ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

Recent Advances in Nonlinear Dimensionality Reduction, Manifold and Topological Learning

Axel Wismüller[1], Michel Verleysen[2], Michael Aupetit[3], and John A. Lee[4] *

1 - University of Rochester, Depts. of Radiology and Biomedical Engineering,
601 Elmwood Avenue, Rochester, NY 14642-8648, USA
2 - Université catholique de Louvain, Machine Learning Group,
Place du Levant 3, B-1348 Louvain-la-Neuve, Belgium
3 - Commissariat à l'Energie Atomique (CEA) - DAM
Département Analyse Surveillance Environnement,
BP 12, 91680 Bruyères-le-Châtel, France
4 - Université catholique de Louvain,
Dept. of Molecular Imaging and Experimental Radiotherapy,
Avenue Hippocrate 55, B-1200 Bruxelles, Belgium

Abstract. The ever-growing amount of data stored in digital databases raises the question of how to organize and extract useful knowledge. This paper outlines some current developments in the domains of dimensionality reduction, manifold learning, and topological learning. Several aspects are dealt with, ranging from novel algorithmic approaches to their real-world applications. The issue of quality assessment is also considered and progress in quantitive as well as visual crieria is reported.

1 Introduction

The transformation of high-dimensional data to lower-dimensional spaces has been a topic of interest for more than a century. Dimensionality reduction pursues several goals: visualizing data in 2- or 3-dimensional spaces, extracting a limited number of relevant features from the original ones, or even simply removing some noise from the data. Principal component analysis (PCA) is probably the first attempt towards dimensionality reduction. It has long been the only method available and used by practitioners, before the advent of multidimensional scaling (MDS) and other more complex techniques. The issues of data representation and dimensionality reduction have been addressed by several communities. PCA and MDS were essentially developed by socio-psychologists. The machine learning community then took the lead; in this community, (nonlinear) dimensionality reduction is often referred to as manifold learning. The topic is also tightly connected to graph embedding techniques.

During the last decades, two revolutions greatly influenced the development of the field: the need to process large datasets, and the advent of nonlinear dimensionality reduction. Nonlinear methods are by definition more powerful than linear methods, as they make fewer hypotheses about the model and/or the manifold. At the same time, they face more difficulties: the need to define

*All authors contributed equally to this publication. J.A. Lee is a Research Associate with the Belgian National Fund of Scientific Research (FNRS).

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

proper objective criteria compatible with the application goal, the use of optimization techniques, the need for evaluation criteria, etc. DR methods can be categorized by their optimization scheme, which can be spectral or non-spectral. As a matter of fact, not all cost functions can be cast within the framework of an eigenproblem. The appealing theoretical properties of spectral techniques, such as the guarantee to find the global optimum, are thus counterbalanced by the bigger flexibility offered by non-spectral optimization.

Dimensionality reduction amounts to associating low-dimensional coordinates to data items, while preserving structural information as much as possible. The latter can be expressed in practice by pairwise distances or, more generally, by (dis)similarities. The methods typically differ in their definition of (dis)similarity measure, and in the weighting of small versus large similarity discrepancies in the cost function. Alternatively, methods can also be driven by topology preservation. They can attempt to reproduce distance ranks in the low-dimensional space, for instance.

The variety of manifold learning techniques also raises the issue of their validation with quality criteria that are both meaningful with respect to the considered application and independent of the compared methods' cost functions.

The remainder of this paper presents a selection of state-of-the-art methods of manifold learning based on distances and similarities (Section 2), as well as recent topology-preserving tools (Section 3). Section 4 deals with quality criteria.

