, page 102

10:00–10:20AM ORAL SESSION 10
Session Chair: Li Fei-Fei

- ***The Multidimensional Wisdom of Crowds***
  Peter Welinder and Pietro Perona, Caltech; Steve Branson and Serge Belongie, UC San Diego

  Distributing labeling tasks among hundreds or thousands of annotators is an increasingly important method for annotating large datasets. We present a method for estimating the underlying value (e.g. the class) of each image from (noisy) annotations provided by multiple annotators. Our method is based on a model of the image formation and annotation process. Each image has different characteristics that are represented in an abstract Euclidean space. Each annotator is modeled as a multidimensional entity with variables representing competence, expertise and bias. This allows the model to discover and represent groups of annotators that have different sets of skills and knowledge, as well as groups of images that differ qualitatively. We find that our model

predicts ground truth labels on both synthetic and real data more accurately than state of the art methods. Experiments also show that our model, starting from a set of binary labels, may discover rich information, such as different "schools of thought" amongst the annotators, and can group together images belonging to separate categories.
Subject Area: Vision

10:50–11:10AM ORAL SESSION 11
Session Chair: Katherine Heller

- ***Construction of Dependent Dirichlet Processes based on Poisson Processes***
  Dahua Lin, Eric Grimson and John Fisher, MIT

  We present a method for constructing dependent Dirichlet processes. The new approach exploits the intrinsic relationship between Dirichlet and Poisson processes in order to create a Markov chain of Dirichlet processes suitable for use as a prior over evolving mixture models. The method allows for the creation, removal, and location variation of component models over time while maintaining the property that the random measures are marginally DP distributed. Additionally, we derive a Gibbs sampling algorithm for model inference and test it on both synthetic and real data. Empirical results demonstrate that the approach is effective in estimating dynamically varying mixture models.
  Subject Area: Probabilistic Models and Methods

11:10–11:30AM SPOTLIGHTS SESSION 9
Session Chair: Katherine Heller

- ***Throttling Poisson Processes***
  Uwe Dick, Peter Haider, Thomas Vanck, Michael Bruckner and Tobias Scheffer, University of Potsdam
  See abstract, page 103

- ***Graph-Valued Regression***
  Han Liu, Xi Chen, John Lafferty and Larry Wasserman, Carnegie Mellon University.
  See abstract, page 100

- ***Switched Latent Force Models for Movement Segmentation***
  Mauricio Alvarez and Neil D Lawrence, University of Manchester; Jan Petersand Bernhard Schoelkopf, MPI for Biological Cybernetics,
  See abstract, page 102

- ***Online Markov Decision Processes under Bandit Feedback***
  Gergely Neu, Budapest U. of Tech. and Econ., Andras Gyorgy and Andras Antos, MTA SZTAKI Institute for Computer Science and Control; Csaba Szepesvari, University of Alberta,
  See abstract, page 104

11:30–11:50AM ORAL SESSION 12
Session Chair: Nando de Freitas

- ***Slice sampling covariance hyperparameters of latent Gaussian models***
  Iain Murray, University of Edinburgh, and Ryan Adams, University of Toronto

  The Gaussian process (GP) is a popular way to specify dependencies between random variables in a probabilistic model. In the Bayesian framework the covariance structure can be specified using unknown hyperparameters. Integrating over these hyperparameters considers different possible explanations for the data when making predictions. This integration is often performed using Markov chain Monte Carlo (MCMC) sampling. However, with non-Gaussian observations standard hyperparameter sampling approaches require careful tuning and may converge slowly. In this paper we present a slice sampling approach that requires little tuning while mixing well in both strong- and weak-data regimes.
  Subject Area: Probabilistic Models and Methods

11:50AM–12:10PM SPOTLIGHTS SESSION 10
Session Chair: Nando de Freitas

- ***Distributed Dual Averaging In Networks***
  John Duchi, Alekh Agarwal and Martin Wainwright, UC Berkeley
  See abstract, page 88

- ***A Family of Penalty Functions for Structured Sparsity***
  Charles A Micchelli, City Univ. of Hong Kong, Jean Morales and Massi Pontil, University College London
  See abstract, page 84

- ***Guaranteed Rank Minimization via Singular Value Projection***
  Prateek Jain, Microsoft Research India Lab, Raghu Meka and Inderjit Dhillon, University of Texas (RIP)
  See abstract, page 95

- ***Online Learning for Latent Dirichlet Allocation***
  Matthew Hoffman, David Blei, Princeton University, and Francis Bach, Ecole Normale Superieure
  See abstract, page 101

2:00–3:10PM ORAL SESSION 13
Session Chair: Nathan Srebro

***INVITED TALK: Statistical Inference of Protein Structure & Function -* Posner Lecture**
Michael I Jordan, UC Berkeley

The study of the structure and function of proteins serves up many problems that offer challenges and opportunities for computational and statistical research. I will overview my experiences in several such problem domains, ranging from domains where off-the-shelf ideas can be fruitfully applied to domains that require new thinking. These are: (1) the identification of active sites in enzymes; (2) the modeling of protein backbone configurations; (3) the prediction of molecular function based on phylogeny; (4) alignment and phylogenetic inference.

Michael I. Jordan is the Pehong Chen Distinguished Professor in the Department of Electrical Engineering and Computer Science and the Department of Statistics at the University of California, Berkeley. His research in recent years has focused on Bayesian nonparametric analysis, probabilistic graphical models, spectral methods, kernel machines and applications to problems in signal processing, statistical genetics, computational biology, information retrieval and natural language processing. Prof. Jordan was named to the National Academy of Sciences (NAS) in 2010 and the National

Academy of Engineering (NAE) in 2010. He is a Fellow of the American Association for the Advancement of Science (AAAS). He has been named a Neyman Lecturer and a Medallion Lecturer by Institute of Mathematical Statistics (IMS). He is a Fellow of the IMS, a Fellow of the IEEE, a Fellow of the AAAI and a Fellow of the ASA

- ***Tree-Structured Stick Breaking for Hierarchical Data***
  Ryan Adams, University of Toronto; Zoubin Ghahramani, Cambridge; Michael I Jordan, UC Berkeley

  Many data are naturally modeled by an unobserved hierarchical structure. In this paper we propose a flexible nonparametric prior over unknown data hierarchies. The approach uses nested stick-breaking processes to allow for trees of unbounded width and depth, where data can live at any node and are infinitely exchangeable. One can view our model as providing infinite mixtures where the components have a dependency structure corresponding to an evolutionary diffusion down a tree. By using a stick-breaking approach, we can apply Markov chain Monte Carlo methods based on slice sampling to perform Bayesian inference and simulate from the posterior distribution on trees. We apply our method to hierarchical clustering of images and topic modeling of text data.
  Subject Area: Probabilistic Models and Methods

3:10–3:30PM SPOTLIGHTS SESSION 11
Session Chair: Nathan Srebro

- ***Structured Determinantal Point Processes***
  Alex Kulesza and Ben Taskar, University of Pennsylvania
  See abstract, page 104

- ***Supervised Clustering***
  Pranjal Awasthi, CMU, and Reza Bosagh Zadeh, Stanford
  See abstract, page 95

- ***Online Learning in The Manifold of Low-Rank Matrices***
  Uri Shalit, Hebrew University of Jerusalem, Daphna Weinshall and Gal Chechik, Google
  See abstract, page 89

- ***Sample Complexity of Testing the Manifold Hypothesis*** Hariharan Narayanan and Sanjoy K Mitter, MIT
  See abstract, page 94

3:30–3:50PM ORAL SESSION 14
Session Chair: Pradeep Ravikumar

- ***Identifying Graph-structured Activation Patterns in Networks***
  James L Sharpnack and Aarti Singh, CMU

  We consider the problem of identifying an activation pattern in a complex, large-scale network that is embedded in very noisy measurements. This problem is relevant to several applications, such as identifying traces of a biochemical spread by a sensor network, expression levels of genes, and anomalous activity or congestion in the Internet. Extracting such patterns is a challenging task specially if the network is large (pattern is very high-dimensional) and the noise is so excessive that it masks the activity at any single node. However, typically there are statistical dependencies in the network activation process that can be leveraged to fuse the measurements of multiple nodes and enable reliable extraction of high-dimensional noisy patterns. In this paper, we analyze an estimator based on the graph Laplacian eigenbasis, and establish the limits of mean square error recovery of noisy patterns arising from a probabilistic (Gaussian or Ising) model based on an arbitrary graph structure. We consider both deterministic and probabilistic network evolution models, and our results indicate that by leveraging the network interaction structure, it is possible to consistently recover high-dimensional patterns even when the noise variance increases with network size.
  Subject Area: Supervised Learning

4:20–5:00PM ORAL SESSION 15
Session Chair: Irina Rish

- ***Phoneme Recognition with Large Hierarchical Reservoirs***
  Fabian Triefenbach, Azarakhsh Jalalvand, Benjamin Schrauwen and Jean-Pierre Martens, Ghent University

  Automatic speech recognition has gradually improved over the years, but the reliable recognition of unconstrained speech is still not within reach. In order to achieve a breakthrough, many research groups are now investigating new methodologies that have potential to outperform the Hidden Markov Model technology that is at the core of all present commercial systems. In this paper, it is shown that the recently introduced concept of Reservoir Computing might form the basis of such a methodology. In a limited amount of time, a reservoir system that can recognize the elementary sounds of continuous speech has been built. The system already achieves a state-of-the-art performance, and there is evidence that the margin for further improvements is still significant.
  Subject Area: Speech and Signal Processing

- ***Identifying Patients at Risk of Major Adverse Cardiovascular Events Using Symbolic Mismatch***
  Zeeshan Syed, Univ. of Michigan, and John Guttag, MIT

  Cardiovascular disease is the leading cause of death globally, resulting in 17 million deaths each year. Despite the availability of various treatment options, existing techniques based upon conventional medical knowledge often fail to identify patients who might have benefited from more aggressive therapy. In this paper, we describe and evaluate a novel unsupervised machine learning approach for cardiac risk stratification. The key idea of our approach is to avoid specialized medical knowledge, and assess patient risk using symbolic mismatch, a new metric to assess similarity in long-term time-series activity. We hypothesize that high risk patients can be identified using symbolic mismatch, as individuals in a population with unusual long-term physiological activity. We describe related approaches that build on these ideas to provide improved medical decision making for patients who have recently suffered coronary attacks. We first describe how to compute the symbolic mismatch between pairs of long term electrocardiographic (ECG) signals. This algorithm maps the original signals into a symbolic domain, and provides a quantitative assessment of the difference between these symbolic representations of the original signals. We then show how this measure can be used with each of a one-class SVM, a nearest neighbor classifier, and hierarchical clustering to improve risk stratification. We evaluated our methods on a population of 686 cardiac patients with available long-term electrocardiographic data. In a univariate analysis, all of the methods provided a statistically significant association with the occurrence of

a major adverse cardiac event in the next 90 days. In a multivariate analysis that incorporated the most widely used clinical risk variables, the nearest neighbor and hierarchical clustering approaches were able to statistically significantly distinguish patients with a roughly two-fold risk of suffering a major adverse cardiac event in the next 90 days.
Subject Area: Applications

5:00–5:20PM SPOTLIGHTS SESSION 12
Session Chair: Killian Weinberger

- ***Functional Geometry Alignment and Localization of Brain Areas***
  Georg Langs and Polina Golland, MIT; Yanmei Tie, Laura Rigolo, Alexandra Golby, Harvard Medical School
  See abstract, page 98
- ***Epitome driven 3-D Diffusion Tensor image segmentation: on extracting specific structures***
  Kamiya Motwani, Nagesh Adluru, Chris Hinrichs, Andrew L Alexander and Vikas Singh, Univ. of Wisconsin
  See abstract, page 97
- ***On a Connection between Importance Sampling and the Likelihood Ratio Policy Gradient***
  Tang Jie and Pieter Abbeel, UC Berkeley
  See abstract, page 87
- ***Shadow Dirichlet for Restricted Probability Modeling***
  Bela A Frigyik, Maya Gupta, University of Washington, and Yihua Chen
  See abstract, page 101

5:20–5:40PM ORAL SESSION 16

- ***On the Convexity of Latent Social Network Inference,***
  Seth Myers and Jure Leskovec, Stanford University

  In many real-world scenarios, it is nearly impossible to collect explicit social network data. In such cases, whole networks must be inferred from underlying observations. Here, we formulate the problem of inferring latent social networks based on network diffusion or disease propagation data. We consider contagions propagating over the edges of an unobserved social network, where we only observe the times when nodes became infected, but not who infected them. Given such node infection times, we then identify the optimal network that best explains the observed data. We present a maximum likelihood approach based on convex programming with a l1-like penalty term that encourages sparsity. Experiments on real and synthetic data reveal that our method near-perfectly recovers the underlying network structure as well as the parameters of the contagion propagation model. Moreover, our approach scales well as it can infer optimal networks on thousands of nodes in a matter of minutes.
  Subject Area: Probabilistic Models and Methods

5:40–6:00PM SPOTLIGHTS SESSION 13

- ***Learning from Logged Implicit Exploration Data***
  Alex Strehl, Facebook, John Langford and Lihong Li, Yahoo! Research; Sham Kakade, Univ. of Pennsylvania
  See abstract, page 87
- ***Deciphering subsampled data: adaptive compressive sampling as a principle of brain communication***
  Guy Isely, Christopher J Hillar and Fritz Sommer, UC Berkeley
  See abstract, page 98
- ***Feature Set Embedding for Incomplete Data***
  David Grangier and Iain Melvin, NEC Laboratories USA
  See abstract, page 86
- ***More data means less inference: A pseudo-max approach to structured learning***
  David Sontag and Tommi Jaakkola MIT; Ofer Meshi and Amir Globerson, Hebrew University
  See abstract, page 85

7:00–11:59PM - POSTER SESSION

W1 **A Primal-Dual Algorithm for Group Sparse Regularization with Overlapping Groups**, sofia mosci, Silvia Villa and Alessandro Verri, DISI Universita’ di Genova; Lorenzo Rosasco, MIT and IIT

W2 **A Family of Penalty Functions for Structured Sparsity**, Charles A Micchelli, City Univ. of Hong Kong; Jean Morales, and Massi Pontil, Univ. College London

W3 **On Herding and the Perceptron Cycling Theorem**, Andrew E Gelfand, Yutian Chen and Max Welling, UC Irvine; Laurens van der Maaten, UC San Diego,

W4 **Collaborative Filtering in a Non-Uniform World: Learning with the Weighted Trace Norm**, Ruslan Salakhutdinov, MIT, and Nathan Srebro, TTI

W5 **Learning Multiple Tasks with a Sparse Matrix-Normal Penalty**, Yi Zhang and Jeff Schneider, Carnegie Mellon University

W6 **Learning Multiple Tasks using Manifold Regularization**, Arvind Agarwal, Hal Daume III and Samuel Gerber, University of Utah

W7 **Label Embedding Trees for Large Multi-Class Tasks**, Samy Bengio, Jason Weston, Google, and David Grangier, NEC Labs America

W8 **Learning via Gaussian Herding**, Koby Crammer, Technion, and Daniel Lee, University of Pennsylvania

W9 **A Novel Kernel for Learning a Neuron Model from Spike Train Data**, Nicholas K Fisher and Arunava Banerjee, University of Florida

W10 **More data means less inference: A pseudo-max approach to structured learning**, David Sontag and Tommi Jaakkola, MIT; Ofer Meshi and Amir Globerson, Hebrew University

W11 **Identifying graph-structured activation patterns in networks**, James L Sharpnack and Aarti Singh, CMU

W12 **Penalized Principal Component Regression on Graphs for Analysis of Sub networks**, Ali Shojaie and George Michailidis, University of Michigan

W13 **Feature Set Embedding for Incomplete Data**, David Grangier and Iain Melvin, NEC Laboratories America

W14 **Learning to combine foveal glimpses with a third-order Boltzmann machine**, Hugo Larochelle and Geoffrey Hinton, University of Toronto

W15 **Feature Construction for Inverse Reinforcement Learning**, Sergey Levine and Vladlen Koltun, Stanford University; Zoran Popovic, University of Washington

W16 **Distributionally Robust Markov Decision Processes**, Huan Xu,The University of Texas, and Shie Mannor, Technion

W17 **A Reduction from Apprenticeship Learning to Classification**, Umar Syed, University of Pennsylvania; Robert E Schapire, Princeton University

W18 **A POMDP Extension with Belief-dependent Rewards**, Mauricio A Araya, Olivier Buffet, Vincent Thomas and Francois Charpillet, Nancy University / INRIA / CNRS

W19 **On a Connection between Importance Sampling and the Likelihood Ratio Policy Gradient**, Tang Jie and Pieter Abbeel, UC Berkeley

W20 **Learning from Logged Implicit Exploration Data**, Alex Strehl, Facebook, John Langford and Lihong Li, Yahoo! Research; Sham Kakade, Univ. of Pennsylvania

W21 **Policy gradients in linearly-solvable MDPs**, Emanuel Todorov, University of Washington

W22 **Distributed Dual Averaging In Networks**, John Duchi, Alekh Agarwal and Martin Wainwright, UC Berkeley .

W23 **Online Learning in The Manifold of Low-Rank Matrices**, Uri Shalit, Hebrew University of Jerusalem, Daphna Weinshall and Gal Chechik, Google

W24 **Moreau-Yosida Regularization for Grouped Tree Structure Learning**, Jun Liu and Jieping Ye, Arizona State University

W25 **A Log-Domain Implementation of the Diffusion Network in Very Large Scale Integration**, Yi-Da Wu, Shi-Jie Lin and Hsin Chen, National Tsing Hua University

W26 **Phoneme Recognition with Large Hierarchical Reservoirs**, FabianTriefenbach, ELIS, Azarakhsh Jalalvand, Benjamin Schrauwen and Jean-Pierre Martens, Ghent University

W27 **Multivariate Dyadic Regression Trees for Sparse Learning Problems**, Han Liu and Xi Chen, CMU

W28 **Sparse Coding for Learning Interpretable Spatio-Temporal Primitives**, Taehwan Kim, Gregory Shakhnarovich and Raquel Urtasun, TTI Chicago

W29 **Efficient and Robust Feature Selection via Joint 2,1-Norms Minimization**, Feiping Nie, Heng Huang, Xiao Cai and Chris Ding, Univ. of Texas at Arlington

W30 **Natural Policy Gradient Methods with Parameter-based Exploration for Control Tasks**, Atsushi Miyamae, Tokyo Institute of Technology/JSPS Research Fellow, Yuichi Nagata, Isao Ono and Shigenobu Kobayashi, Tokyo Institute of Technology

W31 **Linear Complementarity for Regularized Policy Evaluation and Improvement**, Jeffrey T Johns, Christopher Painter-Wakefield and Ronald Parr, Duke University

W32 **Lower Bounds on Rate of Convergence of Cutting Plane Methods**, Xinhua Zhang, AICML, Ankan Saha, University of Chicago, and S.V.N.Vishwanathan, Purdue University

W32 **Large Margin Learning of Upstream Scene Understanding Models**, Jun Zhuand and Eric Xing, Carnegie Mellon University; Li-Jia Li, Li Fei-Fei, Stanford University,

W33 **The Multidimensional Wisdom of Crowds**, Peter Welinder and Pietro Perona, Caltech; Steve Branson and Serge Belongie, UC San Diego,

W35 **Estimating Spatial Layout of Rooms using Volumetric Reasoning about Objects and Surfaces**, David C Lee, Abhinav Gupta, Martial Hebert and Takeo Kanade, Carnegie Mellon University

W36 **Space-Variant Single-Image Blind Deconvolution for Removing Camera Shake**, Stefan Harmeling, Michael Hirsch and Bernhard Schoelkopf, MPI for Biological Cybernetics

W37 **Segmentation as Maximum-Weight Independent Set**, William Brendel, and Sinisa Todorovic, Oregon State University

W38 **Generating more realistic images using gated MRF's**, Marc'Aurelio Ranzato, Volodymyr Mnih and Geoffrey Hinton, University of Toronto

W39 **A Discriminative Latent Model of Image Region and Object Tag Correspondence**, Yang Wang, and Greg Mori,Simon Fraser University

W40 **Using body-anchored priors for identifying actions in single images**, Leonid Karlinsky, Michael Dinerstein and Shimon Ullman, Weizmann Institute of Science

W41 **Simultaneous Object Detection and Ranking with Weak Supervision**, Matthew B Blaschko, Andrea Vedaldi, and Andrew Zisserman, Oxford University

W42 **Feature Transitions with Saccadic Search: Size, Color, and Orientation Are Not Alike**, Stella X Yu, Boston College

W43 **Identifying Patients at Risk of Major Adverse Cardiovascular Events Using Symbolic Mismatch**, Zeeshan Syed, Univ. of Michigan, and John Guttag, MIT

W44 **Sphere Embedding: An Application to Part-of-Speech Induction**,Yariv Maron, Bar Ilan University, Michael Lamar, Saint Louis University, and Elie Bienenstock, Brown University

W45 **Efficient Optimization for Discriminative Latent Class Models**, Armand Joulin and Francis Bach, Ecole Normale Superieure; Jean Ponce

W46 **Sample Complexity of Testing the Manifold Hypothesis**, Hariharan Narayanan and Sanjoy K Mitter, MIT.

