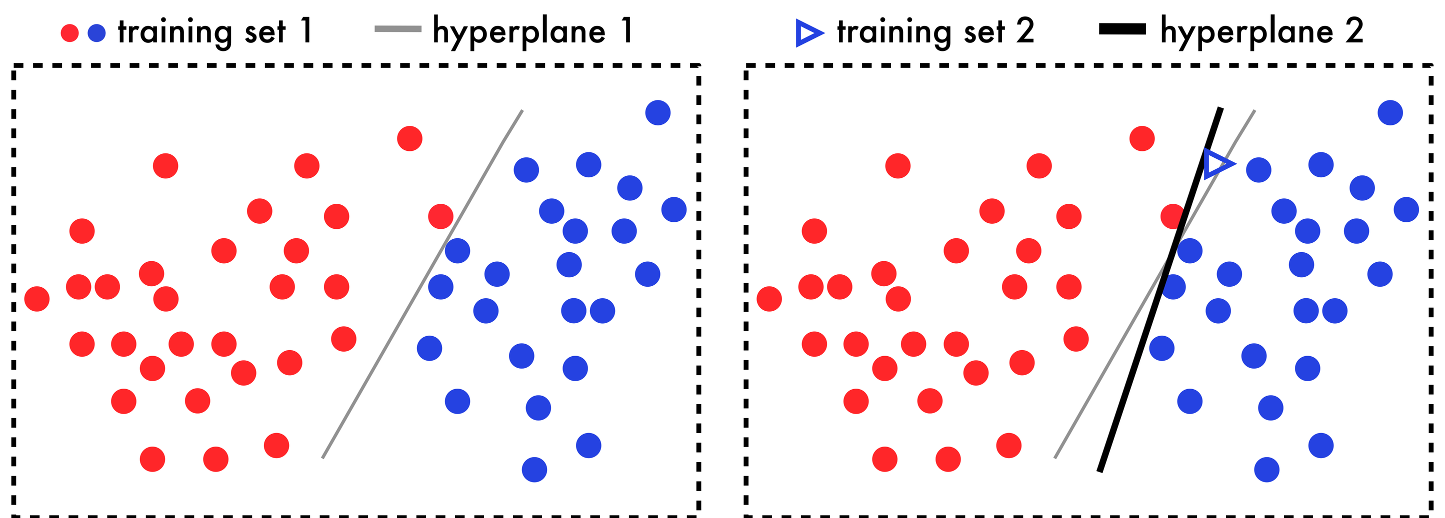
Differentially Private Bayesian Optimization

Matt J. Kusner[1], Jacob R. Gardner[1,2], Roman Garnett[1], Kilian Q. Weinberger[1,2]

[1]Department of Computer Science & Engineering, Washington University in St. Louis, USA