2 Distances and similarities to reduce the dimensionality

Principal component analysis [1, 2, 3] (PCA) is often viewed as a method of representing data set $\mathbf{\Xi} = [\boldsymbol{\xi}_i]_{1 \leq i \leq N}$ in a low-dimensional space while preserving a maximal fraction of the data set variance. Actually, one can also show that PCA is equivalent to classical metric multidimensional scaling [4, 5, 6] (MDS). These two techniques are dual: while PCA involves the covariance matrix $\mathbf{C_{\Xi\Xi}} = \frac{1}{N}(\mathbf{\Xi} - \frac{1}{N}\mathbf{\Xi 1 1}^T)^T(\mathbf{\Xi} - \frac{1}{N}\mathbf{\Xi 1 1}^T)$, MDS relies on the corresponding centered Gram matrix of pairwise inner products $\mathbf{G} = (\mathbf{\Xi} - \frac{1}{N}\mathbf{\Xi 1 1}^T)(\mathbf{\Xi} - \frac{1}{N}\mathbf{\Xi 1 1}^T)^T$. In both cases, a spectral decomposition is used to find low-dimensional coordinates $\mathbf{X} = [\mathbf{x}_i]_{1 \leq i \leq N}$ that correspond to least-square approximations of the mentioned matrices. Formally, these methods find the global optimum of $\min_{\mathbf{X}} \|\mathbf{C_{\Xi\Xi}} - \mathbf{C_{XX}}\|_2$ and $\min_{\mathbf{X}} \|\mathbf{G_{\Xi\Xi}} - \mathbf{G_{XX}}\|_2$, respectively, where $\|\cdot\|_2$ denotes the Frobenius norm. The evolution of classical metric MDS towards nonlinear variants the close relationship between inner products and Euclidean distances. Translating the preservation of inner products into the preservation of the corresponding distances offers a much intuitive and versatile formulation. At the expense of replacing the spectral decomposition with more general optimization tools such as gradient descent, the cost function that formalizes distance preservation can be extended and defined in more flexible ways. For example, $\min_{\mathbf{X}} \|\mathbf{G_{\Xi\Xi}} - \mathbf{G_{XX}}\|_2$ can be replaced with $\min_{\mathbf{X}} \sum_{i<j} w_{ij}(\delta_{ij} - d_{ij})^2$, where the minimized quantity is often called the *stress*, w_{ij} are weights, and distances are denoted by $\delta_{ij} = \|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2$ and $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$. Weight w_{ij} modulates the importance given to the preservation of small distances versus larger ones. This principle is applied in Sammon's nonlinear mapping [7], which fa-

vors the preservation of small distances. In this case, w_{ij} is defined to be equal to $1/\delta_{ij}$. Giving less importance to large distances is supposed to allow data to unfold, in order to make their embedding easier in a low-dimensional space. Curvilinear component analysis [8] follows a similar approach, with the noticible difference that $w_{ij} = f(d_{ij}/\sigma)$, where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a decreasing function of its argument and σ is a neighborhood width. Although at first glance it looks very similar to Sammon's mapping, CCA shows a completely different behavior, due to the dependence of the weights upon the distance in the *low*-dimensional space. This pecularity gives CCA the ability to tear manifolds, which improves their unfolding. Recent studies about quality assesment of dimensionality reduction [9, 10] (see also Section 4) have shown that embedding errors can be divided into two types: either distant points are erroneously embedded close to each other or initially nearby points are mapped too far away. Within this framework, Sammon's mapping and CCA can be shown to tolerate more easily the one type or the other. These antagonist behaviors have been combined in hybrid methods such as Venna's local multidimensional scaling [11, 12], where $w_{ij} = \lambda f(d_{ij}/\sigma) + (1 - \lambda) f(\delta_{ij}/\sigma)$. Parameter λ controls the balance between the two types of errors.

All previously mentioned methods can be extended to other metrics than the Euclidean norm. The most famous example is undoubtedly Isomap [13], which amounts to applying classical metric MDS to a matrix of pairwise geodesic distances. Geodesic distances are measured along the underlying manifold and thus enable a better unfolding. In practice, geodesic distances are approximated by computing shortest paths in a Euclidean graph corresponding to K-ary neighborhoods or ϵ-balls [14]. Geodesic distances have been used in Sammon's mapping as well as in CCA [15].