W47 **Optimal Bayesian Recommendation Sets and Myopically Optimal Choice Query Sets**, Paolo Viappiani and Craig Boutilier, University of Toronto

W48 **Extensions of Generalized Binary Search to Group Identification and Exponential Costs**, Gowtham Bellala and Clayton Scott, University of Michigan Ann Arbor; Suresh Bhavnani, University of Texas,

W49 **Active Learning by Querying Informative and Representative Examples**, Sheng-Jun Huang and Zhi-Hua Zhou, Nanjing University; Rong Jin, Michigan State University,

W50 **Active Instance Sampling via Matrix Partition**, Yuhong Guo, Temple University

W51 **Robust Clustering as Ensembles of Affinity Relations**, Hairong Liu and Shuicheng Yan, National University of Singapore; Longin Jan Latecki, Temple University,

W52 **Supervised Clustering**, Pranjal Awasthi, CMU; Reza Bosagh Zadeh, Stanford University

W53 **Rates of convergence for the cluster tree**, Kamalika Chaudhuri and Sanjoy Dasgupta, UC San Diego

W54 **Random Projection Trees Revisited**, Aman Dhesi, Princeton University, and Purushottam Kar, Indian Institute of Technology

W55 **Worst-Case Linear Discriminant Analysis**, Yu Zhang and Dit-YanYeung, HKUST

W56 **Guaranteed Rank Minimization via Singular Value Projection**, Prateek Jain, Microsoft Research India Lab; Raghu Meka and Inderjit Dhillon, Univ. of Texas (RIP).

W57 **An Inverse Power Method for Nonlinear Eigen problems with Applications in 1-Spectral Clustering and Sparse PCA**, Matthias Hein and Thomas Buhler, Saarland University

W58 **Word Features for Latent Dirichlet Allocation**, James Petterson, Tiberio Caetano and Wray Buntine, NICTA; Alexander J Smola, Yahoo! Research; Shravan M Narayanamurthy,Yahoo! Labs Bangalore

W59 **A novel family of non-parametric cumulative based divergences for point processes**, Sohan Seth, Park Il, Austin Brockmeier and and Jose Principe, University of Florida; Mulugeta Semework, John Choi, Joseph T Francis, SUNY Downstate and NYU-Poly,

W60 **Brain covariance selection: better individual functional connectivity models using population prior**, Gael Varoquaux, Alexandre Gramfort and Bertrand Thirion, INRIA; Jean-Baptiste Poline, CEA

W61 **Epitome driven 3-D Diffusion Tensor image segmentation: on extracting specific structures**, Kamiya Motwani, Nagesh Adluru, Chris Hinrichs, Andrew L Alexander and Vikas Singh, Univ. of Wisconsin Madison

W62 **Spatial and anatomical regularization of SVM for brain image analysis**, Remi Cuingnet, Marie Chupin, CRICM, Habib Benali,INSERM LIF, and Olivier Colliot, CRICM / CNRS / Universite Pierre et Marie Curie / Paris France.

W63 **Infinite Relational Modeling of Functional Connectivity in Resting State fMRI**, Morten Mørup and Lars K Hansen, DTU Informatics; Kristoffer H Madsen, Anne-Marie Dogonowski, Hartwig R Siebner, Danish Research Centre for Magnetic Resonance,

W64 **Functional Geometry Alignment and Localization of Brain Areas**, Georg Langs and Polina Golland, MIT; Yanmei Tie, Laura Rigolo, Alexandra Golby, Harvard Medical School

W65 **Effects of Synaptic Weight Diffusion on Learning in Decision Making Networks**, Kentaro Katahira, and Kazuo Okanoya, JST ERATO Okanoya Emotional Information Project; Masato Okada, The University of Tokyo / RIKEN

W66 **Deciphering subsampled data: adaptive compressive sampling asa principle of brain communication**, Guy Isely, Christopher JHillar, and Fritz Sommer, UC Berkeley

W67 **Accounting for network effects in neuronal responses using L1 regularized point process models**, Ryan C Kelly and Robert E Kass, Carnegie Mellon University, Matthew A Smith, University of Pittsburgh; Tai Sing Lee

W68 **Humans Learn Using Manifolds, Reluctantly**, Bryan R Gibson, Xiaojin (Jerry) Zhu, Tim Rogers, Univ. of Wisonsin-Madison, Chuck Kalish, and JosephHarrison

W69 **Linear readout from a neural population with partial correlation data**, Adrien Wohrer and and Christian K Machens, Ecole Normale Superieure; Ranulfo Romo, Universidad Autonoma de Mexico,

W70 **Implicit encoding of prior probabilities in optimal neural populations**, Deep Ganguli and Eero P Simoncelli, New York University

W71 **Fractionally Predictive Spiking Neurons**, Sander Bohte and Jaldert Rombouts, CWI

W72 **Learning Efficient Markov Networks**, Vibhav G Gogate, William Webb and Pedro Domingos, University of Washington

W73 **Variational bounds for mixed-data factor analysis**, Mohammad Emtiyaz Khan, Benjamin Marlin and Kevin Murphy, University of British Columbia; Guillaume Bouchard, XRCE,

W74 **Probabilistic latent variable models for distinguishing between cause and effect**, Joris Mooij, Oliver Stegle, Dominik Janzing, Kun Zhang and Bernhard Schoelkopf, MPI for Biological Cybernetics

W75 **Graph-Valued Regression**, Han Liu, Xi Chen, John Lafferty and Larry Wasserman, Carnegie Mellon Univ.

W76 **Learning Networks of Stochastic Differential Equations**, Jos´eBento Ayres Pereira, Morteza Ibrahimi and Andrea Montanari, Stanford University

W77 **Copula Bayesian Networks**, Gal Elidan, Hebrew University

.

W78 **t-logistic regression**, Nan Ding and S.V.N. Vishwanathan, Purdue University

W79 **Shadow Dirichlet for Restricted Probability Modeling**, Bela A Frigyik and Maya Gupta, Univ. of Washington; Yihua Chen

W80 **Online Learning for Latent Dirichlet Allocation**, Matthew Hoffman, David Blei, Princeton University; Francis Bach, Ecole Normale Superieure

W81 **Approximate inference in continuous time Gaussian-Jump processes**, Manfred Opper, Technische Universitaet Berlin, Andreas Ruttor, TU Berlin; Guido Sanguinetti, University of Edinburgh

W82 **Global Analytic Solution for Variational Bayesian Matrix Factorization**, Shinichi Nakajima, Nikon Corporation, Masashi Sugiyama, Tokyo Institute of Technology, and Ryota Tomioka, University of Tokyo.

W83 **Tree-Structured Stick Breaking for Hierarchical Data**, RyanAdams, University of Toronto, Zoubin Ghahramani, Cambridge, and Michael I Jordan, UC Berkeley

W84 **Construction of Dependent Dirichlet Processes based on PoissonProcesses**, Dahua Lin, Eric Grimson and John Fisher, MIT

W85 **Switched Latent Force Models for Movement Segmentation**, Mauricio Alvarez and and Neil D Lawrence, University of Manchester; Jan Peters and Bernhard Schoelkopf, MPI for Biological Cybernetics,

W86 **Slice sampling covariance hyperparameters of latent Gaussianmodels**, Iain Murray, University of Edinburgh, and Ryan Adams,University of Toronto

W87 **Exact inference and learning for cumulative distribution functionson loopy graphs**, Jim C Huang, Nebojsa Jojic and Chris Meek, Microsoft Research

W88 **Evidence-Specific Structures for Rich Tractable CRFs**, Anton Chechetka and Carlos Guestrin, Carnegie Mellon University

W89 **On the Convexity of Latent Social Network Inference**, Seth Myers and Jure Leskovec, Stanford University

W90 **Structured Determinantal Point Processes**, Alex Kulesza, and Ben Taskar, University of Pennsylvania

W91 **Throttling Poisson Processes**, Uwe Dick, Peter Haider, Thomas Vanck, Michael Bruckner and Tobias Scheffer, UniPotsdam, University of Potsdam

W92 **Dynamic Infinite Relational Model for Time-varying RelationalData Analysis**, Katsuhiko Ishiguro, Tomoharu Iwata and Naonori Ueda, NTT Communication Science Laboratories; JoshuaTenenbaum, Massachusetts Institute of Technology

W93 **Universal Consistency of Multi-Class Support VectorClassification**, Tobias Glasmachers, IDSIA

W94 **Random Walk Approach to Regret Minimization**, HariharanNarayanan, MIT, and Alexander Rakhlin, University of Pennsylvania

W95 **Online Markov Decision Processes under Bandit Feedback**,Gergely Neu, Budapest U. of Tech. and Econ., Andras Gyorgy and and Andras Antos, MTASZTAKI Institute for Computer Science and Control, Csaba Szepesvari, University of Alberta

W96 **New Adaptive Algorithms for Online Classification**, Francesco Orabona, University of Milano, and Koby Crammer, Technion

W97 **Online Classification with Specificity Constraints**, Andrey Bernstein, Shie Mannor and Nahum Shimkin, Technion

W98 **The LASSO risk: asymptotic results and real world examples**, Mohsen Bayati, Jose Bento Ayres Pereira and Andrea Montanari, Stanford University

W99 **Tight Sample Complexity of Large-Margin Learning**, Sivan Sabato and Naftali Tishby, The Hebrew University; Nathan Srebro, TTI,

W100 **Universal Kernels on Non-Standard Input Spaces**, Andreas Christmann, University of Bayreuth, and Ingo Steinwart, University of Stuttgart

WEDNESDAY - DEMONSTRATIONS

7:30–11:59PM DEMONSTRATIONS

D8 **BCI Demonstration using a Dry-Electrode**, Achim Hornecker, sr@brainproducts.com, Brain Products Gmb

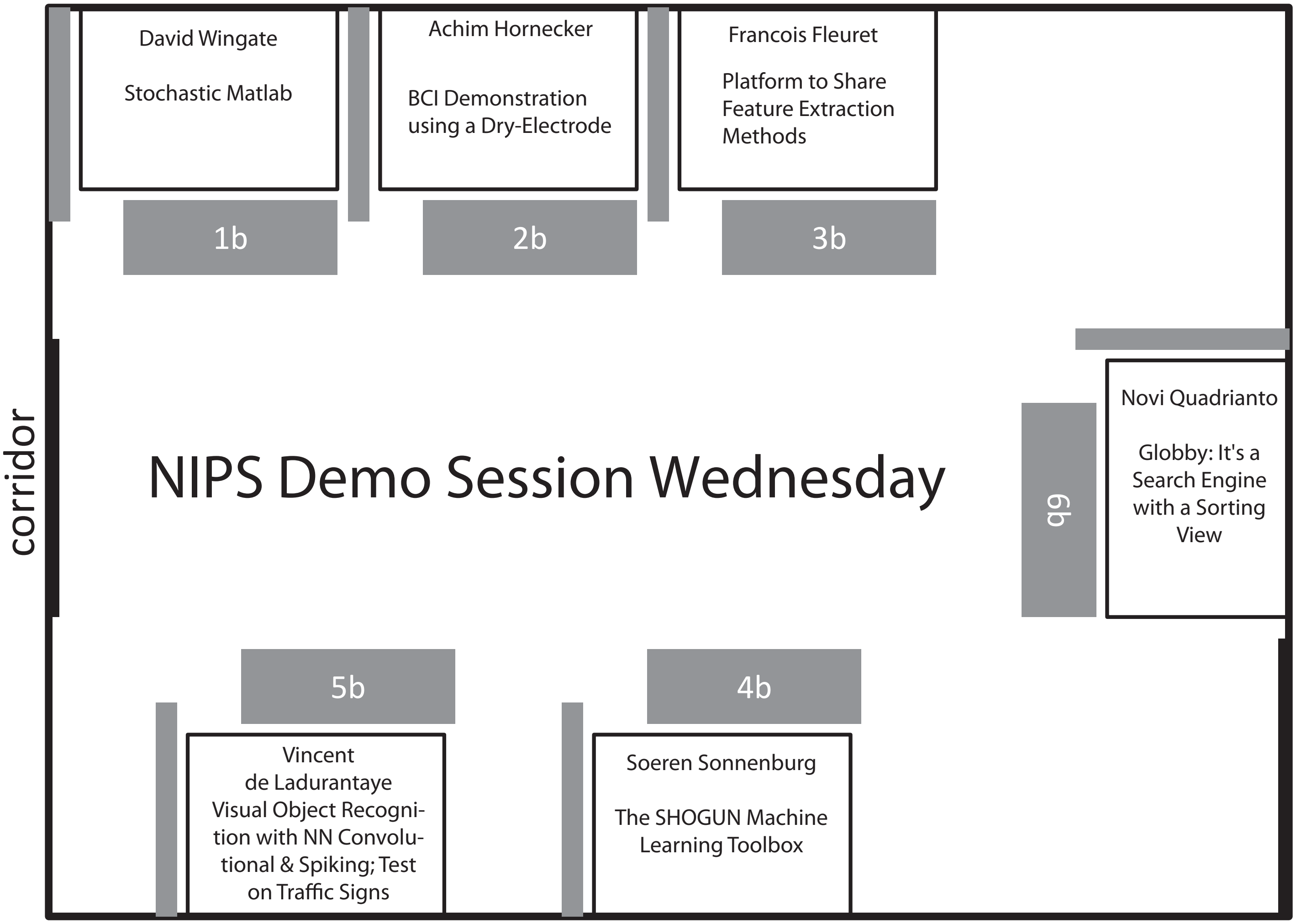
D9 **Globby: It's a Search Engine with a Sorting View**, Novi Quadrianto, novi.quad@gmail.com, NICTA

D10 **Platform to Share Feature Extraction Methods**, Francois Fleuret, francois.fleuret@idiap.ch, IDIAP Research Institute

D11 **Stochastic Matlab**, David Wingate, wingated@mit.edu, Massachusetts Institute of Technology

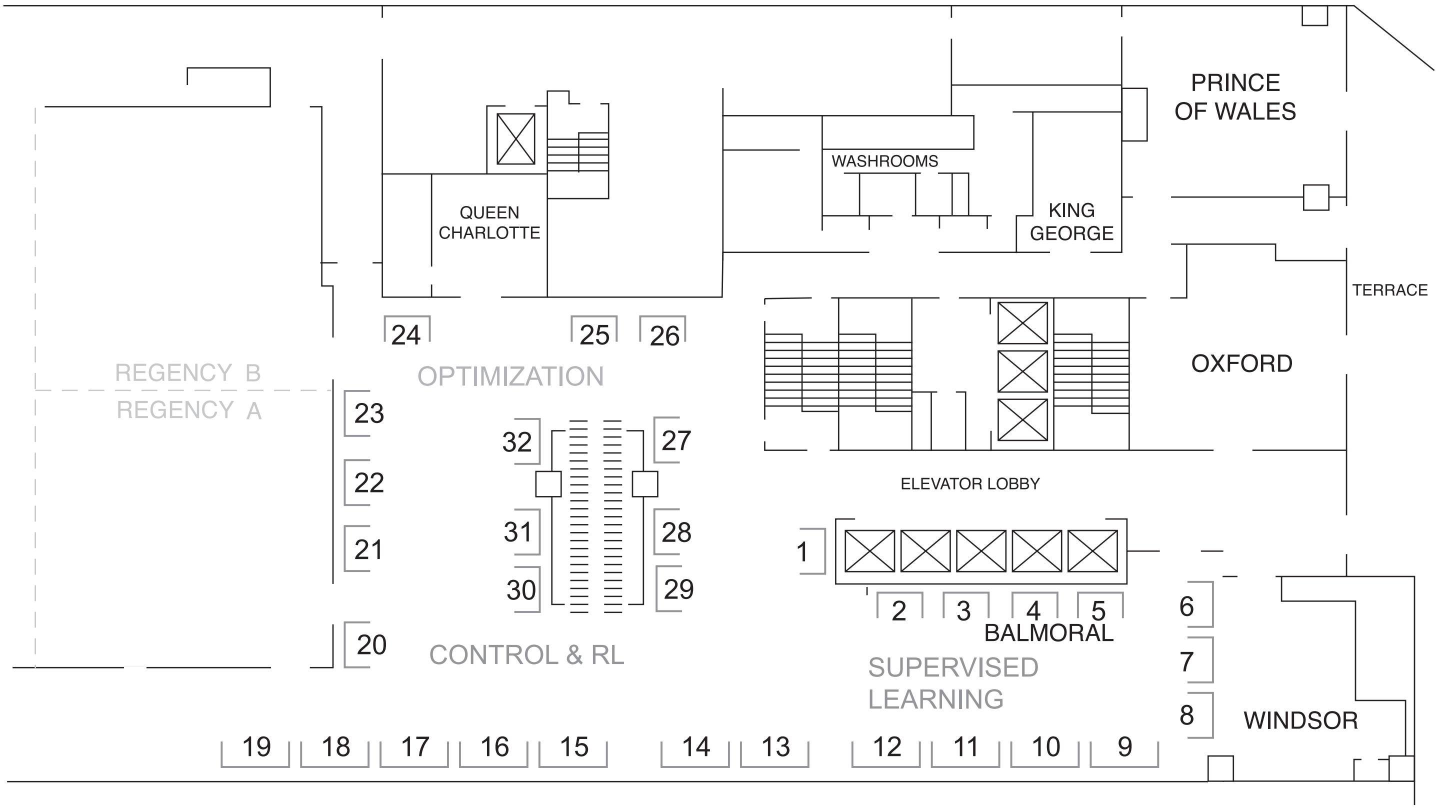
D12 **The SHOGUN Machine Learning Toolbox**, Soeren Sonnenburg, Sonnenburg@tu-berlin.de, Berlin Institute of Technology

D13 **Visual Object Recognition with NN Convolutional & Spiking; Test on Traffic Signs**, Vincent de Ladurantaye, vincent.de.ladurantaye@usherbrooke.ca, Universite de Sherbrooke

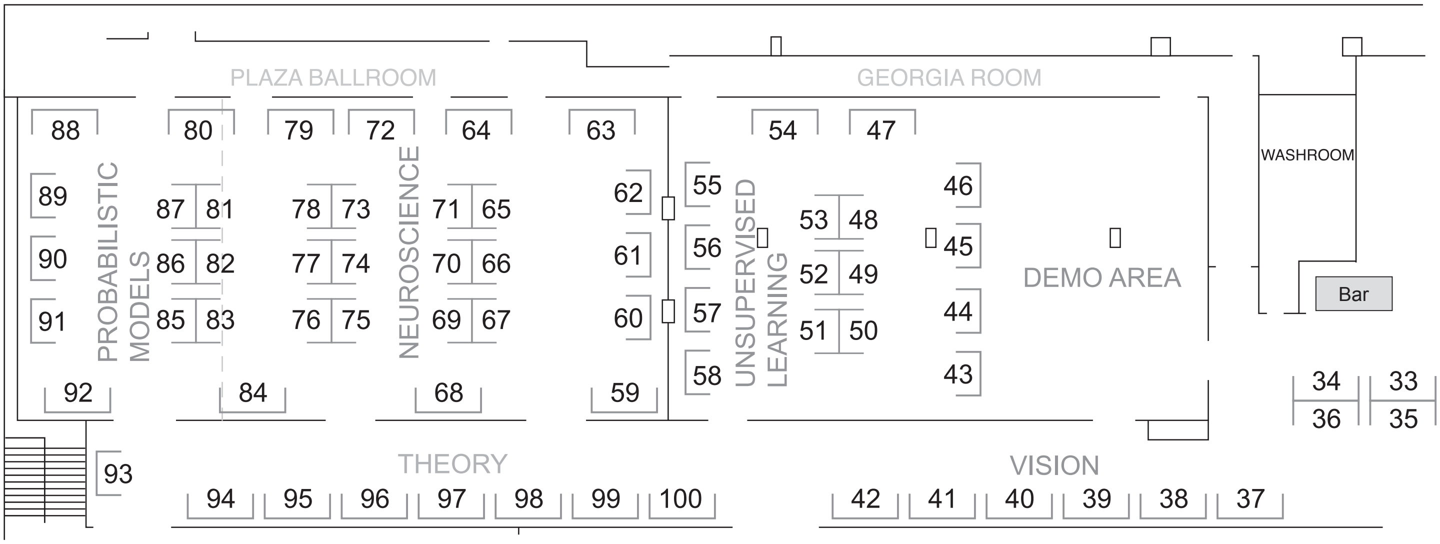

David Wingate
Stochastic Matlab
Achim Hornecker
BCI Demonstration using a Dry-Electrode
Francois Fleuret
Platform to Share Feature Extraction Methods
1b
2b
3b
corridor
NIPS Demo Session Wednesday
6b
Novi Quadrianto
Globby: It's a Search Engine with a Sorting View
5b
4b
Vincent de Ladurantaye
Visual Object Recognition with NN Convolutional & Spiking; Test on Traffic Signs
Soeren Sonnenburg
The SHOGUN Machine Learning Toolbox

NIPS POSTER BOARDS - WEDNESDAY, DECEMBER 8TH

CONVENTION LEVEL (3RD FLOOR)



PLAZA LEVEL (2ND FLOOR)



W1 A Primal-Dual Algorithm for Group Sparse Regularization with Overlapping Groups

sofia mosci mosci@disi.unige.it disi
Silvia Villa villa@dima.unige.it
Alessandro Verri verri@disi.unige.it
DISI Universita' di Genova
Lorenzo Rosasco lrosasco@mit.edu
MIT and IIT

We deal with the problem of variable selection when variables must be selected group-wise with possibly overlapping groups defined a priori. In particular we propose a new optimization procedure for solving the regularized algorithm presented in Jacob et al. 09 where the group lasso penalty is generalized to overlapping groups of variables. While in Jacob et al. 09 the proposed implementation requires explicit replication of the variables belonging to more than one group our iterative procedure is based on a combination of proximal methods in the primal space and constrained Newton method in a reduced dual space corresponding to the active groups. This procedure provides a scalable alternative with no need for data duplication and allows to deal with high dimensional problems without pre-processing to reduce the dimensionality of the data. The computational advantages of our scheme with respect to state-of-the-art algorithms using data duplication are shown empirically with numerical simulations.
Subject Area: Supervised Learning

W2 A Family of Penalty Functions for Structured Sparsity

Charles A Micchelli cmicchel@cityu.edu.hk
City Univ. of Hong Kong
Jean Morales j morales@cs.ucl.ac.uk
Massi Pontil m.pontil@cs.ucl.ac.uk
University College London

We study the problem of learning a sparse linear regression vector under additional conditions on the structure of its sparsity pattern. We present a family of convex penalty functions, which encode this prior knowledge by means of a set of constraints on the absolute values of the regression coefficients. This family subsumes the `1 norm and is flexible enough to include different models of sparsity patterns, which are of practical and theoretical importance. We establish some important properties of these functions and discuss some examples where they can be computed explicitly. Moreover, we present a convergent optimization algorithm for solving regularized least squares with these penalty functions. Numerical simulations highlight the benefit of structured sparsity and the advantage offered by our approach over the Lasso and other related methods.
Subject Area: Supervised Learning
Spotlight presentation, Wednesday, 11:55.