[2]Department of Computer Science, Cornell University, USA

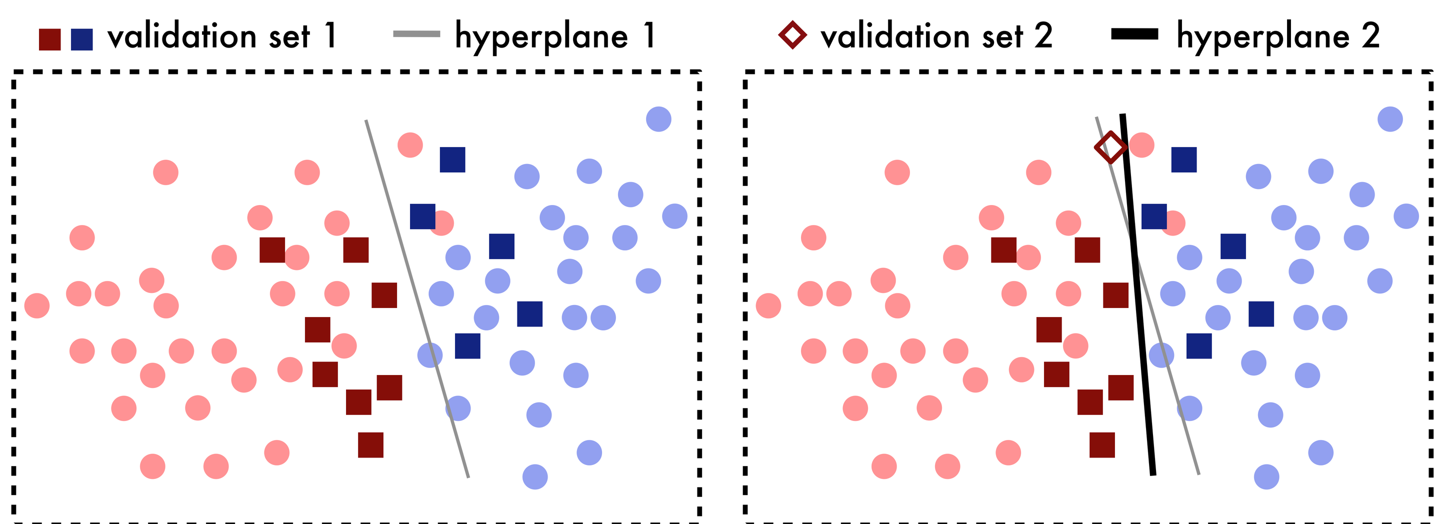
Privacy in Learning



learning can reveal information about the **training set!**

[Kasiviswanathan et al., 2008; Dwork & Lei, 2009; Chaudhuri et al., 2011; Chaudhuri & Hsu, 2012; Jain et al., 2012; Kifer et al., 2012; Smith & Thakurta, 2013; Jain & Thakurta, 2014; Bassily et al., 2014]

Privacy in Hyperparameter Tuning



selecting hyperparameters can reveal information about the **validation set!**

[Chaudhuri & Vinterbo, 2013]

Selecting Hyperparameters

e.g., RBF Kernel SVM has hyperparameters: (λ, σ)

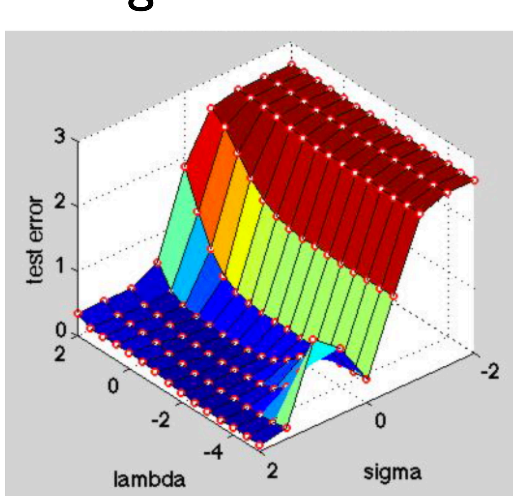
regularization trade-off

kernel width

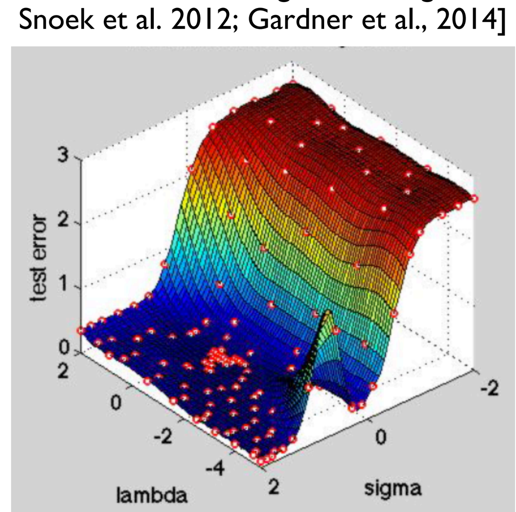
grid search

Bayesian Optimization (BO)

[Hutter et al. 2011; Bergstra & Bengio, 2012; Snoek et al. 2012; Gardner et al., 2014]