Isomap also turns out to be a nonlinear generalization of classical metric MDS that keeps using a spectral decomposition in its optimization process. Very few other methods have succeeded in owning this advantage. Laplacian eigenmaps [16], for instance, tries to unfold and project data by minimizing small distances only. Formally, Laplacian eigenmaps uses a spectral decomposition to solve $\min_{\mathbf{X}} \sum_{i<j} w_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|_2^2$, subject to $\mathbf{1}^T\mathbf{X} = \mathbf{0}$ and $\mathbf{C}_{\mathbf{XX}} = \mathbf{I}$, where $w_{ij} > 0$ if and only if $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_j$ are neighbors. (K-ary neighborhoods or ϵ-balls can be used such as in Isomap.) While the connection between Laplacian eigenmaps and distance preservation might seem unclear, several authors have shown that it actually amounts to applying classical metric MDS to commute-time distances [17], that is, to distances related to random walks in a graph. The connection with distance preservation is perhaps more straightforward in maximum variance unfolding [18]. The idea behind this spectral method is somehow dual to that of Laplacian eigenmaps: MVU seeks to unfold and project data by preserving the distances between neighboring points and maximizing all other ones. Formally, it solves $\max_{\mathbf{X}} \sum_{i<j} \|\mathbf{x}_i - \mathbf{x}_j\|_2^2$, subject to $\mathbf{1}^T\mathbf{X} = \mathbf{0}$ and $\|\mathbf{x}_i - \mathbf{x}_j\|_2 = \delta_{ij}$ if $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_j$ are neighbors. In practice, it amounts to modifying a Gram matrix by means of semidefinite programming before applying classical metric MDS on it.

Since a few years, the interest in distance preservation is slowly evolving toward similarity preservation. Whereas a pairwise dissimilarity typically grows with its corresponding distance, a similarity is usually defined to be a decreasing

function of the distance. In the context of dimensionality reduction, the use of similarities is increasingly perceived as more consistent with the intuition that local properties such as K-ary neighborhoods should be preserved prior to global properties. This idea underlies all weighting schemes that are used in MDS, Sammon's mapping, CCA, and their variants. By using similarities, the dominating terms in a cost function are naturally associated with small distances. For instance, let us define normalized pairwise similarites with $\pi_{ij} = \gamma(\delta_{ij}^2)/\sum_{k<l}\gamma(\delta_{kl}^2)$ and $p_{ij} = g(d_{ij}^2)/\sum_{k<l} g(d_{kl}^2)$, where γ and g are positive and decreasing functions of their arguments. Following the idea of stochastic neighbor embedding [19], the Kullback-Leibler divergence written as $D(\mathbf{X};\mathbf{\Xi}) = \sum_{i<j}\pi_{ij}\log(\pi_{ij}/p_{ij})$ can be minimized by gradient descent. The formula of the partial derivative w.r.t. the low-dimensional coordinates turns out to be surprisingly concise and elegant:

$$\frac{\partial D(\mathbf{X};\mathbf{\Xi})}{\partial \mathbf{x}_i} = \sum_j (\pi_{ij} - p_{ij})\frac{g'(d_{ij}^2)}{g(d_{ij}^2)}(\mathbf{x}_i - \mathbf{x}_j) \ .$$

It also shows that the gradient is negligible for large distances, that is, for small similarities, provided $k'(d_{ij}) \leq k(d_{ij})$. Recent papers investigates the choice of the similarity functions [20] and the definition of the cost function [21]. As the KL divergence is not symmetric, the authors of [21] consider a weighted combination of two divergences, based on the same principle as their distance preserving method in [11, 12]. In particular, this allows them to cast their method within the framework of statistical information retrieval.