W3 On Herding and the Perceptron Cycling Theorem

Andrew E Gelfand agelfand@uci.edu
Max Welling welling@ics.uci.edu
Univ. of California Irvine
Yutian Chen YUTIANC@ICS.UCI.EDU
Laurens van der Maaten lvdmaaten@gmail.com
University of California San Diego

The paper develops a connection between traditional perceptron algorithms and recently introduced herding algorithms. It is shown that both algorithms can be viewed as an application of the perceptron cycling theorem. This connection strengthens some herding results and suggests new (supervised) herding algorithms that, like CRFs or discriminative RBMs, make predictions by conditioning on the input attributes. We develop and investigate variants of conditional herding, and show that conditional herding leads to practical algorithms that perform better than or on par with related classifiers such as the voted perceptron and the discriminative RBM.
Subject Area: Supervised Learning

W4 Collaborative Filtering in a Non-Uniform World: Learning with the Weighted Trace Norm

Ruslan Salakhutdinov rsalakhu@mit.edu
MIT
Nathan Srebro nati@ttic.edu
TTI

We show that matrix completion with trace-norm regularization can be significantly hurt when entries of the matrix are sampled non-uniformly, but that a properly weighted version of the trace-norm regularizer works well with non-uniform sampling. We show that the weighted trace-norm regularization indeed yields significant gains on the highly non-uniformly sampled Netflix dataset.
Subject Area: Unsupervised & Semi-supervised Learning

W5 Learning Multiple Tasks with a Sparse Matrix-Normal Penalty

Yi Zhang yizhang1@cs.cmu.edu
Jeff Schneider schneide@cs.cmu.edu
Carnegie Mellon University

In this paper, we propose a matrix-variate normal penalty with sparse inverse covariances to couple multiple tasks. Learning multiple (parametric) models can be viewed as estimating a matrix of parameters, where rows and columns of the matrix correspond to tasks and features, respectively. Following the matrix-variate normal density, we design a penalty that decomposes the full covariance of matrix elements into the Kronecker product of row covariance and column covariance, which characterizes both task relatedness and feature representation. Several recently proposed methods are variants of the special cases of this formulation. To address the overfitting issue and select meaningful task and feature structures, we include sparse covariance selection into our matrix-

normal regularization via L-1 penalties on task and feature inverse covariances. We empirically study the proposed method and compare with related models in two real-world problems: detecting landmines in multiple fields and recognizing faces between different subjects. Experimental results show that the proposed framework provides an effective and flexible way to model various different structures of multiple tasks.
Subject Area: Supervised Learning

W6 Learning Multiple Tasks using Manifold Regularization

Arvind Agarwal arvind385@gmail.com
Hal Daume III hal@cs.utah.edu
Samuel Gerber sgerber@cs.utah.edu
University of Utah

We present a novel method for multitask learning (MTL) based on manifold regularization: assume that all task parameters lie on a manifold. This is the generalization of a common assumption made in the existing literature: task parameters share a common linear subspace. One proposed method uses the projection distance from the manifold to regularize the task parameters. The manifold structure and the task parameters are learned using an alternating optimization framework. When the manifold structure is fixed, our method decomposes across tasks which can be learnt independently. An approximation of the manifold regularization scheme is presented that preserves the convexity of the single task learning problem, and makes the proposed MTL framework efficient and easy to implement. We show the efficacy of our method on several datasets.
Subject Area: Supervised Learning

W7 Label Embedding Trees for Large Multi-Class Tasks

Samy Bengio bengio@google.com
Jason Weston jaseweston@gmail.com
Google David Grangier dgrangier@nec-labs.com
NEC Labs America

Multi-class classification becomes challenging at test time when the number of classes is very large and testing against every possible class can become computationally infeasible. This problem can be alleviated by imposing (or learning) a structure over the set of classes. We propose an algorithm for learning a tree-structure of classifiers which, by optimizing the overall tree loss, provides superior accuracy to existing tree labeling methods. We also propose a method that learns to embed labels in a low dimensional space that is faster than non-embedding approaches and has superior accuracy to existing embedding approaches. Finally we combine the two ideas resulting in the label embedding tree that outperforms alternative methods including One-vs-Rest while being orders of magnitude faster.
Subject Area: Supervised Learning

W8 Learning via Gaussian Herding

Koby Crammer koby@ee.technion.ac.il
Technion Daniel Lee ddlee@seas.upenn.edu
University of Pennsylvania

We introduce a new family of online learning algorithms based upon constraining the velocity flow over a distribution of weight vectors. In particular we show how to effectively herd a Gaussian weight vector distribution by trading off velocity constraints with a loss function. By uniformly bounding this loss function we demonstrate how to solve the resulting optimization analytically. We compare the resulting algorithms on a variety of real world datasets and demonstrate how these algorithms achieve state-of-the-art robust performance especially with high label noise in the training data.
Subject Area: Supervised Learning

W9 A Novel Kernel for Learning a Neuron Model from Spike Train Data

Nicholas K Fisher nfisher@cise.ufl.edu
Arunava Banerjee arunava@cise.ufl.edu
University of Florida

From a functional viewpoint, a spiking neuron is a device that transforms input spike trains on its various synapses into an output spike train on its axon. We demonstrate in this paper that the function mapping underlying the device can be tractably learned based on input and output spike train data alone. We begin by posing the problem in a classification based framework. We then derive a novel kernel for an SRM0 model that is based on PSP and AHP like functions. With the kernel we demonstrate how the learning problem can be posed as a Quadratic Program. Experimental results demonstrate the strength of our approach.
Subject Area: Supervised Learning

W10 More data means less inference: A pseudo-max approach to structured learning

David Sontag dsontag@csail.mit.edu MIT
Ofer Meshi meshi@cs.huji.ac.il
Tommi Jaakkola tommi@csail.mit.edu MIT
Amir Globerson gamir@cs.huji.ac.il
Hebrew University

The problem of learning to predict structured labels is of key importance in many applications. However, for general graph structure both learning and inference in this setting are intractable. Here we show that it is possible to circumvent this difficulty when the input distribution is rich enough via a method similar in spirit to pseudo-likelihood. We show how our new method achieves consistency, and illustrate empirically that it indeed performs as well as exact methods when sufficiently large training sets are used.
Subject Area: Supervised Learning
Spotlight presentation, Wednesday, 5:55.

W11 Identifying graph-structured activation patterns in networks

James L Sharpnack and Aarti Singh, Carnegie Mellon U.
Subject Area: Supervised Learning
See abstract, page 78

W12 Penalized Principal Component Regression on Graphs for Analysis of Subnetworks

Ali Shojaie shojaie@umich.edu
George Michailidis gmichail@umich.edu
University of Michigan

Network models are widely used to capture interactions among component of complex systems, such as social and biological. To understand their behavior, it is often necessary to analyze functionally related components of the system, corresponding to subsystems. Therefore, the analysis of subnetworks may provide additional insight into the behavior of the system, not evident from individual components. We propose a novel approach for incorporating available network information into the analysis of arbitrary subnetworks. The proposed method offers an efficient dimension reduction strategy using Laplacian eigenmaps with Neumann boundary conditions, and provides a flexible inference framework for analysis of subnetworks, based on a group-penalized principal component regression model on graphs. Asymptotic properties of the proposed inference method, as well as the choice of the tuning parameter for control of the false positive rate are discussed in high dimensional settings. The performance of the proposed methodology is illustrated using simulated and real data examples from biology.
Subject Area: Supervised Learning

W13 Feature Set Embedding for Incomplete Data

David Grangier dgrangier@nec-labs.com
Iain Melvin iain@nec-labs.com
NEC Laboratories America

We present a new learning strategy for classification problems in which train and/or test data suffer from missing features. In previous work, instances are represented as vectors from some feature space and one is forced to impute missing values or to consider an instance-specific subspace. In contrast, our method considers instances as sets of (feature,value) pairs which naturally handle the missing value case. Building onto this framework, we propose a classification strategy for sets. Our proposal maps (feature,value) pairs into an embedding space and then non-linearly combines the set of embedded vectors. The embedding and the combination parameters are learned jointly on the final classification objective. This simple strategy allows great flexibility in encoding prior knowledge about the features in the embedding step and yields advantageous results compared to alternative solutions over several datasets.
Subject Area: Supervised Learning
Spotlight presentation, Wednesday, 5:50.

W14 Learning to combine foveal glimpses with a third-order Boltzmann machine

Hugo Larochelle and Geoffrey Hinton, University of Toronto.
Subject Area: Supervised Learning
See abstract, page 109

W15 Feature Construction for Inverse Reinforcement Learning

Sergey Levine svlevine@stanford.edu
Vladlen Koltun vladlen@stanford.edu
Stanford University
Zoran Popovic zoran@cs.washington.edu
University of Washington

The goal of inverse reinforcement learning is to find a reward function for a Markov decision process, given example traces from its optimal policy. Current IRL techniques generally rely on user-supplied features that form a concise basis for the reward. We present an algorithm that instead constructs reward features from a large collection of component features, by building logical conjunctions of those component features that are relevant to the example policy. Given example traces, the algorithm returns a reward function as well as the constructed features. The reward function can be used to recover a full, deterministic, stationary policy, and the features can be used to transplant the reward function into any novel environment on which the component features are well defined.
Subject Area: Control and Reinforcement Learning

W16 Distributionally Robust Markov Decision Processes

Huan Xu huan.xu@mail.utexas.edu
The University of Texas
Shie Mannor shie@ee.technion.ac.il
Technion

We consider Markov decision processes where the values of the parameters are uncertain. This uncertainty is described by a sequence of nested sets (that is, each set contains the previous one), each of which corresponds to a probabilistic guarantee for a different confidence level so that a set of admissible probability distributions of the unknown parameters is specified. This formulation models the case where the decision maker is aware of and wants to exploit some (yet imprecise) a-priori information of the distribution of parameters, and arises naturally in practice where methods to estimate the confidence region of parameters abound. We propose a decision criterion based on *distributional robustness*: the optimal policy maximizes the expected total reward under the most adversarial probability distribution over realizations of the uncertain parameters that is admissible (i.e. it agrees with the a-priori information). We show that finding the optimal distributionally robust policy can be reduced to a standard robust MDP where the parameters belong to a single uncertainty set hence it can be computed in polynomial time under mild technical conditions.
Subject Area: Control and Reinforcement Learning

W17 A Reduction from Apprenticeship Learning to Classification

Umar Syed usyed@cis.upenn.edu
University of Pennsylvania
Robert E Schapire schapire@cs.princeton.edu
Princeton University

We provide new theoretical results for apprenticeship learning, a variant of reinforcement learning in which the true reward function is unknown, and the goal is to perform well relative to an observed expert. We study a common approach to learning from expert demonstrations: using a classification algorithm to learn to imitate the expert's behavior. Although this straightforward learning strategy is widely-used in practice, it has been subject to very little formal analysis. We prove that, if the learned classifier has error rate ε, the difference between the value of the apprentice's policy and the expert's policy is $O(\sqrt{\varepsilon})$. Further, we prove that this difference is only $O(\varepsilon)$ when the expert's policy is close to optimal. This latter result has an important practical consequence: Not only does imitating a near-optimal expert result in a better policy, but far fewer demonstrations are required to successfully imitate such an expert. This suggests an opportunity for substantial savings whenever the expert is known to be good, but demonstrations are expensive or difficult to obtain.
Subject Area: Control and Reinforcement Learning

W18 A POMDP Extension with Belief-dependent Rewards

Mauricio A Araya mauricio.araya@loria.fr
Olivier Buffet olivier.buffet@loria.fr
Vincent Thomas vincent.thomas@loria.fr
Francois Charpillet francois.charpillet@loria.fr
Nancy University / INRIA / CNRS

Partially Observable Markov Decision Processes (POMDPs) model sequential decision-making problems under uncertainty and partial observability. Unfortunately some problems cannot be modeled with state-dependent reward functions e.g. problems whose objective explicitly implies reducing the uncertainty on the state. To that end we introduce rho-POMDPs an extension of POMDPs where the reward function rho depends on the belief state. We show that under the common assumption that rho is convex the value function is also convex what makes it possible to (1) approximate rho arbitrarily well with a piecewise linear and convex (PWLC) function and (2) use state-of-the-art exact or approximate solving algorithms with limited changes.
Subject Area: Control and Reinforcement Learning

W19 On a Connection between Importance Sampling and the Likelihood Ratio Policy Gradient

Tang Jie jietang@eecs.berkeley.edu
UC Berkeley
Pieter Abbeel pabbeel@cs.berkeley.edu
University of California

Likelihood ratio policy gradient methods have been some of the most successful reinforcement learning algorithms, especially for learning on physical systems. We describe how the likelihood ratio policy gradient can be derived from an importance sampling perspective. This derivation highlights how likelihood ratio methods under-use past experience by (a) using the past experience to estimate only the gradient of the expected return $U(\Theta)$ at the current policy parameterization Θ, rather than to obtain a more complete estimate of $U(\Theta)$, and (b) using past experience under the current policy only rather than using all past experience to improve the estimates. We present a new policy search method, which leverages both of these observations as well as generalized baselines—a new technique which generalizes commonly used baseline techniques for policy gradient methods. Our algorithm outperforms standard likelihood ratio policy gradient algorithms on several testbeds.
Subject Area: Control and Reinforcement Learning
Spotlight presentation, Wednesday, 5:10.

W20 Learning from Logged Implicit Exploration Data

Alex Strehl astrehl@facebook.com
Facebook
John Langford jl@yahoo-inc.com
Yahoo Research
Lihong Li lihong@yahoo-inc.com
Yahoo! Research
Sham Kakade skakade@wharton.upenn.edu
University of Pennsylvania

We provide a sound and consistent foundation for the use of nonrandom exploration data in "contextual bandit" or "partially labeled" settings where only the value of a chosen action is learned. The primary challenge in a variety of settings is that the exploration policy in which "offline" data is logged is not explicitly known. Prior solutions here require either control of the actions during the learning process recorded random exploration or actions chosen obliviously in a repeated manner. The techniques reported here lift these restrictions allowing the learning of a policy for choosing actions given features from historical data where no randomization occurred or was logged. We empirically verify our solution on two reasonably sized sets of real-world data obtained from an Internet %online advertising company.
Subject Area: Control and Reinforcement Learning
Spotlight presentation, Wednesday, 5:40.

W21 Policy gradients in linearly-solvable MDPs

Emanuel Todorov todorov@cs.washington.edu
University of Washington

We present policy gradient results within the framework of linearly-solvable MDPs. For the first time, compatible function approximators and natural policy gradients are obtained by estimating the cost-to-go function, rather than the (much larger) state-action advantage function as is necessary in traditional MDPs. We also develop the first compatible function approximators and natural policy gradients for continuous-time stochastic systems.
Subject Area: Control and Reinforcement Learning

W22 Distributed Dual Averaging In Networks

John Duchi jduchi@cs.berkeley.edu
Alekh Agarwal alekhagarwal@gmail.com
Martin Wainwright wainwrig@eecs.berkeley.edu
UC Berkeley

The goal of decentralized optimization over a network is to optimize a global objective formed by a sum of local (possibly nonsmooth) convex functions using only local computation and communication. We develop and analyze distributed algorithms based on dual averaging of subgradients, and we provide sharp bounds on their convergence rates as a function of the network size and topology. Our analysis clearly separates the convergence of the optimization algorithm itself from the effects of communication constraints arising from the network structure. We show that the number of iterations required by our algorithm scales inversely in the spectral gap of the network. The sharpness of this prediction is confirmed both by theoretical lower bounds and simulations for various networks.
Subject Area: Optimization
Spotlight presentation, Wednesday, 11:50.

W23 Online Learning in The Manifold of Low-Rank Matrices

Uri Shalit uri.shalit@mail.huji.ac.il
Daphna Weinshall daphna@cs.huji.ac.il
Hebrew University of Jerusalem
Gal Chechik gal.chechik@gmail.com
Google

When learning models that are represented in matrix forms, enforcing a low-rank constraint can dramatically improve the memory and run time complexity while providing a natural regularization of the model. However naive approaches for minimizing functions over the set of low-rank matrices are either prohibitively time consuming (repeated singular value decomposition of the matrix) or numerically unstable (optimizing a factored representation of the low rank matrix). We build on recent advances in optimization over manifolds and describe an iterative online learning procedure consisting of a gradient step followed by a second-order retraction back to the manifold. While the ideal retraction is hard to compute and so is the projection operator that approximates it we describe another second-order retraction that can be computed efficiently with run time and memory complexity of O((n+m)k) for a rank-k matrix of dimension m x n given rank one gradients. We use this algorithm LORETA to learn a matrix-form similarity measure over pairs of documents represented as high dimensional vectors. LORETA improves the mean average precision over a passive- aggressive approach in a factorized model and also improves over a full model trained over pre-selected features using the same memory requirements. LORETA also showed consistent improvement over standard methods in a large (1600 classes) multi-label image classification task.
Subject Area: Optimization
Spotlight presentation, Wednesday, 3:20.

W24 Moreau-Yosida Regularization for Grouped Tree Structure Learning

Jun Liu j.liu@asu.edu
Jieping Ye jieping.ye@asu.edu
Arizona State University

We consider the tree structured group Lasso where the structure over the features can be represented as a tree with leaf nodes as features and internal nodes as clusters of the features. The structured regularization with a pre-defined tree structure is based on a group-Lasso penalty, where one group is defined for each node in the tree. Such a regularization can help uncover the structured sparsity, which is desirable for applications with some meaningful tree structures on the features. However, the tree structured group Lasso is challenging to solve due to the complex regularization. In this paper, we develop an efficient algorithm for the tree structured group Lasso. One of the key steps in the proposed algorithm is to solve the Moreau-Yosida regularization associated with the grouped tree structure. The main technical contributions of this paper include (1) we show that the associated Moreau-Yosida regularization admits an analytical solution, and (2) we develop an efficient algorithm for determining the effective interval for the regularization parameter. Our experimental results on the AR and JAFFE face data sets demonstrate the efficiency and effectiveness of the proposed algorithm.
Subject Area: Optimization

W25 A Log-Domain Implementation of the Diffusion Network in Very Large Scale Integration

Yi-Da Wu yidawu@gmail.com
Shi-Jie Lin g9563509@oz.nthu.edu.tw
Hsin Chen hchne@ee.nthu.edu.tw
National Tsing Hua

University The Diffusion Network(DN) is a stochastic recurrent network which has been shown capable of modeling the distributions of continuous-valued, continuous-time paths. However, the dynamics of the DN are governed by stochastic differential equations, making the DN unfavourable for simulation in a digital computer. This paper presents the implementation of the DN in analogue Very Large Scale Integration, enabling the DN to be simulated in real time. Moreover, the log-domain representation is applied to the DN, allowing the supply voltage and thus the power consumption to be reduced without limiting the dynamic ranges for diffusion processes. A VLSI chip containing a DN with two stochastic units has been designed and fabricated. The design of component circuits will be described, so will the simulation of the full system be presented. The simulation results demonstrate that the DN in VLSI is able to regenerate various types of continuous paths in real-time.