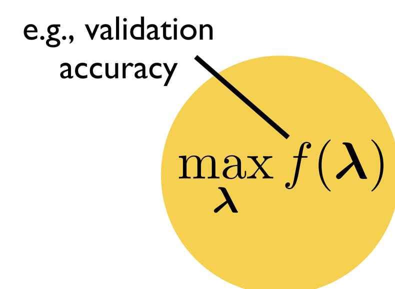


prior work:
private grid search



this work:
private BO

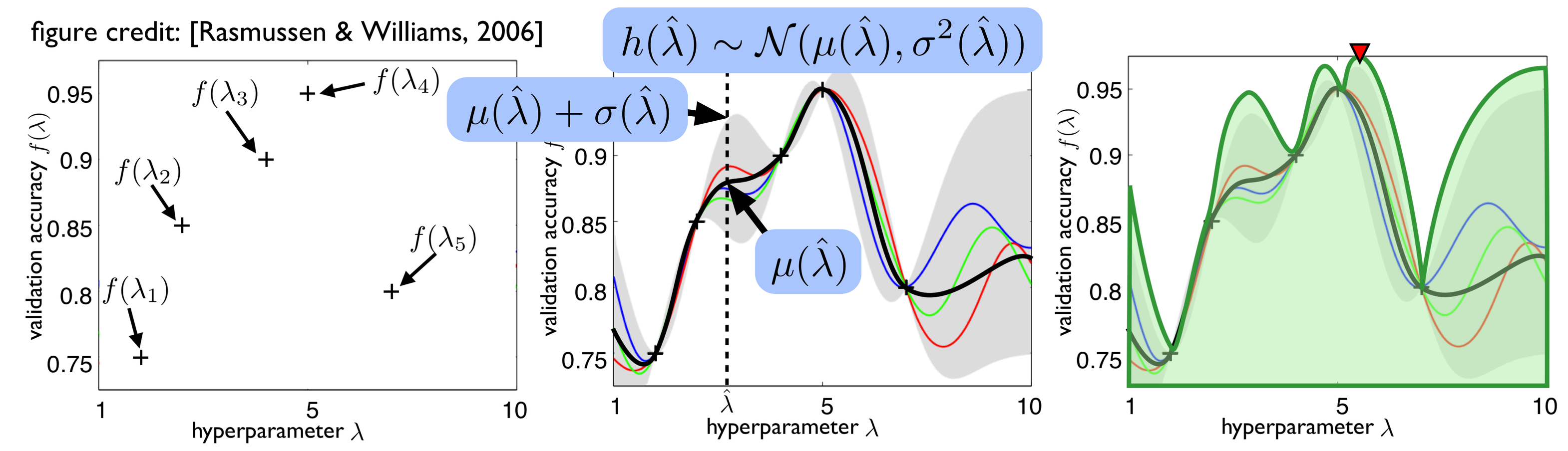
Bayesian Optimization



$f(\lambda)$ is **very expensive** to compute

$f(\lambda)$ is **nonconvex**

idea: model $f(\lambda)$ with an easy-to-evaluate surrogate



1. given samples $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_5)$

2. fit a Gaussian Process $h \sim \mathcal{GP}(0, k(\boldsymbol{\lambda}, \boldsymbol{\lambda}'))$

3. Maximize UCB $\mu(\lambda) + \sqrt{\beta}\sigma(\lambda)$

[Srinivas et al., 2010]