3 Learning topology

Applying geometrical and topological methods in order to analyze high-dimensional data has attracted recent scientific attention in the machine learning community, e.g. [22, 23, 24]. Starting from a finite set of points in a high-dimensional space, several approaches intend to learn, explore and exploit the topology of manifolds, from which these points are supposed to be drawn, or shapes, i.e. topological invariants, such as the intrinsic dimension. There is a wide scope of applications using such topology-based methods ranging from exploratory data analysis [25], pattern recognition [26], process control [27], semi-supervised learning [28, 29], to manifold learning [30, 29] and clustering [31].

In structure-preserving dimensionality reduction, nonlinear embedding techniques are used to represent high-dimensional data or as preprocessing step for supervised or unsupervised learning tasks, e.g. [22, 32]. However, the final dimension of the projected data and the topological properties of the target space are constrained a priori. In spectral methods, it is intended to perform manifold regularization by taking into account the topology of the shapes using the Laplacian of some proximity graph of the data [33, 28]. A similar approach is also used in spectral clustering [34, 35, 36, 37]. Here, choosing an appropriate proximity graph is essential and greatly impacts the results, making these methods sensitive to noise [38] or outliers. Unfortunately, there is no universal objective criterion of how to estimate the quality of such a data-induced graph.

If processing in geometric low-dimensional spaces is addressed, so-called

'computational geometry' approaches can be applied. Relevant concepts range from epsilon-samples [39] and restricted Delaunay triangulations [40] to various concepts for estimating topological and geometrical properties of shapes [39, 41]. Again, to properly reconstruct given real-world data sets, assumptions on the unknown shape as being represented by a smooth manifold have to be made, which frequently will not be adequate in the presence of noise.

In the last few years, various approaches have stimulated the field of topology learning, based on geometric and algebraic ideas. The concept of distance functions, e.g. [42], allows for a re-interpretation of geometric inference [43]. The so-called 'topological persistence' [44] has been applied to noise reduction [45] and to improved visualization methods for 3D image data sets [31]. Manifold reconstruction in high-dimension [46] and the combination of statistical and topological approaches should be mentioned here, extending Voronoï concepts to Bregman divergence [47], or defining generative models based on simplicial complexes [48]. These approaches aim at combining ideas of generative principal manifolds [30] and witness complexes [25].

The most powerful neural network topology learning method is the Self-Organizing Map (SOM) which provides a robust method to visualize essential properties of data [49]. Under certain conditions, it represents a topographic mapping of high-dimensional input data onto a low-dimensional space usually sampled by a regular grid. Here, topographic mapping means the preservation of the continuity of the mapping between the two spaces [50]. After network training, this property can be assessed quantitatively, see e.g. [51]. Various extensions of the basic SOM have been described in the literature, such as magnification control schemes [52], or other modifications related to learning using auxiliary data [53], probability density estimation [54], or kernel methods [55], nonlinear embedding [56], and pattern matching [57, 58]. For a review on the SOM literature, we refer to Kohonen's textbook [59].

Recently, a novel computational approach to topology learning has been proposed that systematically reverses the data-processing workflow in topology-preserving mappings: the Exploration Machine (Exploratory Observation Machine, XOM) [60, 24, 61, 62]. By systematically exchanging functional and structural components of topology-preserving mappings, XOM can be seen as a computational framework for both structure-preserving dimensionality reduction and data clustering [63, 64]. This approach provides conceptual and computational advantages when compared to SOM and other dimensionaliy reduction methods [65], which has been demonstrated by computer simulations and real-world applications, such as in functional MRI and gene expression analysis [65, 61]. Specific advantages refer to (i) concise visualization and resolution of underlying data cluster structures, (ii) substantially reduced computational expense, and (iii) direct applicability to the analysis of non-metric data.

As pointed out in [65], XOM represents the general concept of inverting topology-preserving mappings as a fundamental pattern recognition approach, thus implying novel methods for data clustering, semi-supervised learning [66], analysis of non-metric data, pattern matching, and incremental optimization [60]. Moreover, current research [67] unveils that XOM provides interesting conceptual cross-links between fast sequential online learning known from topology-

preserving mappings (as in SOM) and principled direct optimization of divergence measures (e.g. Kullback-Leibler divergence) which compare neighborhood statistics in data and target spaces, such as in Stochastic Neighbor Embedding (SNE) [68] and its variants.