Subject Area: Hardware

W26 Phoneme Recognition with Large Hierarchical Reservoirs

Fabian Triefenbach, Azarakhsh Jalalvand, Benjamin Schrauwen and Jean-Pierre Martens
Ghent University ELIS

Subject Area: Speech and Signal Processing

See abstract, page 78

W27 Multivariate Dyadic Regression Trees for Sparse Learning Problems

Han Liu hanliu@cs.cmu.edu
Xi Chen xichen@cs.cmu.edu
Carnegie Mellon University

We propose a new nonparametric learning method based on multivariate dyadic regression trees (MDRTs). Unlike traditional dyadic decision trees (DDTs) or classification and regression trees (CARTs), MDRTs are constructed using penalized empirical risk minimization with a novel sparsity-inducing penalty. Theoretically, we show that MDRTs can simultaneously adapt to the unknown sparsity and smoothness of the true regression functions, and achieve the nearly optimal rates of convergence (in a minimax sense) for the class of (α,C)-smooth functions. Empirically, MDRTs can simultaneously conduct function estimation and variable selection in high dimensions. To make MDRTs applicable for large-scale learning problems we propose a greedy heuristics. The superior performance of MDRTs are demonstrated on both synthetic and real datasets.

Subject Area: Supervised Learning

W28 Sparse Coding for Learning Interpretable Spatio-Temporal Primitives

Taehwan Kim taehwan@ttic.edu
Gregory Shakhnarovich greg@ttic.edu
Raquel Urtasun rurtasun@ttic.edu
TTI Chicago

Sparse coding has recently become a popular approach in computer vision to learn dictionaries of natural images. In this paper we extend sparse coding to learn interpretable spatio-temporal primitives of human motion. We cast the problem of learning spatio-temporal primitives as a tensor factorization problem and introduce constraints to learn interpretable primitives. In particular, we use group norms over those tensors, diagonal constraints on the activations as well as smoothness constraints that are inherent to human motion. We demonstrate the effectiveness of our approach to learn interpretable representations of human motion from motion capture data and show that our approach outperforms recently developed matching pursuit and sparse coding algorithms.

Subject Area: Supervised Learning

W29 Efficient and Robust Feature Selection via Joint l2,1-Norms Minimization

Feiping Nie feipingnie@gmail.com
Heng Huang heng@uta.edu
Chris Ding chqding@uta.edu
University of Texas at Arlington
Xiao Cai xiao.cai@mavs.uta.edu
UTA

Feature selection is an important component of many machine learning applications. Especially in many bioinformatics tasks, efficient and robust feature selection methods are desired to extract meaningful features and eliminate noisy ones. In this paper, we propose a new robust feature selection method with emphasizing joint $l_{2,1}$-norm minimization on both loss function and regularization. The l2,1-norm based loss function is robust to outliers in data points and the $l_{2,1}$-norm regularization selects features across all data points with joint sparsity. An efficient algorithm is introduced with proved convergence. Our regression based objective makes the feature selection process more efficient. Our method has been applied into both genomic and proteomic biomarkers discovery. Extensive empirical studies were performed on six data sets to demonstrate the effectiveness of our feature selection method.

Subject Area: Supervised Learning

W30 Natural Policy Gradient Methods with Parameter-based Exploration for Control Tasks

Atsushi Miyamae miyamae@fe.dis.titech.ac.jp
Yuichi Nagata nagata@fe.dis.titech.ac.jp
Isao Ono isao@dis.titech.ac.jp
Shigenobu Kobayashi kobayasi@dis.titech.ac.jp
Tokyo Institute of Technology

In this paper, we propose an efficient algorithm for estimating the natural policy gradient with parameter-based exploration; this algorithm samples directly in the parameter space. Unlike previous methods based on natural gradients, our algorithm calculates the natural policy gradient using the inverse of the exact Fisher information matrix. The computational cost of this algorithm is equal to that of conventional policy gradients whereas previous natural policy gradient methods have a prohibitive computational cost. Experimental results show that the proposed method outperforms several policy gradient methods.
Subject Area: Control and Reinforcement Learning

W31 Linear Complementarity for Regularized Policy Evaluation and Improvement

Jeffrey T Johns, Christopher Painter-Wakefield and
Ronald Parr, Duke University.
Subject Area: Control and Reinforcement Learning
See abstract, page 76

W32 Large Margin Learning of Upstream Scene Understanding Models

Jun Zhu junzhu@cs.cmu.edu
Eric Xing epxing@cs.cmu.edu
Carnegie Mellon University
Li-Jia Li lijiali@cs.stanford.edu
Li Fei-Fei feifeili@cs.stanford.edu
Stanford University

Upstream supervised topic models have been widely used for complicated scene understanding. However, existing maximum likelihood estimation (MLE) schemes can make the prediction model learning independent of latent topic discovery and result in an imbalanced prediction rule for scene classification. This paper presents a joint max-margin and max-likelihood learning method for upstream scene understanding models, in which latent topic discovery and prediction model estimation are closely coupled and well-balanced. The optimization problem is efficiently solved with a variational EM procedure, which iteratively solves an online loss-augmented SVM. We demonstrate the advantages of the large-margin approach on both an 8-category sports dataset and the 67-class MIT indoor scene dataset for scene categorization.
Subject Area: Vision

W32 Lower Bounds on Rate of Convergence of Cutting Plane Methods

Xinhua Zhang xinhua.zhang.cs@gmail.com
AICML
Ankan Saha ankans@gmail.com
University of Chicago
S.V.N. Vishwanathan vishy@stat.purdue.edu
Purdue University

In a recent paper Joachims (2006) presented SVM-Perf, a cutting plane method (CPM) for training linear Support Vector Machines (SVMs) which converges to an accurate solution in $O(1/\varepsilon^2)$ iterations. By tightening the analysis, Teo et al. (2010) showed that $O(1/\varepsilon)$ iterations suffice. Given the impressive convergence speed of CPM on a number of practical problems, it was conjectured that these rates could be further improved. In this paper we disprove this conjecture. We present counter examples which are not only applicable for training linear SVMs with hinge loss, but also hold for support vector methods which optimize a multivariate performance score. However, surprisingly, these problems are not inherently hard. By exploiting the structure of the objective function we can devise an algorithm that converges in $O(1\sqrt{\varepsilon})$ iterations.
Subject Area: Optimization

W33 The Multidimensional Wisdom of Crowds

Peter Welinder and Pietro Perona, Caltech, Steve
Branson and Serge Belongie, UC San Diego,
Subject Area: Vision Oral presentation, Wednesday,
See abstract, page 76.

W35 Estimating Spatial Layout of Rooms using Volumetric Reasoning about Objects and Surfaces

David C Lee dclee@cs.cmu.edu
Abhinav Gupta abhinavg@cs.cmu.edu
Martial Hebert hebert@ri.cmu.edu
Takeo Kanade Takeo.Kanade@cs.cmu.edu
Carnegie Mellon University

There has been a recent push in extraction of 3D spatial layout of scenes. However, none of these approaches model the 3D interaction between objects and the spatial layout. In this paper, we argue for a parametric representation of objects in 3D, which allows us to incorporate volumetric constraints of the physical world. We show that augmenting current structured prediction techniques with volumetric reasoning significantly improves the performance of the state-of-the-art.
Subject Area: Vision

W36 Space-Variant Single-Image Blind Deconvolution for Removing Camera Shake

Stefan Harmeling stefan.harmeling@tuebingen.mpg.de
Hirsch Michael michael.hirsch@tuebingen.mpg.de
Bernhard Schoelkopf bernhard.schoelkopf@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics

Modelling camera shake as a space-invariant convolution simplifies the problem of removing camera shake, but often insufficiently models actual motion blur such as those due to camera rotation and movements outside the sensor plane or when objects in the scene have different distances to the camera. In order to overcome such limitations we contribute threefold: (i) we introduce a taxonomy of camera shakes, (ii) we show how to combine a recently introduced framework for space-variant filtering based on overlap-add from Hirsch et al. and a fast algorithm for single image blind deconvolution for space-invariant filters from Cho and Lee to introduce a method for blind deconvolution for space-variant blur. And (iii), we present an experimental setup for evaluation that allows us to take images with real camera shake while at the same time record the space-variant point spread function corresponding to that blur. Finally, we demonstrate that our method is able to deblur images degraded by spatially-varying blur originating from real camera shake.
Subject Area: Vision

W37 Segmentation as Maximum-Weight Independent Set

William Brendel b rendelw@onid.orst.edu
Sinisa Todorovic sinisa@eecs.oregonstate.edu
Oregon State University

Given an ensemble of distinct, low-level segmentations of an image, our goal is to identify visually ”meaningful” segments in the ensemble. Knowledge about any specific objects and surfaces present in the image is not available. The selection of image regions occupied by objects is formalized as the maximum-weight independent set (MWIS) problem. MWIS is the heaviest subset of mutually non-adjacent nodes of an attributed graph. We construct such a graph from all segments in the ensemble. Then, MWIS selects maximally distinctive segments that together partition the image. A new MWIS algorithm is presented. The algorithm seeks a solution directly in the discrete domain, instead of relaxing MWIS to a continuous problem, as common in previous work. It iteratively finds a candidate discrete solution of the Taylor series expansion of the original MWIS objective function around the previous solution. The algorithm is shown to converge to a maximum. Our empirical evaluation on the benchmark Berkeley segmentation dataset shows that the new algorithm eliminates the need for hand-picking optimal input parameters of the state-ofthe- art segmenters, and outperforms their best, manually optimized results.
Subject Area: Vision

W38 Generating more realistic images using gated MRF’s

Marc’Aurelio Ranzato ranzato@cs.toronto.edu
Volodymyr Mnih vmnih@cs.toronto.edu
Geoffrey Hinton hinton@cs.toronto.edu
University of Toronto

Probabilistic models of natural images are usually evaluated by measuring performance on rather indirect tasks such as denoising and inpainting. A more direct way to evaluate a generative model is to draw samples from it and to check whether statistical properties of the samples match the statistics of natural images. This method is seldom used with high-resolution images because current models produce samples that are very different from natural images as assessed by even simple visual inspection. We investigate the reasons for this failure and we show that by augmenting existing models so that there are two sets of latent variables one set modelling pixel intensities and the other set modelling imagespecific pixel covariances we are able to generate high-resolution images that look much more realistic than before. The overall model can be interpreted as a gated MRF where both pair-wise dependencies and mean intensities of pixels are modulated by the states of latent variables. Finally we confirm that if we disallow weight-sharing between receptive fields that overlap each other the gated MRF learns more efficient internal representations as demonstrated in several recognition tasks.
Subject Area: Vision

W39 A Discriminative Latent Model of Image Region and Object Tag Correspondence

Yang Wang ywang12@cs.sfu.ca
Greg Mori mori@cs.sfu.ca
Simon Fraser University

We propose a discriminative latent model for annotating images with unaligned object-level textual annotations. Instead of using the bag-of-words image representation currently popular in the computer vision community, our model explicitly captures more intricate relationships underlying visual and textual information. In particular, we model the mapping that translates image regions to annotations. This mapping allows us to relate image regions to their corresponding annotation terms. We also model the overall scene label as latent information. This allows us to cluster test images. Our training data consist of images and their associated annotations. But we do not have access to the ground-truth region-to-annotation mapping or the overall scene label. We develop a novel variant of the latent SVM framework to model them as latent variables. Our experimental results demonstrate the effectiveness of the proposed model compared with other baseline methods.
Subject Area: Vision

W40 Using body-anchored priors for identifying actions in single images

Leonid Karlinsky leokarlin@gmail.com
Michael Dinerstein dinerstein@gmail.com
Shimon Ullman shimon.ullman@weizmann.ac.il
Weizmann Institute of Science

This paper presents an approach to the visual recognition of human actions using only single images as input. The task is easy for humans but difficult for current approaches to object recognition, because action instances may be similar in terms of body pose, and often require detailed examination of relations between participating objects and body parts in order to be recognized. The proposed approach applies a two-stage interpretation procedure to each training and test image. The first stage produces accurate detection of the relevant body parts of the actor, forming a prior for the local evidence needed to be considered for identifying the action. The second stage extracts features that are ‘anchored’ to the detected body parts, and uses these features and their feature-to-part relations in order to recognize the action. The body anchored priors we propose apply to a large range of human actions. These priors allow focusing on the relevant regions and relations, thereby significantly simplifying the learning process and increasing recognition performance.
Subject Area: Vision

W41 Simultaneous Object Detection and Ranking with Weak Supervision

Matthew B Blaschko blaschko@robots.ox.ac.uk
Andrea Vedaldi vedaldi@robots.ox.ac.uk
Andrew Zisserman az@robots.ox.ac.uk
Oxford University

A standard approach to learning object category detectors is to provide strong supervision in the form of a region of interest (ROI) specifying each instance of the object in the training images. In this work are goal is to learn from heterogeneous labels, in which some images are only weakly supervised, specifying only the presence or absence of the object or a weak indication of object location, whilst others are fully annotated. To this end we develop a discriminative learning approach and make two contributions: (i) we propose a structured output formulation for weakly annotated images where full annotations are treated as latent variables; and (ii) we propose to optimize a ranking objective function allowing our method to more effectively use negatively labeled images to improve detection average precision performance. The method is demonstrated on the benchmark INRIA pedestrian detection dataset of Dalal and Triggs and the PASCAL VOC dataset and it is shown that for a significant proportion of weakly supervised images the performance achieved is very similar to the fully supervised (state of the art) results.
Subject Area: Vision

W42 Feature Transitions with Saccadic Search: Size, Color, and Orientation Are Not Alike

Stella X Yu syu@cs.bc.edu
Boston College

Size, color, and orientation have long been considered elementary features whose attributes are extracted in parallel and available to guide the deployment of attention. If each is processed in the same fashion with simply a different set of local detectors, one would expect similar search behaviours on localizing an equivalent flickering change among identically laid out disks. We analyze feature transitions associated with saccadic search and find out that size, color, and orientation are not alike in dynamic attribute processing over time. The Markovian feature transition is attractive for size, repulsive for color, and largely reversible for orientation.
Subject Area: Vision

W43 Identifying Patients at Risk of Major Adverse Cardiovascular Events Using Symbolic Mismatch

Zeeshan Syed, Univ. of Michigan, and John Guttag, MIT.
Subject Area: Applications
See abstract, page 78

W44 Sphere Embedding: An Application to Part-of-Speech Induction

Yariv Maron syarivm@yahoo.com
Bar Ilan University
Michael Lamar mlamar@slu.edu
Saint Louis University
Elie Bienenstock elie@brown.edu
Brown University

Motivated by an application to unsupervised part-of-speech tagging, we present an algorithm for the Euclidean embedding of large sets of categorical data based on co-occurrence statistics. We use the CODE model of Globerson et al. but constrain the embedding to lie on a high-dimensional unit sphere. This constraint allows for efficient optimization, even in the case of large datasets and high embedding dimensionality. Using k-means clustering of the embedded data, our approach efficiently produces state-of-the-art results. We analyze the reasons why the sphere constraint is beneficial in this application, and conjecture that these reasons might apply quite generally to other large-scale tasks.
Subject Area: Applications

W45 Efficient Optimization for Discriminative Latent Class Models

Armand Joulin armand.joulin@ens.fr
Francis Bach francis.bach@ens.fr
Ecole Normale Superieure
Jean Ponce jean.ponce@ens.fr

Dimensionality reduction is commonly used in the setting of multi-label supervised classification to control the learning capacity and to provide a meaningful representation of the data. We introduce a simple forward probabilistic model which is a multinomial extension of reduced rank regression; we show that this model provides a probabilistic interpretation of discriminative clustering methods with added benefits in terms of number of hyperparameters and optimization. While expectation-maximization (EM) algorithm is commonly used to learn these models its optimization usually leads to local minimum because it relies on a non-convex cost function with many such local minima. To avoid this problem we introduce a local approximation of this cost function which leads to a quadratic non-convex optimization problem over a product of simplices. In order to minimize such functions we propose an efficient algorithm based on convex relaxation and low-rank representation of our data which allows to deal with large instances. Experiments on text document classification show that the new model outperforms other supervised dimensionality reduction methods while simulations on unsupervised clustering show that our probabilistic formulation has better properties than existing discriminative clustering methods.
Subject Area: Unsupervised & Semi-supervised Learning

W46 Sample Complexity of Testing the Manifold Hypothesis

Hariharan Narayanan har@mit.edu
Sanjoy K Mitter mitter@mit.edu
MIT

The hypothesis that high dimensional data tends to lie in the vicinity of a low dimensional manifold is the basis of a collection of methodologies termed Manifold Learning. In this paper we study statistical aspects of the question of fitting a manifold with a nearly optimal least squared error. Given upper bounds on the dimension volume and curvature we show that Empirical Risk Minimization can produce a nearly optimal manifold using a number of random samples that is independent of the ambient dimension of the space in which data lie. We obtain an upper bound on the required number of samples that depends polynomially on the curvature exponentially on the intrinsic dimension and linearly on the intrinsic volume. For constant error we prove a matching minimax lower bound on the sample complexity that shows that this dependence on intrinsic dimension volume and curvature is unavoidable. Whether the known lower bound of $O(\frac{k}{\epsilon^2} + \frac{log \frac{1}{de}}{\epsilon^2})$for the sample complexity of Empirical Risk minimization on k−means applied to data in a unit ball of arbitrary dimension is tight has been an open question since 1997 [bart2]. Here is the desired bound on the error and de is a bound on the probability of failure. We improve the best currently known upper bound [pontil] of $O(\frac{k}{\epsilon^2} + \frac{log^4 \frac{k}{\epsilon}}{\epsilon^2})$ to $O(\frac{k}{\epsilon^2} (min(k \frac{log^4 \frac{k}{\epsilon}}{\epsilon^2}))+ \frac{log \frac{1}{de}}{\epsilon^2})$. Based on these results we devise a simple algorithm for k−means and another that uses a family of convex programs to fit a piecewise linear curve of a specified length to high dimensional data where the sample complexity is independent of the ambient dimension.
Subject Area: Unsupervised & Semi-supervised Learning
Spotlight presentation, Wednesday, 3:25.

W47 Optimal Bayesian Recommendation Sets and Myopically Optimal Choice Query Sets

Paolo Viappiani paolo.viappiani@gmail.com
Craig Boutilier cebly@cs.toronto.edu
University of Toronto

Bayesian approaches to utility elicitation typically adopt (myopic) expected value of information (EVOI) as a natural criterion for selecting queries. However EVOI-optimization is usually computationally prohibitive. In this paper we examine EVOI optimization using choice queries queries in which a user is ask to select her most preferred product from a set. We show that under very general assumptions the optimal choice query w.r.t. EVOI coincides with optimal recommendation set that is a set maximizing expected utility of the user selection. Since recommendation set optimization is a simpler submodular problem this can greatly reduce the complexity of both exact and approximate (greedy) computation of optimal choice queries. We also examine the case where user responses to choice queries are error-prone (using both constant and follow mixed multinomial logit noise models) and provide worst-case guarantees. Finally we present a local search technique that works well with large outcome spaces.
Subject Area: Unsupervised & Semi-supervised Learning
Spotlight presentation, Wednesday, 9:40.

W48 Extensions of Generalized Binary Search to Group Identification and Exponential Costs

Gowtham Bellala gowtham@umich.edu
Clayton Scott clayscot@umich.edu
University of Michigan
Suresh Bhavnani skbhavnani@gmail.com
University of Texas

Generalized Binary Search (GBS) is a well known greedy algorithm for identifying an unknown object while minimizing the number of ”yes” or ”no” questions posed about that object, and arises in problems such as active learning and active diagnosis. Here, we provide a coding-theoretic interpretation for GBS and show that GBS can be viewed as a top-down algorithm that greedily minimizes the expected number of queries required to identify an object. This interpretation is then used to extend GBS in two ways. First, we consider the case where the objects are partitioned into groups, and the objective is to identify only the group to which the object belongs. Then, we consider the case where the cost of identifying an object grows exponentially in the number of queries. In each case, we present an exact formula for the objective function involving Shannon or Renyi entropy, and develop a greedy algorithm for minimizing it.
Subject Area: Unsupervised & Semi-supervised Learning

W49 Active Learning by Querying Informative and Representative Examples

Sheng-Jun Huang huangsj@lamda.nju.edu.cn
Zhi-Hua Zhou zhouzh@lamda.nju.edu.cn
Nanjing University
Rong Jin rongjin@cse.msu.edu
Michigan State University

Most active learning approaches select either informative or representative unlabeled instances to query their labels. Although several active learning algorithms have been proposed to combine the two criterions for query selection, they are usually ad hoc in finding unlabeled instances that are both informative and representative. We address this challenge by a principled approach, termed QUIRE, based on the min-max view of active learning. The proposed approach provides a systematic way for measuring and combining the informativeness and representativeness of an instance. Extensive experimental results show that the proposed QUIRE approach outperforms several state-of -the-art active learning approaches.