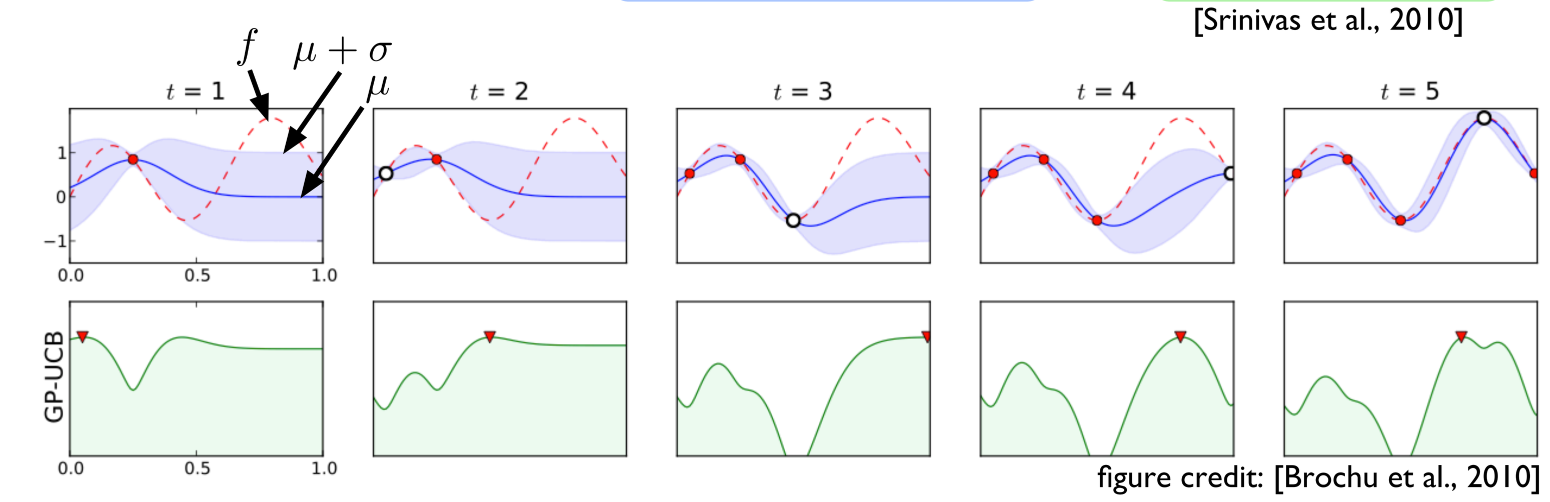


figure credit: [Brochu et al., 2010]

Return: best $f(\lambda)$ seen so far

How to make this procedure private?

Differential Privacy

[Dwork et al., 2006]

A formalization of "privacy through randomness"

BO validation set 1 $\mathcal{V}$

$\mathcal{A}$

age: 20	gender: M	SSN: 42...2
age: 91	gender: M	SSN: 21...0
age: 82	gender: F	SSN: 46...7
age: 39	gender: F	SSN: 91...3
⋮	⋮	⋮

$\approx$

BO validation set 2 $\mathcal{V}'$

$\mathcal{A}$

age: 20	gender: M	SSN: 42...2
age: 91	gender: M	SSN: 21...0
age: 56	gender: F	SSN: 72...4
age: 39	gender: F	SSN: 91...3
⋮	⋮	⋮

Definition 1. *A randomized algorithm $\mathcal{A}$ is (ϵ, δ)-**differentially private** for $\epsilon, \delta \geq 0$ if for all $f(\lambda) \in \mathrm{Range}(\mathcal{A})$ and for all neighboring datasets $\mathcal{V}, \mathcal{V}'$ (i.e., such that $\mathcal{V}$ and $\mathcal{V}'$ differ in the value of one record) we have that*

$$\Pr[\mathcal{A}(\mathcal{V}) = f(\lambda)] \leq e^{\epsilon} \Pr[\mathcal{A}(\mathcal{V}') = f(\lambda)] + \delta.$$

properties

- k (ϵ, δ)-differentially private runs is $(k\epsilon, k\delta)$-diff. private
- post-processing doesn't decrease privacy
- immune to common attacks (e.g., linkage, differencing attacks)

A Private Mechanism

Goal: mask change in $\mathcal{A}$ when run on $\mathcal{V}$ vs. $\mathcal{V}'$

Definition 2. *The **global sensitivity** of an algorithm $\mathcal{A}$ over all neighboring datasets $\mathcal{V}, \mathcal{V}'$ (i.e., $\mathcal{V}, \mathcal{V}'$ differ by the value of one record) is*

$$\Delta_{\mathcal{A}} \triangleq \max_{\mathcal{V}, \mathcal{V}'} \|\mathcal{A}(\mathcal{V}) - \mathcal{A}(\mathcal{V}')\|_1.$$

is $(\epsilon, 0)$-differentially private

Laplace Mechanism

[Dwork et al., 2006]

1. Draw $\omega \sim \mathrm{Laplace}(0, \Delta_{\mathcal{A}}/\epsilon)$
2. Release $\mathcal{A}(\mathcal{V}) + \omega$

Our Results

Assumption 1: $f(\boldsymbol{\lambda})$ and $f'(\boldsymbol{\lambda})$ are GP distributed

Theorem 1. *Given Assumption 1 and the assumptions in Theorem 2 of de Freitas et al. (2012), for neighboring datasets $\mathcal{V}, \mathcal{V}'$ we have the following global sensitivity bound,*

$$|f'(\lambda_T) - f(\lambda_T)| \leq Ae^{-\frac{T\tau}{(\log T)^{d/4}}} + c$$

w.p. at least $1 - \delta$ for c, d as defined in the paper and given constants A and τ in de Freitas et al. (2012), after T rounds of BO.

in the presence of noise...