4 Quality assessment

The variety of methods presented in the previous sections raises the question of quality assessment. Relevant criteria are needed in order to compare methods and evaluate the reliability of their results. For a long time, quality criteria have been closely related to the cost functions of some dimensionality reduction techniques. For instance, PCA variance fraction or stress functions [5, 6, 7] have been very popular. Since the eighties, the SOM community has developed specific criteria based on topological considerations. Trustworthiness and continuity [9] are such criteria based on rank preservation between data and their k-nearest neighbors in original and projection space.

The previous criteria are given as a single or a pair of numbers. While this may be a sufficient summary to compare several mappings and select the best one, this is not enough considering that mappings are to be used as visual support decision tools which must be interpreted with the eyes. We stress that nonlinear maps which display multidimensional data as cloud of points, cannot be trusted as such, because axes have no meaning so we cannot tell about the correlation of some original variables, and distances are not well preserved in general so we cannot tell about the authenticity of the cluster structure we observe.

Several authors [69, 10, 70, 9, 71] provided a taxonomy of the distorsions which might occur. According to the one defined in [69]: compression and stretching of the distances alter the geometry, while tears (nearby data mapped far appart) and false neighborhoods (far appart data mapped as neighbors) alter the topology of the underlying data structure. A statistical interpretation of these different types of errors is given in [21]; it allows the authors to define quality criteria that are closely related to quantities such as precision and recall, which are standard tools in classification and information retrieval.

Not only the quantification of mapping errors is of interest: the location of the errors in the low-dimensional representation proves to be important as well. We must indeed know which part of the display we can trust before willing to infer any property of the original multidimensional data structure. In the sequel, SOM is simply considered as performing a non linear mapping of the neurons instead of the data, so visualization initially dedicated to nonlinear mappings apply to SOM too.

The Shepard diagram can be used as an auxiliary graphic which displays a cloud of $N(N-1)/2$ points having the original and mapped pairwise distances as x and y components respectively. The cloud lays close to the diagonal $y = x$ if no or few distortions occur, above it for a majority of stretching and tears, below for a majority of compressions and false neighborhoods. However, this scatter plot is not visually correlated straight to the map making it difficult to know where exactly in the map the distortions occur.

The problem is that the map shows N points while there are about N^2 distor-

tions to display (N^2 pairwise distances) so one way is to display some statistic about them. Aupetit proposed to visualize local amount of compression and stretching, coloring Voronoï cells of edges in the Delaunay graph of the mapped data[69], which is somehow similar to the U-matrix representation used with SOM [72] where color shows the amount of empty space between the neurons in the data space. However, both these approaches cannot show tears, making hazardous to draw any conclusion about the data cluster structure (A single cluster can be shared in very different parts of the map as Aupetit shows in [69]). Kaski et al. [73] proposed to color SOM neurons based on their similarity in the data space. The SOM is projected both in the data space and in an auxiliary perceptually uniform 2-dimensional color space which visually encodes the similarity. However, the unfolding in the color space is prone to distortions itself and a 2-dimensional color space cannot account for all the topological states the data structure may have. In this special session, Lespinats and Aupetit propose to visualize the average stretching or compression measured at each point through standard trustworthiness and continuity criteria devised by Venna and Kaski [9], by coloring accordingly the Voronoï cell of these points. Thus, showing both kinds of distortions makes visual inference possible in areas free of any of them.

Another way to deal with mappings prone to distortions, is not to show distortions themselves, but to show some measure of the original data co-located within the map. This is a kind of spatial correlation where the topological structures of the original and projection spaces are displayed on top of each other to allow for visual comparison.