Subject Area: Unsupervised & Semi-supervised Learning

W50 Active Instance Sampling via Matrix Partition

Yuhong Guo yuhong@temple.edu
Temple University

Recently, batch-mode active learning has attracted a lot of attention. In this paper, we propose a novel batch-mode active learning approach that selects a batch of queries in each iteration by maximizing a natural form of mutual information criterion between the labeled and unlabeled instances. By employing a Gaussian process framework, this mutual information based instance selection problem can be formulated as a matrix partition problem. Although the matrix partition is an NP-hard combinatorial optimization problem, we show a good local solution can be obtained by exploiting an effective local optimization technique on the relaxed continuous optimization problem. The proposed active learning approach is independent of employed classification models. Our empirical studies show this approach can achieve comparable or superior performance to discriminative batch-mode active learning methods.

Subject Area: Unsupervised & Semi-supervised Learning

W51 Robust Clustering as Ensembles of Affinity Relations

Hairong Liu lhrbss@gmail.com
Shuicheng Yan eleyans@nus.edu.sg
National University of Singapore
Longin Jan Latecki latecki@temple.edu
Temple University

In this paper, we regard clustering as ensembles of k-ary affinity relations and clusters correspond to subsets of objects with maximal average affinity relations. The average affinity relation of a cluster is relaxed and well approximated by a constrained homogenous function. We present an efficient procedure to solve this optimization problem and show that the underlying clusters can be robustly revealed by using priors systematically constructed from the data. Our method can automatically select some points to form clusters leaving other points un-grouped; thus it is inherently robust to large numbers of outliers which has seriously limited the applicability of classical methods. Our method also provides a unified solution to clustering from k-ary affinity relations with $k \geq 2$ that is it applies to both graph-based and hypergraph-based clustering problems. Both theoretical analysis and experimental results show the superiority of our method over classical solutions to the clustering problem especially when there exists a large number of outliers.

Subject Area: Unsupervised & Semi-supervised Learning

W52 Supervised Clustering

Pranjal Awasthi pranjal.iitm@gmail.com
CMU
Reza Bosagh Zadeh rezab@stanford.edu
Stanford University

Despite the ubiquity of clustering as a tool in unsupervised learning, there is not yet a consensus on a formal theory and the vast majority of work in this direction has focused on unsupervised clustering. We study a recently proposed framework for supervised clustering where there is access to a teacher. We give an improved generic algorithm to cluster any concept class in that model. Our algorithm is query-efficient in the sense that it involves only a small amount of interaction with the teacher. We also present and study two natural generalizations of the model. The model assumes that the teacher response to the algorithm is perfect. We eliminate this limitation by proposing a noisy model and give an algorithm for clustering the class of intervals in this noisy model. We also propose a dynamic model where the teacher sees a random subset of the points. Finally for datasets satisfying a spectrum of weak to strong properties we give query bounds and show that a class of clustering functions containing Single-Linkage will find the target clustering under the strongest property.

Subject Area: Unsupervised & Semi-supervised Learning
Spotlight presentation, Wednesday, 3:15.

W53 Rates of convergence for the cluster tree

Kamalika Chaudhuri kamalika@cs.ucsd.edu
Sanjoy Dasgupta dasgupta@cs.ucsd.edu
UC San Diego

For a density f on Rd, a high-density cluster is any connected component of $\{x : f(x) \geq c\}$, for some $c > 0$. The set of all high-density clusters form a hierarchy called the cluster tree of f. We present a procedure for estimating the cluster tree given samples from f. We give finite-sample convergence rates for our algorithm, as well as lower bounds on the sample complexity of this estimation problem.

Subject Area: Unsupervised & Semi-supervised Learning

W54 Random Projection Trees Revisited

aman Dhesi adhesi@princeton.edu
Princeton University
Purushottam Kar purushot@cse.iitk.ac.in
Indian Institute of Technology

The Random Projection Tree (RPTree) structures proposed in [Dasgupta-Freund-STOC- 08] are space partitioning data structures that automatically adapt to various notions of intrinsic dimensionality of data. We prove new results for both the RPTree-Max and the RPTree-Mean data structures. Our result for RPTree-Max gives a near-optimal bound on the number of levels required by this data structure to reduce the size of its cells by a factor s >= 2. We also prove a packing lemma for this data structure. Our final result shows that low-dimensional manifolds possess bounded Local Covariance Dimension. As a consequence we show that RPTree-Mean adapts to manifold dimension as well.
Subject Area: Unsupervised & Semi-supervised Learning

W55 Worst-Case Linear Discriminant Analysis

Yu Zhang zhangyu@cse.ust.hk
Dit-Yan Yeung dyyeung@cse.ust.hk
HKUST

Dimensionality reduction is often needed in many applications due to the high dimensionality of the data involved. In this paper, we first analyze the scatter measures used in the conventional linear discriminant analysis (LDA) model and note that the formulation is based on the average-case view. Based on this analysis, we then propose a new dimensionality reduction method called worst-case linear discriminant analysis (WLDA) by defining new between-class and within-class scatter measures. This new model adopts the worst-case view which arguably is more suitable for applications such as classification. When the number of training data points or the number of features is not very large, we relax the optimization problem involved and formulate it as a metric learning problem. Otherwise, we take a greedy approach by finding one direction of the transformation at a time. Moreover, we also analyze a special case of WLDA to show its relationship with conventional LDA. Experiments conducted on several benchmark datasets demonstrate the effectiveness of WLDA when compared with some related dimensionality reduction methods.
Subject Area: Unsupervised & Semi-supervised Learning

W56 Guaranteed Rank Minimization via Singular Value Projection

Prateek Jain pjain9@gmail.com
Microsoft Research India Lab
Raghu Meka raghu@cs.utexas.edu
Inderjit Dhillon inderjit@cs.utexas.edu
University of Texas Austin

Minimizing the rank of a matrix subject to affine constraints is a fundamental problem with many important applications in machine learning and statistics. In this paper we propose a simple and fast algorithm SVP (Singular Value Projection) for rank minimization under affine constraints ARMP and show that SVP recovers the minimum rank solution for affine constraints that satisfy a Restricted Isometry Property (RIP). Our method guarantees geometric convergence rate even in the presence of noise and requires strictly weaker assumptions on the RIP constants than the existing methods. We also introduce a Newton-step for our SVP framework to speed-up the convergence with substantial empirical gains. Next, we address a practically important application of ARMP - the problem of low-rank matrix completion, for which the defining affine constraints do not directly obey RIP, hence the guarantees of SVP do not hold. However, we provide partial progress towards a proof of exact recovery for our algorithm by showing a more restricted isometry property and observe empirically that our algorithm recovers low-rank Incoherent matrices from an almost optimal number of uniformly sampled entries. We also demonstrate empirically that our algorithms outperform existing methods, such as those of [CaiCS2008,LeeB2009b, KeshavanOM2009], for ARMP and the matrix completion problem by an order of magnitude and are also more robust to noise and sampling schemes. In particular, results show that our SVP-Newton method is significantly robust to noise and performs impressively on a more realistic power-law sampling scheme for the matrix completion problem.
Subject Area: Unsupervised & Semi-supervised Learning
Spotlight presentation, Wednesday, 12:00.

W57 An Inverse Power Method for Nonlinear Eigenproblems with Applications in 1-Spectral Clustering and Sparse PCA

Matthias Hein hein@cs.uni-sb.de
Thomas Buhler tb@cs.uni-saarland.de
Saarland University

Many problems in machine learning and statistics can be formulated as (generalized) eigenproblems. In terms of the associated optimization problem computing linear eigenvectors amounts to finding critical points of a quadratic function subject to quadratic constraints. In this paper we show that a certain class of constrained optimization problems with nonquadratic objective and constraints can be understood as nonlinear eigenproblems. We derive a generalization of the inverse power method which is guaranteed to converge to a nonlinear eigenvector. We apply the inverse power method to 1-spectral clustering and sparse PCA which can naturally be formulated as nonlinear eigenproblems. In both applications we achieve state-of-the-art results in terms of solution quality and runtime. Moving beyond the standard eigenproblem should be useful also in many other applications and our inverse power method can be easily adapted to new problems.
Subject Area: Unsupervised & Semi-supervised Learning

W58 Word Features for Latent Dirichlet Allocation

James Petterson james.petterson@nicta.com.au
Tiberio Caetano Tiberio.Caetano@nicta.com.au
Wray Buntine wray.buntine@nicta.com.au
NICTA
Alexander J Smola alex@smola.org
Yahoo! Research
Shravan M Narayanamurthy shravanm@yahoo-inc.com
Yahoo! Labs Bangalore

We extend Latent Dirichlet Allocation (LDA) by explicitly allowing for the encoding of side information in the distribution over words. This results in a variety of new capabilities, such as improved estimates for infrequently occurring words, as well as the ability to leverage thesauri and dictionaries in order to boost topic cohesion within and across languages. We present experiments on multi-language topic synchronisation where dictionary information is used to bias corresponding words towards similar topics. Results indicate that our model substantially improves topic cohesion when compared to the standard LDA model.
Subject Area: Unsupervised & Semi-supervised Learning

W59 A novel family of non-parametric cumulative based divergences for point processes

Sohan Seth sohan@cnel.ufl.edu
Park Il memming@cnel.ufl.edu
Austin Brockmeier ajbrockmeier@ufl.edu
University of Florida
Mulugeta Semework mulugeta.semework@downstate.edu
John Choi john.choi@downstate.edu
Joseph T Francis joe.francis@downstate.edu
SUNY Downstate Medical Center

SUNY Downstate and NYU-Poly Jose Principe principe@ cnel.ufl.edu University of Florida at Gainesville Hypothesis testing on point processes has several applications such as model fitting, plasticity detection, and non-stationarity detection. Standard tools for hypothesis testing include tests on mean firing rate and time varying rate function. However, these statistics do not fully describe a point process and thus the tests can be misleading. In this paper, we introduce a family of non-parametric divergence measures for hypothesis testing. We extend the traditional Kolmogorov–Smirnov and Cramer–von-Mises tests for point process via stratification. The proposed divergence measures compare the underlying probability structure and, thus, is zero if and only if the point processes are the same. This leads to a more robust test of hypothesis. We prove consistency and show that these measures can be efficiently estimated from data. We demonstrate an application of using the proposed divergence as a cost function to find optimally matched spike trains.
Subject Area: Neuroscience

W60 Brain covariance selection: better individual functional connectivity models using population prior

Gael Varoquaux gael.varoquaux@normalesup.
org
Alexandre Gramfort alexandre.gramfort@inria.fr
INRIA
Jean-Baptiste Poline jbpoline@cea.fr
CEA Bertrand Thirion bertrand.thirion@inria.fr
INRIA

Spontaneous brain activity, as observed in functional neuroimaging, has been shown to display reproducible structure that expresses brain architecture and carries markers of brain pathologies. An important view of modern neuroscience is that such large-scale structure of coherent activity reflects modularity properties of brain connectivity graphs. However, to date, there has been no demonstration that the limited and noisy data available in spontaneous activity observations could be used to learn full-brain probabilistic models that generalize to new data. Learning such models entails two main challenges: i) modeling full brain connectivity is a difficult estimation problem that faces the curse of dimensionality and ii) variability between subjects coupled with the variability of functional signals between experimental runs makes the use of multiple datasets challenging. We describe subject-level brain functional connectivity structure as a multivariate Gaussian process and introduce a new strategy to estimate it from group data by imposing a common structure on the graphical model in the population. We show that individual models learned from functional Magnetic Resonance Imaging (fMRI) data using this population prior generalize better to unseen data than models based on alternative regularization schemes. To our knowledge this is the first report of a cross-validated model of spontaneous brain activity. Finally we use the estimated graphical model to explore the large-scale characteristics of functional architecture and show for the first time that known cognitive networks appear as the integrated communities of functional connectivity graph.
Subject Area: Neuroscience

W61 Epitome driven 3-D Diffusion Tensor image segmentation: on extracting specific structures

Kamiya Motwani kmotwani@cs.wisc.edu
Department of Computer Sciences Univ. of Wisconsin Madison
Nagesh Adluru adluru@wisc.edu
Chris Hinrichs hinrichs@cs.wisc.edu
Vikas Singh vsingh@biostat.wisc.edu
Univ. of Wisconsin Madison
andrew L Alexander alalexander2@wisc.edu
Waisman Laboratory for Brain Imaging and Behavior
Univ. of Wisconsin-Madison

We study the problem of segmenting specific white matter structures of interest from Diffusion Tensor (DT-MR) images of the human brain. This is an important requirement in many Neuroimaging studies: for instance, to evaluate whether a brain structure exhibits group level differences as a function of disease in a set of images. Typically, interactive expert guided segmentation has been the method of choice for such applications, but this is tedious for large datasets common today. To address this problem, we endow an image segmentation algorithm with 'advice' encoding some global characteristics of the region(s) we want to extract. This is accomplished by constructing (using expert-segmented images) an epitome of a specific region - as a histogram over a bag of 'words' (e.g.,suitable feature descriptors). Now, given such a representation, the problem reduces to segmenting new brain image with additional constraints that enforce consistency between the segmented foreground and the pre-specified histogram over features. We present combinatorial approximation algorithms to incorporate such domain specific constraints for Markov Random Field (MRF) segmentation. Making use of recent results on image co-segmentation, we derive effective solution strategies for our problem. We provide an analysis of solution quality, and present promising experimental evidence showing that many structures of interest in Neuroscience can be extracted reliably from 3-D brain image volumes using our algorithm.

Subject Area: Neuroscience
Spotlight presentation, Wednesday, 5:05.

W62 Spatial and anatomical regularization of SVM for brain image analysis

Remi Cuingnet remi.cuingnet@gmail.com
Marie Chupin marie.chupin@upmc.fr
Habib Benali habib.benali@imed.jussieu.fr
INSERM LIF
Olivier Colliot olivier.colliot@upmc.fr
CRICM / CNRS / Universite Pierre et Marie Curie / Paris France

Support vector machines (SVM) are increasingly used in brain image analyses since they allow capturing complex multivariate relationships in the data. Moreover, when the kernel is linear, SVMs can be used to localize spatial patterns of discrimination between two groups of subjects. However, the features' spatial distribution is not taken into account. As a consequence, the optimal margin hyperplane is often scattered and lacks spatial coherence, making its anatomical interpretation difficult. This paper introduces a framework to spatially regularize SVM for brain image analysis. We show that Laplacian regularization provides a flexible framework to integrate various types of constraints and can be applied to both cortical surfaces and 3D brain images. The proposed framework is applied to the classification of MR images based on gray matter concentration maps and cortical thickness measures from 30 patients with Alzheimer's disease and 30 elderly controls. The results demonstrate that the proposed method enables natural spatial and anatomical regularization of the classifier.

Subject Area: Neuroscience

W63 Infinite Relational Modeling of Functional Connectivity in Resting State fMRI

Morten Mørup mm@imm.dtu.dk
DTU Informatics
Lars K Hansen lkh@imm.dtu.dk
Kristoffer H Madsen khm@imm.dtu.dk
Anne-Marie Dogonowski annemd@drcmr.dk
Hartwig R Siebner h.siebner@drcmr.dk
Danish Research Centre for Magnetic Resonance

DTU Informatics Functional magnetic resonance imaging (fMRI) can be applied to study the functional connectivity of the neural elements which form complex network at a whole brain level. Most analyses of functional resting state networks (RSN) have been based on the analysis of correlation between the temporal dynamics of various regions of the brain. While these models can identify coherently behaving groups in terms of correlation they give little insight into how these groups interact. In this paper we take a different view on the analysis of functional resting state networks. Starting from the definition of resting state as functional coherent groups we search for functional units of the brain that communicate with other parts of the brain in a coherent manner as measured by mutual information. We use the infinite relational model (IRM) to quantify functional coherent groups of resting state networks and demonstrate how the extracted component interactions can be used to discriminate between functional resting state activity in multiple sclerosis and normal subjects.

Subject Area: Neuroscience

W64 Functional Geometry Alignment and Localization of Brain Areas

Georg Langs	langs@csail.mit.edu
Polina Golland	polina@csail.mit.edu
MIT	
Yanmei Tie	ytie@bwh.harvard.edu
Laura Rigolo	lrigolo@partners.org
Alexandra Golby	AGOLBY@PARTNERS.ORG
Harvard Medical School	

Matching functional brain regions across individuals is a challenging task, largely due to the variability in their location and extent. It is particularly difficult, but highly relevant, for patients with pathologies such as brain tumors, which can cause substantial reorganization of functional systems. In such cases spatial registration based on anatomical data is only of limited value if the goal is to establish correspondences of functional areas among different individuals, or to localize potentially displaced active regions. Rather than rely on spatial alignment, we propose to perform registration in an alternative space whose geometry is governed by the functional interaction patterns in the brain. We first embed each brain into a functional map that reflects connectivity patterns during a fMRI experiment. The resulting functional maps are then registered, and the obtained correspondences are propagated back to the two brains. In application to a language fMRI experiment, our preliminary results suggest that the proposed method yields improved functional correspondences across subjects. This advantage is pronounced for subjects with tumors that affect the language areas and thus cause spatial reorganization of the functional regions.

Subject Area: Neuroscience

Spotlight presentation, Wednesday, 5:00.

W65 Effects of Synaptic Weight Diffusion on Learning in Decision Making Networks

Kentaro Katahira	katahira@mns.k.u-tokyo.ac.jp
Kazuo Okanoya	okanoya@brain.riken.jp
JST ERATO Okanoya Emotional Information Project	
Masato Okada	okada@k.u-tokyo.ac.jp
The University of Tokyo / RIKEN	

When animals repeatedly choose actions from multiple alternatives, they can allocate their choices stochastically depending on past actions and outcomes. It is commonly assumed that this ability is achieved by modifications in synaptic weights related to decision making. Choice behavior has been empirically found to follow Herrnstein's matching law. Loewenstein & Seung (2006) demonstrated that matching behavior is a steady state of learning in neural networks if the synaptic weights change proportionally to the covariance between reward and neural activities. However, their proof did not take into account the change in entire synaptic distributions. In this study, we show that matching behavior is not necessarily a steady state of the covariance-based learning rule when the synaptic strength is sufficiently strong so that the fluctuations in input from individual sensory neurons influence the net input to output neurons. This is caused by the increasing variance in the input potential due to the diffusion of synaptic weights. This effect causes an undermatching phenomenon, which has been observed in many behavioral experiments. We suggest that the synaptic diffusion effects provide a robust neural mechanism for stochastic choice behavior.

Subject Area: Neuroscience

W66 Deciphering subsampled data: adaptive compressive sampling as a principle of brain communication

Guy Isely	guyi@berkeley.edu
Christopher J Hillar	chillar@msri.org
Fritz Sommer	fsommer@berkeley.edu
UC Berkeley	

A new algorithm is proposed for a) unsupervised learning of sparse representations from subsampled measurements and b) estimating the parameters required for linearly reconstructing signals from the sparse codes. We verify that the new algorithm performs efficient data compression on par with the recent method of compressive sampling. Further, we demonstrate that the algorithm performs robustly when stacked in several stages or when applied in undercomplete or overcomplete situations. The new algorithm can explain how neural populations in the brain that receive subsampled input through fiber bottlenecks are able to form coherent response properties.

Subject Area: Neuroscience

Spotlight presentation, Wednesday, 5:45.

W67 Accounting for network effects in neuronal responses using L1 regularized point process models

Ryan C Kelly	rkelly@cs.cmu.edu
Tai Sing Lee	tai@cnbc.cmu.edu
Robert E Kass	kass@stat.cmu.edu
Carnegie Mellon University	
Matthew A Smith	masmith@cnbc.cmu.edu
University of Pittsburgh	

Activity of a neuron, even in the early sensory areas, is not simply a function of its local receptive field or tuning properties, but depends on global context of the stimulus, as well as the neural context. This suggests the activity of the surrounding neurons and global brain states can exert considerable influence on the activity of a neuron. In this paper we implemented an L1 regularized point process model to assess the contribution of multiple factors to the firing rate of many individual units recorded simultaneously from V1 with a 96-electrode ”Utah” array. We found that the spikes of surrounding neurons indeed provide strong predictions of a neuron's response, in addition to the neuron's receptive field transfer function. We also found that the same spikes could be accounted for with the local field potentials, a surrogate measure of global network states. This work shows that accounting for network fluctuations can improve estimates of single trial firing rate and stimulus-response transfer functions.

Subject Area: Neuroscience

W68 Humans Learn Using Manifolds, Reluctantly

Bryan R Gibson, Univ. of Wisonsin-Madison, Xiaojin (Jerry) Zhu, U. Wisconsin-Madison, Tim Rogers, UW-Madison, Chuck Kalish, , and Joseph Harrison, .