Theorem 2. *Given Assumption 1, and neighboring $\mathcal{V}, \mathcal{V}'$, we have the following global sensitivity bound for the maximum v (noisy validation accuracy) after BO, w.p. $\geq 1 - \delta$*

$$|\max_{t \leq T} v'_t - \max_{t \leq T} v_t| \leq \frac{\sqrt{C_1 \beta_T \gamma_T}}{\sqrt{T}} + c + q.$$

(for c, q, C_1, β_T as defined in the paper), where γ_T is bounded above for the squared exponential and Matérn kernels (Srinivas et al., 2010).

Assumption 2: $f(\boldsymbol{\lambda})$ and is L-Lipschitz (additionally training loss is L-Lipschitz and convex)

Theorem 3. *Given Assumption 2, for neighboring $\mathcal{V}, \mathcal{V}'$ and arbitrary $\lambda < \lambda'$ (and $\lambda_{\min}$ is the smallest hyperparameter) we have that,*

$$|f(\lambda) - f'(\lambda')| \leq \frac{(\lambda' - \lambda)L}{\lambda' \lambda} + \min\left\{\frac{g^*}{m}, \frac{L}{m\lambda_{\min}}\right\}$$

where L is the Lipschitz constant of f, m is the size of $\mathcal{V}$, and g is defined in the paper.

References

1. S. A. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? In Proc. of FOCS, 2008.
2. C. Dwork and J. Lei. Differential privacy and robust statistics. In Proceedings of the 41st ACM Symposium on Theory of Computing (STOC), 2009.
3. Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. Differentially private empirical risk minimization. JMLR, 12:1069–1109, 2011.
4, K. Chaudhuri and D. Hsu. Convergence rates for differentially private statistical estimation. ICML, 2012.
5. Jain, Prateek, Kothari, Pravesh, and Thakurta, Abhradeep. Differentially private online learning. In COLT, 2012.
6. Daniel Kifer, Adam Smith, and Abhradeep Thakurta. Private convex empirical risk minimization and high-dimensional regression. In COLT, 2012.
7. Adam Smith and Abhradeep Thakurta. Nearly optimal algorithms for private online learning in full-information and bandit settings. In NIPS, 2013.
8. Jain, P., and Thakurta, A. (2014).(Near) Dimension independent risk bounds for differentially private learning, ICML, 2013.
9. Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In FOCS, pages 464–473. IEEE, October 18–21 2014.
10. Kamalika Chaudhuri and Staal A. Vinterbo. A stability-based validation procedure for differentially private machine learning. pages 2652–2660, 2013.
11. Hutter, F., Hoos, H. H., and Leyton-Brown, K.. Sequential model-based optimization for general algorithm configuration. In Learning and Intelligent Optimization, pp. 507–523. Springer, 2011.
12. Bergstra, James and Bengio, Yoshua. Random search for hyper-parameter optimization. JMLR, 13:281–305, 2012.
13. Snoek, Jasper, Larochelle, Hugo, and Adams, Ryan P. Practical bayesian optimization of machine learning algorithms. In NIPS, pp. 2951–2959, 2012.
14.Gardner, Jacob, Kusner, Matt, Xu, Zhixiang, Weinberger, Kilian, and Cunningham, John. Bayesian optimization with inequality constraints. In ICML, pp. 937–945, 2014.
15, Dwork, Cynthia, McSherry, Frank, Nissim, Kobbi, and Smith, Adam. Calibrating noise to sensitivity in private data analysis. In Theory of Cryptography, pp. 265–284. Springer, 2006
16.Rasmussen, Carl Edward and Williams, Christopher K. I. Gaussian processes for machine learning. 2006.
17. Eric Brochu, Vlad M. Cora, and Nando de Freitas. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. pre-print, 2010. arXiv:1012.2599.
18.Srinivas, Niranjan, Krause, Andreas, Kakade, Sham M, and Seeger, Matthias. Gaussian process optimization in the bandit setting: No regret and experimental design. In ICML, 2010.
19. de Freitas, Nando, Smola, Alex, and Zoghi, Masrour. Exponential regret bounds for gaussian process bandits with deterministic observations. In ICML, 2012.