Rousset et al. [74] are the first to implement this idea with a SOM by replacing each neuron with a small clone of the map itself which displays as a color, the original distances between the neuron and each neuron of the small map. Neurons which look similar are close in the data space. However the approach is limited to small maps. Pölzlbauer et al. [75] proposed to visualize a SOM with a graph structure on top of it whose edges connect two neurons if some of their data are neighbors based on a proximity criterion (k-nearest or ϵ-ball neighborhoods). In this case, any kind of topological structure can be represented but the method is prone to the hairball effect : many links crossing each other through the whole map can hide distortion-free areas. In a similar way, a recent paper by Tasdemir and Merenyi [76] shows the Induced Delaunay Triangulation (IDT) [77] of the neurons built in the data space. Two neurons are connected by an edge of the IDT if they are first and second best matching units of some data points called the witnesses of this edge [25]. The edges of the graph are weighted with respect to the number of witnesses they have, and colored accordingly. However, the IDT is known to be prone to topological artefacts [78] so may not show some topological distortion of the SOM. Aupetit [69] proposed the proximity measure for nonlinear projection methods, which considers a reference point, and displays its original distance to the other points as a color of their Voronoï cells. This is similar to only displaying the neighborhood graph of one neuron in the Pölzlbauer approach. Therefore the proximity measure cannot show at once all the original topology, but that one can be discovered step by step selecting reference points throughout the map. In these four methods the original similarity is visualized (up to some quantization in SOM), so even with

many mapping distortions, it is still possible to recover the original topology of the data or neuron structure in the data space.

The main conclusion to draw from these last works its that mappings are not an end, but only a means to display useable and useful information on top of them. This is a usual way of thinking for SOM practitionners because the location of the neurons on the map is not sufficient to show cluster structures, for instance. But in any cases, practitionners should be aware of distorsions, because eyes are prone to see patterns even in random clouds of points, so we advise them not to use maps without being confident about what the map shows. Displaying on the map the distortions at first, and the original similarities at best are two ways to strengthen the relevance of their conclusions.

References

[1] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572, 1901.
[2] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.
[3] I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, NY, 1986.
[4] G. Young and A.S. Householder. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3:19–22, 1938.
[5] T.F. Cox and M.A.A. Cox. *Multidimensional Scaling*. Chapman & Hall, London, 1995.
[6] I. Borg and P. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer-Verlag, New York, 1997.
[7] J.W. Sammon. A nonlinear mapping algorithm for data structure analysis. *IEEE Transactions on Computers*, CC-18(5):401–409, 1969.
[8] P. Demartines and J. Hérault. Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. *IEEE Transactions on Neural Networks*, 8(1):148–154, January 1997.
[9] J. Venna and S. Kaski. Neighborhood preservation in nonlinear projection methods: An experimental study. In *ICANN*, pages 485–491, 2001.
[10] J.A. Lee and M. Verleysen. Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomputing*, 72(7-9):1431–1443, 2009.
[11] J. Venna and S. Kaski. Visualizing gene interaction graphs with local multidimensional scaling. In M. Verleysen, editor, *Proc. ESANN 2006, 14th European Symposium on Artificial Neural Networks*, pages 557–562. d-side, Bruges, Belgium, April 2006.
[12] J. Venna and S. Kaski. Local multidimensional scaling. *Neural Networks*, 19:889–899, 2006.
[13] J.B. Tenenbaum, V. de Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000.
[14] M. Bernstein, V. de Silva, J.C. Langford, and J.B. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, Stanford University, Palo Alto, CA, December 2000.
[15] J.A. Lee and M. Verleysen. Curvilinear distance analysis versus isomap. *Neurocomputing*, 57:49–76, March 2004.
[16] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, June 2003.
[17] L. Yen, D. Vanvyve, F. Wouters, F. Fouss, M. Verleysen, and M. Saerens. Clustering using a random-walk based distance measure. In M. Verleysen, editor, *Proc. ESANN 2005, 13th European Symposium on Artificial Neural Networks*, pages 317–324, Bruges, Belgium, April 2005. d-side.
[18] K.Q. Weinberger and L.K. Saul. Unsupervised learning of image manifolds by semidefinite programming. *International Journal of Computer Vision*, 70(1):77–90, 2006.
[19] G. Hinton and S.T. Roweis. Stochastic neighbor embedding. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems (NIPS 2002)*, volume 15, pages 833–840. MIT Press, 2003.
[20] L. van der Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.
[21] J. Venna, J. Peltonen, K. Nybo, H. Aidos, and S. Kaski. Information retrieval perspective to nonlinear dimensionality reduction for data visualization. *Journal of Machine Learning Research: Workshop and Conference Proceedings*, 11:(to appear), 2010.
[22] J.A. Lee and M. Verleysen. *Nonlinear Dimensionality Reduction*. Springer, New York, 2007.