Subject Area: Cognitive Science

W69 Linear readout from a neural population with partial correlation data

Adrien Wohrer adrien.wohrer@ens.fr
Christian K Machens christian.machens@ens.fr
Ecole Normale Superieure
Ranulfo Romo rromo@ifc.unam.mx
Universidad Autonoma de Mexico

How much information does a neural population convey about a stimulus? Answers to this question are known to strongly depend on the correlation of response variability in neural populations. These noise correlations however are essentially immeasurable as the number of parameters in a noise correlation matrix grows quadratically with population size. Here we suggest to bypass this problem by imposing a parametric model on a noise correlation matrix. Our basic assumption is that noise correlations arise due to common inputs between neurons. On average noise correlations will therefore reflect signal correlations which can be measured in neural populations. We suggest an explicit parametric dependency between signal and noise correlations. We show how this dependency can be used to ”fill the gaps” in noise correlations matrices using an iterative application of the Wishart distribution over positive definitive matrices. We apply our method to data from the primary somatosensory cortex of monkeys performing a two-alternative-forced choice task. We compare the discrimination thresholds read out from the population of recorded neurons with the discrimination threshold of the monkey and show that our method predicts different results than simpler average schemes of noise correlations.

Subject Area: Neuroscience

W70 Implicit encoding of prior probabilities in optimal neural populations

Deep Ganguli dganguli@cns.nyu.edu
Eero P Simoncelli eero.simoncelli@cns.nyu.edu
New York University

Optimal coding provides a guiding principle for understanding the representation of sensory variables in neural populations. Here we consider the influence of a prior probability distribution over sensory variables on the optimal allocation of cells and spikes in a neural population. We model the spikes of each cell as samples from an independent Poisson process with rate governed by an associated tuning curve. For this response model, we approximate the Fisher information in terms of the density and amplitude of the tuning curves, under the assumption that tuning width varies inversely with cell density. We consider a family of objective functions based on the expected value, over the sensory prior, of a functional of the Fisher information. This family includes lower bounds on mutual information and perceptual discriminability as special cases. In all cases, we find a closed form expression for the optimum, in which the density and gain of the cells in the population are power law functions of the stimulus prior. This also implies a power law relationship between the prior and perceptual discriminability. We show preliminary evidence that the theory successfully predicts the relationship between empirically measured stimulus priors, physiologically measured neural response properties (cell density, tuning widths, and firing rates), and psychophysically measured discrimination thresholds.

Subject Area: Neuroscience

W71 Fractionally Predictive Spiking Neurons

Sander Bohte sbohte@cwi.nl
Jaldert Rombouts jaldert@gmail.com
CWI

Recent experimental work has suggested that the neural firing rate can be interpreted as a fractional derivative, at least when signal variation induces neural adaptation. Here, we show that the actual neural spike-train itself can be considered as the fractional derivative, provided that the neural signal is approximated by a sum of power-law kernels. A simple standard thresholding spiking neuron suffices to carry out such an approximation, given a suitable refractory response. Empirically, we find that the online approximation of signals with a sum of power-law kernels is beneficial for encoding signals with slowly varying components, like long-memory self-similar signals. For such signals, the online power-law kernel approximation typically required less than half the number of spikes for similar SNR as compared to sums of similar but exponentially decaying kernels. As power-law kernels can be accurately approximated using sums or cascades of weighted exponentials, we demonstrate that the corresponding decoding of spike-trains by a receiving neuron allows for natural and transparent temporal signal filtering by tuning the weights of the decoding kernel.

Subject Area: Neuroscience

W72 Learning Efficient Markov Networks

Vibhav G Gogate vgogate@cs.washington.edu
William Webb williamaustinwebb@gmail.com
Pedro Domingos pedrod@cs.washington.edu
University of Washington Seattle

We present an algorithm for learning high-treewidth Markov networks where inference is still tractable. This is made possible by exploiting context specific independence and determinism in the domain. The class of models our algorithm can learn has the same desirable properties as thin junction trees: polynomial inference, closed form weight learning, etc., but is much broader. Our algorithm searches for a feature that divides the state space into subspaces where the remaining variables decompose into independent subsets (conditioned on the feature or its negation) and recurses on each subspace/subset of variables until no useful new features can be found. We provide probabilistic performance guarantees for our algorithm under the assumption that the maximum feature length is k (the treewidth can be much larger) and dependences are of bounded strength. We also propose a greedy version of the algorithm that, while forgoing these guarantees, is much more efficient.Experiments on a variety of domains show that our approach compares favorably with thin junction trees and other Markov network structure learners.

Subject Area: Probabilistic Models and Methods

W73 Variational bounds for mixed-data factor analysis

Mohammad Emtiyaz Khan emtiyaz@cs.ubc.ca
Benjamin Marlin bmarlin@cs.ubc.ca
Kevin Murphy murphyk@cs.ubc.ca
University of British Columbia
Guillaume Bouchard guillaume.bouchard@xerox.com
XRCE

We propose a new variational EM algorithm for fitting factor analysis models with mixed continuous and categorical observations. The algorithm is based on a simple quadratic bound to the log-sum-exp function. In the special case of fully observed binary data the bound we propose is significantly faster than previous variational methods. We show that EM is significantly more robust in the presence of missing data compared to treating the latent factors as parameters which is the approach used by exponential family PCA and other related matrix-factorization methods. A further benefit of the variational approach is that it can easily be extended to the case of mixtures of factor analyzers as we show. We present results on synthetic and real data sets demonstrating several desirable properties of our proposed method.

Subject Area: Probabilistic Models and Methods

W74 Probabilistic latent variable models for distinguishing between cause and effect

Joris Mooij joris.mooij@tuebingen.mpg.de
Oliver Stegle oliver.stegle@tuebingen.mpg.de
Dominik Janzing dominik.janzing@tuebingen.mpg.de
Kun Zhang kun.zhang@tuebingen.mpg.de
Bernhard Schoelkopf bernhard.schoelkopf@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics

We propose a novel method for inferring whether X causes Y or vice versa from joint observations of X and Y. The basic idea is to model the observed data using probabilistic latent variable models, which incorporate the effects of unobserved noise. To this end, we consider the hypothetical effect variable to be a function of the hypothetical cause variable and an independent noise term (not necessarily additive). An important novel aspect of our work is that we do not restrict the model class, but instead put general non-parametric priors on this function and on the distribution of the cause. The causal direction can then be inferred by using standard Bayesian model selection. We evaluate our approach on synthetic data and real-world data and report encouraging results.

Subject Area: Probabilistic Models and Methods

W75 Graph-Valued Regression

Han Liu hanliu@cs.cmu.edu
Xi Chen xichen@cs.cmu.edu
John Lafferty lafferty@cs.cmu.edu
Larry Wasserman larry@stat.cmu.edu
Carnegie Mellon University

Undirected graphical models encode in a graph G the dependency structure of a random vector Y . In many applications, it is of interest to model Y given another random vector X as input. We refer to the problem of estimating the graph G(x) of Y conditioned on X = x as “graph-valued regression”. In this paper, we propose a semiparametric method for estimating G(x) that builds a tree on the X space just as in CART (classification and regression trees), but at each leaf of the tree estimates a graph. We call the method “Graph-optimized CART”, or Go-CART. We study the theoretical properties of Go-CART using dyadic partitioning trees, establishing oracle inequalities on risk minimization and tree partition consistency. We also demonstrate the application of Go-CART to a meteorological dataset, showing how graph-valued regression can provide a useful tool for analyzing complex data.

Subject Area: Probabilistic Models and Methods
Spotlight presentation, Wednesday, 11:15.

W76 Learning Networks of Stochastic Differential Equations

Jos´e Bento Ayres Pereira jbento@stanford.edu
Morteza Ibrahimi ibrahimi@stanford.edu
Andrea Montanari montanari@stanford.edu
Stanford University

We consider linear models for stochastic dynamics. Any such model can be associated a network (namely a directed graph) describing which degrees of freedom interact under the dynamics. We tackle the problem of learning such a network from observation of the system trajectory over a time interval T.We analyse the l1-regularized least squares algorithm and in the setting in which the underlying network is sparse we prove performance guarantees that are uniform in the sampling rate as long as this is sufficiently high. This result substantiates the notion of a well defined 'time complexity' for the network inference problem.
Subject Area: Probabilistic Models and Methods

W77 Copula Bayesian Networks

Gal Elidan galel@huji.ac.il
Hebrew University

We present the Copula Bayesian Network model for representing multivariate continuous distributions. Our approach builds on a novel copula-based parameterization of a conditional density that, joined with a graph that encodes independencies, offers great flexibility in modeling high-dimensional densities, while maintaining control over the form of the univariate marginals. We demonstrate the advantage of our framework for generalization over standard Bayesian networks as well as tree structured copula models for varied real-life domains that are of substantially higher dimension than those typically considered in the copula literature.
Subject Area: Probabilistic Models and Methods

W78 t-logistic regression

Nan Ding ding10@purdue.edu
S.V.N. Vishwanathan vishy@stat.purdue.edu
Purdue University

We extend logistic regression by using t-exponential families which were introduced recently in statistical physics. This gives rise to a regularized risk minimization problem with a non-convex loss function. An efficient block coordinate descent optimization scheme can be derived for estimating the parameters. Because of the nature of the loss function, our algorithm is tolerant to label noise. Furthermore, unlike other algorithms which employ non-convex loss functions our algorithm is fairly robust to the choice of initial values. We verify both these observations empirically on a number of synthetic and real datasets.
Subject Area: Probabilistic Models and Methods

W79 Shadow Dirichlet for Restricted Probability Modeling

Bela A Frigyik frigyik@gmail.com
Maya Gupta gupta@ee.washington.edu
Yihua Chen yhchen@ee.washington.edu
University of Washington

Although the Dirichlet distribution is widely used, the independence structure of its components limits its accuracy as a model. The proposed shadow Dirichlet distribution manipulates the support in order to model probability mass functions (pmfs) with dependencies or constraints that often arise in real world problems such as regularized pmfs monotonic pmfs and pmfs with bounded variation. We describe some properties of this new class of distributions provide maximum entropy constructions give an expectation-maximization method for estimating the mean parameter and illustrate with real data.
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Wednesday, 5:15.

W80 Online Learning for Latent Dirichlet Allocation

Matthew Hoffman mdhoffma@princeton.edu
David Blei blei@cs.princeton.edu
Princeton University
Francis Bach francis.bach@ens.fr
Ecole Normale Superieure

We develop an online variational Bayes (VB) algorithm for Latent Dirichlet Allocation (LDA). Online LDA is based on online stochastic optimization with a natural gradient step, which we show converges to a local optimum of the VB objective function. It can handily analyze massive document collections, including those arriving in a stream. We study the performance of online LDA in several ways, including by fitting a 100-topic topic model to 3.3M articles from Wikipedia in a single pass. We demonstrate that online LDA finds topic models as good or better than those found with batch VB, and in a fraction of the time.
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Wednesday, 12:05.

W81 Approximate inference in continuous time Gaussian-Jump processes

Manfred Opper opperm@cs.tu-berlin.de
Andreas Ruttor andreas.ruttor@tu-berlin.de
Technische Universitaet Berlin
Guido Sanguinetti gsanguin@inf.ed.ac.uk
University of Edinburgh

We present a novel approach to inference in conditionally Gaussian continuous time stochastic processes where the latent process is a Markovian jump process. We first consider the case of jump-diffusion processes where the drift of a linear stochastic differential equation can jump at arbitrary time points. We derive partial differential equations for exact inference and present a very efficient mean field approximation. By introducing a novel lower bound on the free energy we then generalise our approach to Gaussian processes with arbitrary covariance such as the non-Markovian RBF covariance. We present results on both simulated and real data showing that the approach is very accurate in capturing latent dynamics and can be useful in a number of real data modelling tasks.
Subject Area: Probabilistic Models and Methods

W82 Global Analytic Solution for Variational Bayesian Matrix Factorization

Shinichi Nakajima nakajima.s@nikon.co.jp
Nikon Corporation
Masashi Sugiyama sugi@cs.titech.ac.jp
Tokyo Institute of Technology
Ryota Tomioka tomioka@mist.i.u-tokyo.ac.jp
University of Tokyo

Bayesian methods of matrix factorization (MF) have been actively explored recently as promising alternatives to classical singular value decomposition. In this paper we show that despite the fact that the optimization problem is non-convex the global optimal solution of variational Bayesian (VB) MF can be computed analytically by solving a quartic equation. This is highly advantageous over a popular VBMF algorithm based on iterated conditional modes since it can only find a local optimal solution after iterations. We further show that the global optimal solution of empirical VBMF (hyperparameters are also learned from data) can also be analytically computed. We illustrate the usefulness of our results through experiments.
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Wednesday, 9:55.

W83 Tree-Structured Stick Breaking for Hierarchical Data

Ryan Adams, University of Toronto, Zoubin Ghahramani, Cambridge, and Michael I Jordan, UC Berkeley.
Subject Area: Probabilistic Models and Methods
See abstract, page 78

W84 Construction of Dependent Dirichlet Processes based on Poisson Processes

Dahua Lin, Eric Grimson and John Fisher, MIT.
Subject Area: Probabilistic Models and Methods
See abstract, page 77

W85 Switched Latent Force Models for Movement Segmentation

Mauricio Alvarez alvarezm@cs.man.ac.uk
Neil D Lawrence neill@cs.man.ac.uk
University of Manchester
Jan Peters jan.peters@tuebingen.mpg.de
Bernhard Schoelkopf bernhard.schoelkopf@tuebingen.mpg.de
MPI for Biological Cybernetics

Latent force models encode the interaction between multiple related dynamical systems in the form of a kernel or covariance function. Each variable to be modeled is represented as the output of a differential equation and each differential equation is driven by a weighted sum of latent functions with uncertainty given by a Gaussian process prior. In this paper we consider employing the latent force model framework for the problem of determining robot motor primitives. To deal with discontinuities in the dynamical systems or the latent driving force we introduce an extension of the basic latent force model that switches between different latent functions and potentially different dynamical systems. This creates a versatile representation for robot movements that can capture discrete changes and non-linearities in the dynamics. We give illustrative examples on both synthetic data and for striking movements recorded using a Barrett WAM robot as haptic input device. Our inspiration is robot motor primitives but we expect our model to have wide application for dynamical systems including models for human motion capture data and systems biology.
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Wednesday, 11:20.

W86 Slice sampling covariance hyperparameters of latent Gaussian models

Iain Murray, University of Edinburgh
Ryan Adams, University of Toronto.
Subject Area: Probabilistic Models and Methods
See abstract, page 77

W87 Exact inference and learning for cumulative distribution functions on loopy graphs

Jim C Huang jimhua@microsoft.com
Nebojsa Jojic jojic@microsoft.com
Chris Meek meek@microsoft.com
Microsoft Research

Probabilistic graphical models use local factors to represent dependence among sets of variables. For many problem domains, for instance climatology and epidemiology, in addition to local dependencies, we may also wish to model heavy-tailed statistics, where extreme deviations should not be treated as outliers. Specifying such distributions using graphical models for probability density functions (PDFs) generally lead to intractable inference and learning. Cumulative distribution networks (CDNs) provide a means to tractably specify multivariate heavy-tailed models as a product of cumulative distribution functions (CDFs). Currently, algorithms for inference and learning, which correspond to computing mixed derivatives, are exact only for tree-structured graphs. For graphs of arbitrary topology, an efficient algorithm is needed that takes advantage of the sparse structure of the model, unlike symbolic differentiation programs such as Mathematica and D* that do not. We present an algorithm for recursively decomposing the computation of derivatives for CDNs of arbitrary topology, where the decomposition is naturally described using junction trees. We compare the performance of the resulting algorithm to Mathematica and D*, and we apply our method to learning models for rainfall and H1N1 data, where we show that CDNs with cycles are able to provide a significantly better fits to the data as compared to tree-structured and unstructured CDNs and other heavy-tailed multivariate distributions such as the multivariate copula and logistic models.
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Wednesday, 9:50.

W88 Evidence-Specific Structures for Rich Tractable CRFs

Anton Chechetka antonc@cs.cmu.edu
Carlos Guestrin guestrin@cs.cmu.edu
Carnegie Mellon University

We present a simple and effective approach to learning tractable conditional random fields with structure that depends on the evidence. Our approach retains the advantages of tractable discriminative models, namely efficient exact inference and exact parameter learning. At the same time, our algorithm does not suffer a large expressive power penalty inherent to fixed tractable structures. On real-life relational datasets, our approach matches or exceeds state of the art accuracy of the dense models, and at the same time provides an order of magnitude speedup
Subject Area: Probabilistic Models and Methods

W89 On the Convexity of Latent Social Network Inference

Seth Myers and Jure Leskovec, Stanford University.
Subject Area: Probabilistic Models and Methods
See abstract, page 79

W90 Structured Determinantal Point Processes

Alex Kulesza kulesza@cis.upenn.edu
Ben Taskar taskar@cis.upenn.edu
University of Pennsylvania

We present a novel probabilistic model for distributions over sets of structures – for example, sets of sequences, trees, or graphs. The critical characteristic of our model is a preference for diversity: sets containing dissimilar structures are more likely. Our model is a marriage of structured probabilistic models, like Markov random fields and context free grammars, with determinantal point processes, which arise in quantum physics as models of particles with repulsive interactions. We extend the determinantal point process model to handle an exponentially-sized set of particles (structures) via a natural factorization of the model into parts. We show how this factorization leads to tractable algorithms for exact inference, including computing marginals, computing conditional probabilities, and sampling. Our algorithms exploit a novel polynomially-sized dual representation of determinantal point processes, and use message passing over a special semiring to compute relevant quantities. We illustrate the advantages of the model on tracking and articulated pose estimation problems.
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Wednesday, 3:10.

W91 Throttling Poisson Processes

Uwe Dick uwedick@cs.uni-potsdam.de
Peter Haider haider@cs.uni-potsdam.de
Thomas Vanck vanck@cs.uni-potsdam.de
Michael Bruckner mibrueck@cs.uni-potsdam.de
Tobias Scheffer scheffer@cs.uni-potsdam.de
University of Potsdam

We study a setting in which Poisson processes generate sequences of decision-making events. The optimization goal is allowed to depend on the rate of decision outcomes; the rate may depend on a potentially long backlog of events and decisions. We model the problem as a Poisson process with a throttling policy that enforces a data-dependent rate limit and reduce the learning problem to a convex optimization problem that can be solved efficiently. This problem setting matches applications in which damage caused by an attacker grows as a function of the rate of unsuppressed hostile events. We report on experiments on abuse detection for an email service.
Subject Area: Probabilistic Models and Methods
Spotlight presentation, Wednesday, 11:10.

W92 Dynamic Infinite Relational Model for Time-varying Relational Data Analysis

Katsuhiko Ishiguro ishiguro@cslab.kecl.ntt.co.jp
Tomoharu Iwata iwata@cslab.kecl.ntt.co.jp
Naonori Ueda ueda@cslab.kecl.ntt.co.jp
NTT Communication Science Laboratories
Joshua Tenenbaum jbt@mit.edu
Massachusetts Institute of Technology

We propose a new probabilistic model for analyzing dynamic evolutions of relational data, such as additions, deletions and split & merge, of relation clusters like communities in social networks. Our proposed model abstracts observed time-varying object-object relationships into relationships between object clusters. We extend the infinite Hidden Markov model to follow dynamic and time-sensitive changes in the structure of the relational data and to estimate a number of clusters simultaneously. We show the usefulness of the model through experiments with synthetic and real-world data sets.
Subject Area: Probabilistic Models and Methods

W93 Universal Consistency of Multi-Class Support Vector Classification

Tobias Glasmachers tobias@idsia.ch

IDSIA Steinwart was the first to prove universal consistency of support vector machine classification. His proof analyzed the 'standard' support vector machine classifier which is restricted to binary classification problems. In contrast recent analysis has resulted in the common belief that several extensions of SVM classification to more than two classes are inconsistent. Countering this belief we proof the universal consistency of the multi-class support vector machine by Crammer and Singer. Our proof extends Steinwart's techniques to the multi-class case.
Subject Area: Theory

W94 Random Walk Approach to Regret Minimization

Hariharan Narayanan har@mit.edu
MIT
Alexander Rakhlin rakhlin@gmail.com
University of Pennsylvania

We propose a computationally efficient random walk on a convex body which rapidly mixes to a time-varying Gibbs distribution. In the setting of online convex optimization and repeated games, the algorithm yields low regret and presents a novel efficient method for implementing mixture forecasting strategies.
Subject Area: Theory

W95 Online Markov Decision Processes under Bandit Feedback

Gergely Neu neu.gergely@gmail.com
Budapest U. of Tech. and Econ.
Andras Gyorgy gya@szit.bme.hu
Andras Antos antos@szit.bme.hu
MTA SZTAKI Institute for Computer Science and Control
Csaba Szepesvari szepesva@ualberta.ca
University of Alberta

We consider online learning in finite stochastic Markovian environments where in each time step a new reward function is chosen by an oblivious adversary. The goal of the learning agent is to compete with the best stationary policy in terms of the total reward received. In each time step the agent observes the current state and the reward associated with the last transition, however, the agent does not observe the rewards associated with other state-action pairs. The agent is assumed to know the transition probabilities. The state of the art result for this setting is a no-regret algorithm. In this paper we propose a new learning algorithm and assuming that stationary policies mix uniformly fast, we show that after T time steps, the expected regret of the new algorithm is $O(T^{2/3}(\mathit{lnT})^{1/3})$, giving the first rigorously proved convergence rate result for the problem.
Subject Area: Theory
Spotlight presentation, Wednesday, 11:25.

W96 New Adaptive Algorithms for Online Classification

Francesco Orabona orabona@dsi.unimi.it
University of Milano
Koby Crammer koby@ee.technion.ac.il
Technion

We propose a general framework to online learning for classification problems with time-varying potential functions in the adversarial setting. This framework allows to design and prove relative mistake bounds for any generic loss function. The mistake bounds can be specialized for the hinge loss allowing to recover and improve the bounds of known online classification algorithms. By optimizing the general bound we derive a new online classification algorithm called NAROW that hybridly uses adaptive- and fixed- second order information. We analyze the properties of the algorithm and illustrate its performance using synthetic dataset. Subject Area: Theory

W97 Online Classification with Specificity Constraints

Andrey Bernstein andreyb@tx.technion.ac.il
Shie Mannor shie@ee.technion.ac.il
Nahum Shimkin shimkin@ee.technion.ac.il
Technion

We consider the online binary classification problem, where we are given m classifiers. At each stage, the classifiers map the input to the probability that the input belongs to the positive class. An online classification meta-algorithm is an algorithm that combines the outputs of the classifiers in order to attain a certain goal, without having prior knowledge on the form and statistics of the input, and without prior knowledge on the performance of the given classifiers. In this paper, we use sensitivity and specificity as the performance metrics of the meta-algorithm. In particular, our goal is to design an algorithm which satisfies the following two properties (asymptotically): (i) its average false positive rate (fp-rate) is under some given threshold, and (ii) its average true positive rate (tp-rate) is not worse than the tp-rate of the best convex combination of the m given classifiers that satisfies fp-rate constraint, in hindsight. We show that this problem is in fact a special case of the regret minimization problem with constraints, and therefore the above goal is not attainable. Hence, we pose a relaxed goal and propose a corresponding practical online learning meta-algorithm that attains it. In the case of two classifiers we show that this algorithm takes a very simple form. To our best knowledge this is the first algorithm that addresses the problem of the average tp-rate maximization under average fp-rate constraints in the online setting.
Subject Area: Theory
Spotlight presentation, Wednesday, 9:45.

W98 The LASSO risk: asymptotic results and real world examples

Mohsen Bayati bayati@stanfordalumni.org
Jos´e Bento Ayres Pereira jbento@stanford.edu
Andrea Montanari montanar@stanford.edu
Stanford University

We consider the problem of learning a coefficient vector x0 from noisy linear observation y=Ax0+w. In many contexts (ranging from model selection to image processing) it is desirable to construct a sparse estimator. In this case, a popular approach consists in solving an l1-penalized least squares problem known as the LASSO or BPDN. For sequences of matrices A of increasing dimensions with iid gaussian entries we prove that the normalized risk of the LASSO converges to a limit and we obtain an explicit expression for this limit. Our result is the first rigorous derivation of an explicit formula for the asymptotic risk of the LASSO for random instances. The proof technique is based on the analysis of AMP a recently developed efficient algorithm that is inspired from graphical models ideas. Through simulations on real data matrices (gene expression data and hospital medical records) we observe that these results can be relevant in a broad array of practical applications.
Subject Area: Theory

W99 Tight Sample Complexity of Large-Margin Learning

Sivan Sabato sivan_sabato@cs.huji.ac.il
Naftali Tishby tishby@cs.huji.ac.il
The Hebrew University Jerusalem
Nathan Srebro nati@ttic.edu
TTI

We obtain a tight distribution-specific characterization of the sample complexity of largemargin classification with L2 regularization: We introduce the gamma-adapted-dimension, which is a simple function of the spectrum of a distribution's covariance matrix, and show distribution-specific upper and lower bounds on the sample complexity, both governed by the gamma-adapted-dimension of the source distribution. We conclude that this new quantity tightly characterizes the true sample complexity of large-margin classification. The bounds hold for a rich family of sub-Gaussian distributions.
Subject Area: Theory

W100 Universal Kernels on Non-Standard Input Spaces

Andreas Christmann andreas.christmann@uni-bayreuth.de
University of Bayreuth
Ingo Steinwart ingo.steinwart@mathematik.uni-stuttgart.de
University of Stuttgart

During the last years support vector machines (SVMs) have been successfully applied even in situations where the input space X is not necessarily a subset of R^d. Examples include SVMs using probability measures to analyse e.g. histograms or coloured images, SVMs for text classification and web mining, and SVMs for applications from computational biology using, e.g., kernels for trees and graphs. Moreover, SVMs are known to be consistent to the Bayes risk, if either the input space is a complete separable metric space and the reproducing kernel Hilbert space (RKHS) $H \subset L_p(P_X)$ is dense, or if the SVM is based on a universal kernel k. So far however there are no RKHSs of practical interest known that satisfy these assumptions on H or k if $X\ 6 \not\subset R^d$. We close this gap by providing a general technique based on Taylor-type kernels to explicitly construct universal kernels on compact metric spaces which are not subset of R^d. We apply this technique for the following special cases: universal kernels on the set of probability measures universal kernels based on Fourier transforms and universal kernels for signal processing.
Subject Area: Theory

D8 BCI Demonstration using a Dry-Electrode

Achim Hornecker sr@brainproducts.com
Brain Products GmbH

The subject will control a game (levitate a ball within a tube) only by the means of his/her thoughts (no action by any extremities required) and electrodes/electrode cap that doesn't need to be prepared using gel will be used for the acquisition of the EEG signals.

D9 Globby: It's a Search Engine with a Sorting View

Novi Quadrianto novi.quad@gmail.com
NICTA

Bridging the gap between keyword-based and content-based search engines is the next big thing in enhancing the user experience and serving up more systematic search results. Globby achieves this by returning keyword-based query relevant images in a set of pages, where each page contains several images with content alike images are placed at proximal locations. Globby is based on a provably locally convergent statistical method to arrange similar images close to each other in predefined but arbitrary structures with an additional ranking constraint that the most relevant image to the query is placed at, for example, top left corner.

D10 Platform to Share Feature Extraction Methods

Francois Fleuret francois.fleuret@idiap.ch
IDIAP Research Institute

The MASH project is a three year European initiative which started early in 2010. Its main objective is to develop new tools and methods to help the collaborative design of very large families of features extractors for machine learning. This is investigated through the development of an open web platform which allows to submit implementations of feature extractors, browse extractors already contributed to the system, launch experiments, and analyze previous experimental results to focus the design on the identified weakness of the system. Performance of the designed learning architectures are evaluated on image recognition, object detection, and goal planning. The former targeting both a video-gamelike 3d simulator and a real robot controlled remotely. We want mainly to demonstrate the features of the platform to the NIPS attendees, and show both the development tools, and the applications we have already implemented.

D11 Stochastic Matlab

David Wingate wingated@mit.edu
Massachusetts Institute of Technology

We'll be demonstrating Stochastic Matlab, a new probabilistic programming language. Stochastic Matlab allows users to write generative models in Matlab (freely mixing deterministic and random primitives, MEX files, and any other Matlab construct) and then condition the model on data. Stochastic Matlab then performs inference automatically using a variety of inference methods, including standard MCMC, parallel tempering, and hamiltonian Monte Carlo. Lightweight GPU integration is also provided. Probabilistic programming languages allow users to rapidly iterate models, testing them under a variety of inference methods. The goal of Stochastic Matlab is to scale to large datasets by taking advantage of GPUs and clusters. (Note: this project is an open-source project, and is not affiliated with Mathworks.)

D12 The SHOGUN Machine Learning Toolbox

Soeren Sonnenburg Soeren.Sonnenburg@tu-berlin.de
Berlin Institute of Technology

We have developed a machine learning toolbox, called SHOGUN, which is designed for unified large-scale learning for a broad range of feature types and learning settings. It offers a considerable number of machine learning models such as support vector machines for classification and regression, hidden Markov models, multiple kernel learning, linear discriminant analysis, linear programming machines, and perceptrons. Most of the specific algorithms are able to deal with several different data classes, including dense and sparse vectors and sequences using floating point or discrete data types. We have used this toolbox in several applications from computational biology, some of them coming with no less than 50 million training examples and others with 7 billion test examples. With more than a thousand installations worldwide, SHOGUN is already widely adopted in the machine learning community and beyond. SHOGUN is implemented in C++ and interfaces to MATLAB, R, Octave, Python, and has a stand-alone command line interface. The source code is freely available under the GNU General Public License, Version 3 at http://www.shogun-toolbox.org.

D13 Visual Object Recognition with NN Convolutional & Spiking; Test on Traffic Signs

Vincent de Ladurantaye vincent.de.ladurantaye@usherbrooke.ca
Universite de Sherbrooke

We propose a new method of realizing visual object recognition by combining hierarchical feature extraction and temporal binding. In a first time, our system uses a convolutive hierarchical neural network similar to HMAX (Riesenhuber, 1999) to extract features from the input image. Feature extractors are learned using unsupervised learning methods (Kavukcuoglu, 2009). However, instead of using conventional classification methods like SVM, we use a spiking neural network to realize temporal binding association as proposed by (Milner, 1974) and (Malsburg, 1981). Features of the input image are matched to a reference image using a modified version of the model proposed by (Pichevar, 2006). Similar object parts will synchronize to realize recognition. The advantage of using such architecture, as opposed to classical classification methods, is that the local organization of the features is conserved, yielding a more robust recognition. The system is also able to recognize objects from very few training samples. Another advantage of our model is that the matching process with a reference image can help interpolate missing or occluded part of the objects.

THURSDAY CONFERENCE

ORAL SESSION 17 - (9:00–10:10AM):
Session Chair: Rob Fergus

INVITED TALK: Perceptual Bases for Rules of Thumb in Photography

Martin S Banks martybanks@berkeley.edu
University of California, Berkeley

Photographers utilize many rules of thumb for creating natural-looking pictures. The explanations for these guidelines are vague and probably incorrect. I will explore two common photographic rules and argue that they are understandable from a consideration of the perceptual mechanisms involved and peoples‘; viewing habits. The first rule of thumb concerns the lens focal length required to produce pictures that are not spatially distorted. Photography textbooks recommend choosing a focal length that is 3/2 the film width. The text books state vaguely that the rule creates ”a field of view that corresponds to that of normal vision” (Giancoli, 2000), ”the same perspective as the human eye” (Alesse, 1989), or ”;approximates the impression human vision gives“; (London et al., 2005). There are two phenomena related to this rule. One is perceived spatial distortions in wide-angle (short focal length) pictures. I argue that the perceived distortions are caused by the perceptual mechanisms people employ to take into account oblique viewing positions. I present some demonstrations that validate this explanation. The second phenomenon is perceived depth in pictures taken with different focal lengths. The textbooks argue that pictures taken with short focal lengths expand perceived depth and those taken with long focal lengths compress it. I argue that these effects are due to a combination of the viewing geometry and the way people typically look at pictures. I present demonstrations to validate this. The second rule of thumb concerns the camera aperture and depth-of-field blur. Photography textbooks do not describe a quantitative rule and treat the magnitude of depth-of-field blur as arbitrary. I examine the geometry of apertures, lenses, and image formation. From that analysis, I argue that there is a natural relationship between depthof- field blur and the 3D layout of the photographed scene. I present demonstrations that human viewers are sensitive to this relationship. In particular, depicted scenes are perceived differently depending on the relationship between blur and 3D layout.

Martin Banks is Professor of Optometry, Vision Science, Psychology, and Neuroscience at UC Berkeley. He received his Bachelor‘;s degree in 1970 from Occidental College, majoring in psychology and minoring in physics. He then spent a year in Germany as a grade-school teacher. He then entered graduate school at UC San Diego where he received a Master‘;s degree in experimental psychology in 1973. He entered the doctoral program at University of Minnesota receiving a PhD. in developmental psychology in 1976. He was Assistant and Associate Professor of Psychology at University of Texas Austin before moving to Berkeley in 1985. Professor Banks is known for research on human visual perception, particularly the perception of depth and for research on the integration of cues from different sensory organs. He was involved in the development of novel stereo displays that present nearly correct focus cues and other stereo displays that bypass the optics of the human eye. Professor Banks is a Fellow of the American Association for the Advancement of Science, a Fellow of the American Psychological Society, Fellow of the Center for the Advanced Study of the Behavioral Sciences, recipient of the McCandless Award for Early Scientific Contribution, recipient of the Gottsdanker and Howard lectureships, the first recipient of the Koffka Award for Contribution in Perception and Development, and an Honorary Professor of the University of Wales, Cardiff.

ORAL SESSION - 9:50AM

Learning to combine foveal glimpses with a third-order Boltzmann machine

Hugo Larochelle larocheh@cs.toronto.edu
Geoffrey Hinton hinton@cs.toronto.edu
University of Toronto

We describe a model based on a Boltzmann machine with third-order connections that can learn how to accumulate information about a shape over several fixations. The model uses a retina that only has enough high resolution pixels to cover a small area of the image, so it must decide on a sequence of fixations and it must combine the ”glimpse” at each fixation with the location of the fixation before integrating the information with information from other glimpses of the same object. We evaluate this model on a synthetic dataset and two image classification datasets, showing that it can perform at least as well as a model trained on whole images.
Subject Area: Supervised Learning

ORAL SESSION 18 - (10:40–11:50AM):
Session Chair: David Blei

INVITED TALK: How to Grow a Mind: Statistics, Structure and Abstraction , Posner Lecture

Josh Tenenbaum
MIT

How do humans come to know so much about the world from so little data? Even young children can infer the meanings of words, the hidden properties of objects, or the existence of causal relations from just one or a few relevant observations -- far outstripping the capabilities of conventional learning machines. How do they do it -- and how can we bring machines closer to these human-like learning abilities? I will argue that people’s everyday inductive leaps can be understood in terms of (approximations to) probabilistic inference over generative models of the world. These models can have rich latent structure based on abstract knowledge representations, what cognitive psychologists have sometimes called “intuitive theories”, “mental models”, or “schemas”. They also typically have a hierarchical structure supporting inference at multiple levels, or “learning to learn”, where abstract knowledge may itself be learned from experience at the same time as it guides more specific generalizations from sparse data. This talk will focus on models of learning and “learning to learn” about categories, word meanings and causal relations. I will show in each of these settings how human learners can balance the need for strongly constraining inductive biases -- necessary for rapid generalization -- with the flexibility to adapt to the structure of new environments, learning new inductive biases for which our minds could not have been pre-programmed. I will also discuss briefly how this approach extends to richer forms of knowledge, such as intuitive psychology and social inferences, or physical reasoning. Finally, I will raise some challenges for our current state of understanding about learning in the brain, and neurally inspired computational models.

Josh Tenenbaum is an Associate Professor of Computational Cognitive Science at MIT in the Department of Brain and Cognitive Sciences and the Computer Science and Artificial Intelligence Laboratory (CSAIL). He received his PhD from MIT in 1999, and was an Assistant Professor at Stanford University from 1999 to 2002. He studies learning and inference in humans and machines, with the twin goals of understanding human intelligence in computational terms and bringing computers closer to human capacities. He focuses on problems of inductive generalization from limited data -- learning concepts and word meanings, inferring causal relations or goals -- and learning abstract knowledge that supports these inductive leaps in the form of probabilistic generative models or 'intuitive theories'. He has also developed several novel machine learning methods inspired by human learning and perception, most notably Isomap, an approach to unsupervised learning of nonlinear manifolds in high-dimensional data. He has been Associate Editor for the journal Cognitive Science, has been active on program committees for the CogSci and NIPS conferences, and has co-organized a number of workshops, tutorials and summer schools in human and machine learning. Several of his papers have received outstanding paper awards or best student paper awards at the IEEE Computer Vision and Pattern Recognition (CVPR), NIPS, and Cognitive Science conferences. He is the recipient of the New Investigator Award from the Society for Mathematical Psychology (2005), the Early Investigator Award from the Society of Experimental Psychologists (2007), and the Distinguished Scientific Award for Early Career Contribution to Psychology (in the area of cognition and human learning) from the American Psychological Association (2008).

ORAL SESSION - 10:40AM
Session Chair: Maneesh Sahani

Humans Learn Using Manifolds, Reluctantly

Bryan R Gibson	bgibson@cs.wisc.edu
Xiaojin (Jerry) Zhu	jerryzhu@cs.wisc.edu U.
Tim Rogers	ttrogers@wisc.edu
Chuck Kalish	cwkalish@wisc.edu
Joseph Harrison	jharrison@wisc.edu

Univ. of Wisconsin-Madison

When the distribution of unlabeled data in feature space lies along a manifold, the information it provides may be used by a learner to assist classification in a semi-supervised setting. While manifold learning is well-known in machine learning, the use of manifolds in human learning is largely unstudied. We perform a set of experiments which test a human's ability to use a manifold in a semi-supervised learning task, under varying conditions. We show that humans may be encouraged into using the manifold, overcoming the strong preference for a simple, axis-parallel linear boundary.

Subject Area: Cognitive Science

THE SAM ROWEIS SYMPOSIUM

THE SAM ROWEIS SYMPOSIUM

(2:00–5:00PM)

MS1 Unifying Views in Unsupervised Learning

Zoubin Ghahramani zoubin@eng.cam.ac.uk
University of Cambridge & CMU

The NIPS community has benefited greatly from Sam Roweis' insights into the connections between different models and algorithms. I will review our work on a 'unifying' framework for linear Gaussian models, which formed the backbone of the NIPS Tutorial Sam and I gave in 1999. This framework highlighted connections between factor analysis, PCA, mixture models, HMMs, state-space models, and ICA, had the EM algorithm as the allpurpose swiss-army-knife of learning algorithms, and culminated in a 'graphical model for graphical models' depicting the connections. Though perhaps well-known now, those connections were surprising at the time (at least to us) and resulted in a more coherent and systematic view of statistical machine learning that has endured to this day. Inspired by this approach, I will present some newer unifying views, of kernel methods, and of nonparametric Bayesian models.

MS2 Manifold Learning

Lawrence Saul saul@cs.ucsd.edu
UC San Diego

How can we detect low dimensional structure in high dimensional data? Sam and I worked feverishly on this problem for a number of years. We were particularly interested in analyzing high dimensional data that lies on or near a low dimensional manifold. I will describe the algorithm, locally linear embedding (LLE), that we developed for this problem. I will conclude by relating LLE to more recent work in manifold learning and sketching some future directions for research.

MS3 A Probabilistic Approach to Data Visualization

Geoffrey E Hinton hinton@cs.toronto.edu
University of Toronto

Dimensionality reduction methods allow us to visualize the structure of large, high-dimensional datasets by giving each data-point a location in a two-dimensional map. Sam Roweis was involved in the development of several different methods for producing maps that preserve local similarity by displaying very similar data-points at nearby locations in the map without worrying too much about the map distances between dissimilar data-points. One of these methods, called Stochastic Neighbor Embedding, converts the problem of finding a good map into the problem of matching two probability distributions. It uses the density under a high-dimensional Gaussian centered at each data-point to determine the probability of picking each of the other data-points as a neighbor. It then uses exactly the same method to determine neighbor probabilities using the two-dimensional locations of the corresponding map points. The aim is to move the map points so that the neighbor probabilities computed in the high-dimensional data-space are well-modeled by the neighbor probabilities computed in the low-dimensional map. This leads to very nice maps for a variety of datasets. I will describe some further developments of this method that lead to even better maps.

MS4 Learning Structural Sparsity

Yoram Singer singer@google.com
Google

In the past years my work focused on algorithms for learning high dimensional yet sparse models from very large datasets. During the two years that Sam spent at Google, he greatly influenced my course of research on large scale learning of structural sparsity. He was too humble and too busy to formally co-author any of the papers that constitute the talk (see http://magicbroom.info/Sparsity.html). Yet, many parts of this talk would not have materialized without his encouragement, feedback, and ideas. In the talk I review the design, analysis and implementation of sparsity promoting learning algorithms, including coordinate and mirror descent with non-smooth regularization, forward-backward splitting algorithms, and other recently devised algorithms for sparse models. I will conclude with an overview of new work on learning self pruning decision trees and structured histograms by combining exponential models with sparsity promoting regularization.

MS5 Automating Astronomy

David W. Hogg david.hogg@nyu.edu
New York University

Telescopes and their instruments produce digital images. This makes astronomy a playground for computer vision. It is ”easy” because the viewpoint is fixed, the illumination is fixed (on human timescales), and the range of objects viewed is limited (at finite resolution). It is hard because the data are beyond multi-spectral (current data span 17 orders of magnitude in wavelength) and we care deeply about the objects imaged at extremely low signal-to-noise. In precise contexts we are compelled to model the data probabilistically; this requires techniques of machine learning. Successes in this area have implications for our understanding of the fundamental laws of physics, the dark matter, and the initial conditions of the Universe.

IN MEMORIAM

Sam Roweis 1972-2010

The Neural Information Processing Systems Foundation mourns the untimely death of Sam T. Roweis on January 12, 2010. Sam was a brilliant scientist and engineer whose work deeply influenced the fields of artificial intelligence, machine learning, applied mathematics, neural computation, and observational science. He was also a strong advocate for the use of machine learning and computational statistics for scientific data analysis and discovery.

Sam T. Roweis was born on April 27, 1972. He graduated from secondary school as valedictorian of the University of Toronto Schools in 1990, and obtained a bachelor's degree with honors from the University of Toronto Engineering Science Program four years later. In 1994 he joined the Computation and Neural Systems PhD program at the California Institute of Technology, working under the supervision of John J. Hopfield. After earning his PhD in 1999, Sam took a postdoctoral position in London with the Gatsby Computational Neuroscience Unit. After his postdoc, some time at MIT, and a stint with the startup company WhizBang! Labs, Sam took a faculty job at the University of Toronto. In 2005 Sam spent a semester at MIT and in 2006 he was named a fellow of the Canadian Institute for Advanced Research (CIfAR) and received tenure at Toronto. He joined Google's research labs in San Francisco and Mountain View in 2007. Sam moved to the Computer Science Department at NYU's Courant Institute as an Associate Professor in September 2009.

Among Sam's many achievements in machine learning was a new form of Independent Component Analysis (ICA) that could be used to separate multiple audio sources from a single microphone signal, the Locally Linear Embedding algorithm (LLE) with Lawrence Saul, which revolutionized the field of dimensionality reduction, Stochastic Neighborhood Embedding (SNE) and Neighborhood Component Analysis (NCA), and the astrometry.net system with David W. Hogg that can take any picture of the sky from any source, and instantly identify the location, orientation, and magnification of the image, as well as name each object (star, galaxy, nebula) it contains.

Sam had a singular gift: to him, any complex concept was naturally reduced to a simple set of ideas, each of which had clear analogies in other (often very distant) realms. This gift allowed him to explain the key idea behind anything in just a few minutes. Combined with contagious enthusiasm, this made him an unusually gifted teacher and speaker. His talks and discussions were clear and highly entertaining. His tutorial lectures on graphical models and metric learning, available on video at videolectures.net, have been viewed over 25,000 times. He would often begin group meetings by giving a puzzle, the solution of which was always beautiful, enlightening, or hilarious.

Many members of the research community became friends with Sam, because of his warm and friendly personality, his communicative smile, and his natural inclination to engagement and enthusiasm. Sam inspired many students to pursue a career in research, and to focus their research on machine learning and artificial intelligence because of his broad interests, his clear-sightedness, his sense of humor, his warmth and his infectious enthusiasm.

In the last year of his life, Sam was battling severe depression but his wonderfully professional demeanor concealed this from most of his friends and colleagues. He is greatly mourned by his colleagues and students at NYU, who extend their sympathy to his many friends in the broader research community, especially at the University of Toronto, the Gatsby Neuroscience Unit, and Google Research. Most of all, we express our deepest sympathy to his wife Meredith, his twin baby daughters Aya and Orli, and his father Shoukry.

Yann LeCun, David Hogg, Zoubin Ghahramani, Geoffrey Hinton.

Partha Niyogi, 1967-2010

Partha Niyogi, Louis Block Professor in Computer Science and Statistics at the University of Chicago, passed away on October 1, 2010 after a struggle with brain cancer. He was a rising star in machine learning and his research in language learning was seminal.

He was born July 31, 1967, in Calcutta, India and received his bachelor's degree in electrical engineering from the Indian Institute of Technology in New Delhi, India, in 1989. His thesis was on the automatic recognition of beats on the tabla, a percussion instrument of northern India. This led him to the study of perception, recognition and learning, as well as acoustics, music and language. He earned his master's degree in 1992 and his doctorate in 1995 from the Massachusetts Institute of Technology under the direction of Tomaso Poggio. He was a postdoctoral fellow and research associate in MIT's Brain and Cognitive Science Department, and then joined the technical staff at Bell Laboratories in Murray Hill, N.J. before moving to the University of Chicago in 2000.

Much of his research addressed problems related to language acquisition and how they might be replicated in a machine. His ultimate goal was to build computer systems that could interact with and learn from humans. He was the author of The Computational Nature of Language Learning and Evolution that it brings together in a common framework the problems of language learning (in linguistics) and function learning (in learning theory and neural networks) and shows how one can treat them with the same style of analysis. This was a seminal contribution, especially given that the communities working in these two fields seldom overlap and often misrepresent or misunderstand each other's work.

Continued on page 115

UPCOMING CONFERENCES

2011
GRANADA
SPAIN

2012 - 2014
LAKE TAHOE
NEVADA

In another major line of research, Partha made fundamental contributions to geometrically based methods for inferring hidden patterns from complex data sets related to a variety of problems, including those in image recognition and analysis of spoken language. Together with Federico Girosi he brought together techniques from approximation theory and statistics to address the key question of generalization performance of learning algorithms. They showed how in order to understand the complexity of the problem one needed results from both approximation and statistical learning theory. It is a paper that has helped to get many mathematicians involved in learning theory, presenting them with a clean description of a challenging problem. Together with Kah Kay Sung, Partha studied one of the main ways to reduce the complexity of learning, which is incorporating prior knowledge in creating or choosing new examples. He formulated theoretically principled frameworks for "active learning" to choose new examples in function learning or pattern recognition. This approach has been applied successfully to problems of object detection in computer vision.

Partha was a beacon of clarity in a complex intellectual landscape where it is easy to get lost in the jungle of technical details and drown in an ocean of marginal improvements. He was curious of all things scientific, and never afraid of learning new tools, no matter how complex they were, if he thought they were needed. The integrity he showed on the scientific field was mirrored in all aspects of his life. He was guided and motivated by a deep respect and concern for others. Always an advocate for the disadvantaged and the oppressed, he was a passionate observer of the political reality and a strong and informed debater. If his intellectual tools were sharp like a cold razor, his personal manners were gentle and round, shrouded in a deep generosity and a warm smile.

Tomaso Poggio, Federico Girosi, and Terry Sejnowski

REVIEWERS

Pieter Abbeel
Jacob Abernethy
Margarita Ackerman
Ryan Adams
Alekh Agarwal
Deepak Agarwal
Shivani Agarwal
John Mark Agosta
Kunal Agrawal
Edo Airoldi
Shotaro Akaho
Mauricio Alvarez
Carlos Alzate
Christophe Andrieu
Andras Antos
Cedric Archambeau
Andreas Argyriou
Maleki Arian
Artin Armagan
Hideki Asoh
Jean-Yves Audibert
Peter Auer
Doru C Balcan
Maria Florina Balcan
Dana Ballard
Laura Balzano
Arindam Banerjee
Richard Baraniuk
David Barber
Evgeniy Bart
Sumit Basu
Peter Battaglia
Jonathan Baxter
Jeff Beck
Mikhail Belkin
Shai Ben-David
Asa Ben-Hur
Samy Bengio
Philip Berens
Alex Berg
Tamara Berg
Amine Bermak
Michel Besserve
Matthias Bethge
Alina Beygelzimer
Jinbo Bi
Misha Bilenko
Jeff A Bilmes
Horst Bischof
Andrew Blake
Mathieu Blanchette
Matthew B Blaschko
David Blei
John Blitzer
Andrew Bolstad
Karsten Borgwardt
Leon Bottou
Matthew Botvinick
Guillaume Bouchard
Alexandre Bouchard-Cote
Jordan Boyd-Graber
Tim Brady
Ulf Brefeld
Emma Brunskill
Sebastien Bubeck
Wolfram Burgard
Chris Burges
Colin Campbell
Kevin Canini
Stephane Canu
Olivier Cappe
Constantine Caramanis
Lawrence Carin
Miguel A Carreira-Perpinan
Melissa Carroll
Rich Caruana
Carlos Carvalho
Gert Cauwenberghs
Gavin Cawley
Lawrence Cayton
Nicolo Cesa-Bianchi
Volkan Cevher
Karthekeyan Chandrasekaran
Jonathan Chang
Olivier Chapelle
Gal Chechik
Yixin Chen
David Chiang
Elisabetta Chicca
Bhattacharyya Chiranjib
Andreas Christmann
Mario Christoudias
Wei Chu
Ken Church
Anton Civit
Mark Coates
Ruben Coen-Cagli
Ronan Collobert
Greg Corrado
Corinna Cortes
Timothee Cour
Aaron Courville
Koby Crammer
Mark Craven
Daniel Cremers
John P Cunningham
Marco Cuturi
Arnak Dalalyan
David Danks
Sanjoy Dasgupta
Hal Daume III
Mark Davenport
Jason Davis
Nathaniel Daw
Peter Dayan
Tijl De Bie
Luc De Raedt
Virginia De Sa
Dennis Decoste
Marc P Deisenroth
Ofer Dekel
Olivier Delalleau
Jia Deng
Renaud Detry
Inderjit Dhillon
James Dicarlo
Tom Diethe
Thomas Dietterich
Chris Ding
Francesco Dinuzzo
Carlos Diuk-Wasser
Brent Doiron
Piotr Dollar
Finale P Doshi-Velez
Arnaud Doucet
Petros Drineas
Jan Drugowitsch
John Duchi
Piotr Dudek
Miroslav Dudik
Delbert Dueck
David Dunson
Jason Eisner
Gal Elidan
Charles Elkan
Yaakov Engel
Barbara Engelhardt
Tom Erez
Brian Eriksson
Damien Ernst
Eleazar Eskin
Li Fei-Fei
Jacob Feldman
Rob Fergus
Alan Fern
Vittorio Ferrari
Steve Fienberg
Mario Figueiredo
Jozsef Fiser
Andrew Fitzgibbon
David Forsyth
Charless Fowlkes
Emily Fox
Rina Foygel
Vojtech Franc
Alexander Fraser
Peter Frazier
Jeremy Freeman
William Freeman
Yoav Freund
Michael Friedlander
Karl Friston
Kenji Fukumizu
Thomas Gabel
Kuzman Ganchev
Surya Ganguli

REVIEWERS

Peter Gehler
Claudio Gentile
Samuel Gershman
Sebastian Gerwinn
Olivier Gevaert
Mohammad Ghavamzadeh
Ran Gilad-Bachrach
Benoit Girard
Mark Girolami
Inmar Givoni
Sally A Goldman
Manuel Gomez Rodriguez
Stephen Gould
Joao Graca
Thore Graepel
Kristen Grauman
Alexander G Gray
Russ Greiner
Remi Gribonval
Jim Griffin
Tom Griffiths
Moritz Grosse-Wentrup
Amit Gruber
Steffen Gruenewaelder
Todd Gureckis
Dilan GoRuR
Gholamreza Haffari
Patrick Haffner
Attias T Hagai
Jihun Hamm
Firas Hamze
Lauren Hannah
Steve Hanneke
zaid Harchaoui
Stefan Harmeling
Matthew Harrison
Jarvis Haupt
Milos Hauskrecht
Tamir Hazan
Xiaofei He
Martial Hebert
Matthias Hein
Katherine Heller
Mark Herbster
Daniel Hernandez-Lobato
Tom Heskes
Jeremy Hill
Geoffrey Hinton
Jesse Hoey
Matt Hoffman
Matthew Hoffman
Thomas Hofmann
Steven Chu-Hong Hoi
Derek Hoiem
Jean F Honorio
Paul Horton
Eric Horvitz
Matthew Howard
Chun-Nan Hsu
Bert Huang
Frank Hutter
Quentin Huys
Christian Igel
Alexander Ihler

Shiro Ikeda
Park Il
Tsang Ivor Wai-Hung
Tommi Jaakkola
Nathan Jacobs
Robert Jacobs
Frank Jaekel
Dominik Janzing
Tony Jebara
Stefanie Jegelka
Rong Jin
Jason K Johnson
Michael I Jordan
Frederic Jurie
Leslie Kaelbling
Adam Kalai
Takafumi Kanamori
Ashish Kapoor
Bert Kappen
Nikos Karampatziakis
Yan Karklin
Hisashi Kashima
Samuel Kaski
Yoshinobu Kawahara
Motoaki Kawanabe
Michael Kearns
Sathiya Keerthi
Balazs Kegl
Charles Kemp
Kristian Kersting
Roozbeh Kiani
Sergey Kirshner
Jyrki Kivinen
David Knowles
Jens Kober
Kilian Koepsell
Konrad Koerding
Daphne Koller
Vladimir Kolmogorov
J. Zico Kolter
Risi Kondor
James Kozloski
Andreas Krause
Josic Kresimir
Balaji Krishnapuram
Oliver Kroemer
Alex Kulesza
Brian Kulis
M. Pawan Kumar
Sanjiv Kumar
James Kwok
Sebastien Lahaie
Peng Lai
Christoph Lampert
John Langford
Ivan Laptev
Pavel Laskov
Francois Laviolette
Alessandro Lazaric
Svetlana Lazebnik
Nicolas Le Roux
Matt Lease
Guy Lebanon
Honglak Lee

Tai Sing Lee
Wee Sun Lee
Yoonkyung Lee
Yuh-Jye Lee
Victor Lempitsky
Mate Lengyel
Vincent Lepetit
Alex Lerchner
Christina Leslie
Roger Levy
Li-Jia Li
Lihong Li
Ping Li
Wei Li
Xiao Li
Chih-Jen Lin
Hsuan-Tien Lin
Yuanqing Lin
Gustavo Linan
Alejandro Linares-Barranco
Bernabe Linares-Barranco
Christoph Lippert
Michael Littman
Ce Liu
Han Liu
Shih-Chii Liu
Tie-Yan Liu
Yan Liu
Dan Lizotte
Phil Long
Manuel Lopes
Aurelie C Lozano
Hongjing Lu
Tyler Lu
Jorg Lucke
Elliot A Ludvig
Jakob Macke
Lester Mackey
Sridhar Mahadevan
Julien Mairal
Gideon Mann
Shie Mannor
Vikash Mansinghka
Yishay Mansour
Davy Manuel
Yi Mao
Benjamin Marlin
Ruben Martinez-Cantin
Jiri Matas
Clerc Maureen
Jon D McAuliffe
Rob Mcculloch
Chris Meek
Nishant A Mehta
Ron Meir
Francisco S Melo
Hirsch Michael
Brian Milch
Kurt Miller
David Mimno
Djordje Mitrovic
Andriy Mnih
Daichi Mochihashi
Claire Monteleoni

Greg Mori
Quaid Morris
Sayan Mukherjee
Remi Munos
Art Munson
Art Munson
Noboru Murata
Kevin Murphy
Iain Murray
Boaz Nadler
Hariharan Narayanan
Rama Natarajan
Bobak Nazer
Sahand Negahban
Jonathan Nelson
Bernhard Nessler
David Newman
Andrew Ng
Long Nguyen
Duy Nguyen-Tuong
Hannes Nickisch
Alex Niculescu-Mizil
Juan Carlos Niebles
Yael Niv
William S Noble
Rob Nowak
Sebastian Nowozin
Guillaume R Obozinski
Uwe Ohler
Bruno A Olshausen
Cheng Soon Ong
Sylvie C.W. Ong
Manfred Opper
Francesco Orabona
Gergo Orban
Peter Orbanz
Satoshi Osaga
Umut Ozertem
Benjamin Packer
David Page
John Paisley
David Pal
Jason A Palmer
Stephanie Palmer
Liam Paninski
Ulrich Paquet
Devi Parikh
Sylvain Paris
Ronald Parr
Emilio Parrado-Hernandez
Kristiaan Pelckmans
Fernando Pereira
Francisco Pereira
Fernando Perez-Cruz
Jan Peters
Slav Petrov
Jean-Pascal Pfister
Justus Piater
Jonathan W Pillow
Joelle Pineau
Robert Pless
Russell Poldrack
Massi Pontil
Pascal Poupart

REVIEWERS

Doina Precup
Philippe Preux
Yanjun Qi
Yuan Qi
Ariadna Quattoni
Michael Rabbat
Agnes Radl
Maxim Raginsky
Ali Rahimi
Alain Rakotomamonjy
Haefner M Ralf
Deva Ramanan
Marc'Aurelio Ranzato
Vinayak Rao
Garvesh Raskutti
Nathan Ratliff
Magnus Rattray
Pradeep Ravikumar
Mark Reid
Lev Reyzin
Elisa Ricci
Irina Rish
Abel Rodriguez
Heiko Roeglin
Lorenzo Rosasco
Michal Rosen-Zvi
Saharon Rosset
Afshin Rostamizadeh
Aaron Roth
Arnd Roth
Stefan Roth
Yasser Roudi
Cynthia Rudin
Bryan Russell
Nicole Rust
Daniil Ryabko
Hannes P Saal
Sivan Sabato
Kate Saenko
Yvan Saeys
Maneesh Sahani
Toshimichi Saito
Ruslan Salakhutdinov
Ruslan Salakhutdinov
Mathieu Salzmann
Dimitris Samaras
Adam Sanborn
Sujay Sanghavi
Guido Sanguinetti
Scott Sanner
Ben Sapp
Stefan Schaal
Robert E Schapire
Tobias Scheffer
Katya Scheinberg
Bruno Scherrer
Bernt Schiele
Alexander Schliep
Jeff Schneider
Paul Schrater
Dale Schuurmans
Odelia Schwartz
Clayton Scott
James Scott
MicheLe Sebag
Per B Sederberg
Matthias W Seeger
Yevgeny Seldin
Teresa Serrano-Gotarredona
Thomas Serre
Patrick Shafto
Mohak Shah
Gregory Shakhnarovich
Ohad Shamir
Mohit Sharma
Tatyana Sharpee
Blake Shaw
Or Sheffet
Dan Sheldon
Xiaotong Shen
Nino Shervashidze
Shirish Shevade
Tadashi Shibata
Shohei Shimizu
Pannaga Shivaswamy
Jon Shlens
Sajid Siddiqi
Ricardo Silva
Eero Simoncelli
Eero P Simoncelli
Vikas Sindhwani
Yoram Singer
Aarti Singh
Ajit Singh
Kaushik Sinha
Mathieu Sinn
Josef Sivic
William Smart
Cristian Sminchisescu
Noah Smith
Alexander J Smola
Peter Sollich
Fritz Sommer
Le Song
Soeren Sonnenburg
David Sontag
Finnegan Southey
Matthijs Spaan
Suvrit Sra
Nathan Srebro
Karthik Sridharan
Bharath Sriperumbudur
Oliver Stegle
Florian Steinke
Ingo Steinwart
Mark Steyvers
Alan Stocker
Amos J Storkey
Amarnag Subramanya
Erik Sudderth
Masashi Sugiyama
Ilya Sutskever
Rich Sutton
Johan Suykens
Taiji Suzuki
Umar Syed
Sandor Szedmak
Csaba Szepesvari
Matt Taddy
Prasad Tadepalli
Ameet Talwalkar
Vincent Tan
Toshiyuki Tanaka
Marshall Tappen
Ben Taskar
Nikolaj Tatti
Graham Taylor
Yee Whye Teh
Joshua Tenenbaum
Christian Thurau
Jo-Anne Ting
Michalis Titsias
Michael Todd
Surya Tokdar
Andrea Tolias
Ryota Tomioka
Zhang Tong
Antonio Torralba
Lorenzo Torresani
Marc Toussaint
Volker Tresp
Alessandro Treves
Bill Triggs
Wilson Truccolo
Koji Tsuda
Zhuowen Tu
Rich Turner
Naonori Ueda
Lyle Ungar
Raquel Urtasun
Sam Usalam
Manik Varma
Andrea Vedaldi
Jakob Verbeek
Jean-Philippe Vert
S.V.N. Vishwanathan
Nikos Vlassis
Ed Vul
Christian Walder
Thomas Walsh
Jack Wang
Liwei Wang
Larry Wasserman
Markus Weimer
Kilian Weinberger
David Weiss
REVIEWERS 233
Yair Weiss
Max Welling
Rebecca Willett
Jason William
Ross S Williamson
Sinead Williamson
Robert Wilson
David Wingate
John Winn
Ole Winther
David Wipf
Frank Wood
Steve Wright
Mingrui Wu
Wei Wu
Ying-Nian Wu
Lin Xiao
Eric Xing
Ming-Hsuan Yang
Jieping Ye
Yiming Ying
Ni Yizhao
Angela Yu
Byron Yu
Chun-Nam Yu
Kai Yu
Shi Yu
Shipeng Yu
Yisong Yue
Carlos Zamarre no-Ramos
Luke Zettlemoyer
Hongyuan Zha
Jian Zhang
Kun Zhang
Ya Zhang
Alice Zheng
Dengyong Zhou
Zhi-Hua Zhou
Hongtu Zhu
Jun Zhu
Xiaojin (Jerry) Zhu
Martin Zinkevich
Larry Zitnick
Onno Zoeter
Alexandre d'Aspremont
Nando de Freitas
David A van Dyk
Laurens van der Maaten

AUTHOR INDEX

AUTHOR INDEX

AUTHOR INDEX

N I
2010
P S

N I
2010
P S

N I
2010
P S