[23] G. Haro, G. Randall, and G. Sapiro. Stratification learning: Detecting mixed density and dimensionality in high dimensional point clouds. In *Proc. NIPS*, 2007.
[24] A. Wismüller. The exploration machine - a novel method for data visualization. In *Lecture Notes in Computer Science. Advances in Self-Organizing Maps*, pages 344–352, 2009.
[25] V. de Silva and G. Carlsson. Topological estimation using witness complexes. In *Eurographics Symposium on Point-Based Graphics*, ETH, Zürich, Switzerland, June 2-4, 2004.
[26] A. Collins, A. Zomorodian, G. Carlsson, and L. Guibas. A barcode shape descriptor for curve point cloud data. In *Eurographics Symposium on Point-Based Graphics*, ETH, Zürich, Switzerland, June 2-4, 2004.
[27] M. Zeller, R. Sharma, and K. Schulten. Topology representing network for sensor-based robot motion planning. In *Proc. WCNN*. INNS Press.
[28] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7:2399–2434, 2006.
[29] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-Supervised Learning*. MIT Press, Cambridge, MA, 2006.
[30] R. Tibshirani. Principal curves revisited. *Statistics and Computing*, 2:183–190, 1992.
[31] D. Laney, P.T. Bremer, A. Mascarenhas, P. Miller, and V. Pascucci. Understanding the structure of the turbulent mixing layer in hydrodynamic instabilities. *IEEE Trans. on Vis. and Comp. Graph.*, 12(5):1053–1060, 2006.
[32] M. Vlachos, C. Domeniconi, D. Gunopulos, G. Kollios, and N. Koudas. Nonlinear dimensionality reduction techniques for classification and visualization. In *Proc. KDD*, 2002.
[33] A. Argyriou, M. Herbster, and M. Pontil. Combining graph laplacians for semi-supervised learning. In *Proc. NIPS*, 2005.
[34] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Proc. NIPS*, 2002.
[35] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.
[36] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
[37] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: analysis and an anlgorithm. In *Proc. NIPS*, 2002.
[38] M. Hein and M. Maier. Manifold denoising. *Advances in Neural Information Processing Systems*, 19, 2007.
[39] N. Amenta and M. Bern. Surface reconstruction by Voronoi filtering. *Discrete and Computational Geometry*, 22(4):481–504, 1999.
[40] H. Edelsbrunner and N.R. Shah. Triangulating topological spaces. *Int. Journal on Computational Geometry and Applications*, 7:365–378, 1997.
[41] H. Edelsbrunner, J. Harer, V. Natarajan, and V. Pascucci. Morse-smale complexes for piecewise linear 3-manifolds. In *Eurographics Symposium on Point-Based Graphics*, Proc. 19th Ann. Symp. Comp. Geom. (SOCG), 2003.
[42] F. Chazal, D. Cohen-Steiner, and A. Lieutier. A sampling theory for compact sets in euclidean spaces. *Discrete and Computational Geometry*, 41(3):461–479, 2007.
[43] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete and Computational Geometry*, 39(1-3):419–441, 2006.
[44] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. In *IEEE Symp. on Found. of Comp. Sci.*, pages 454–463, 2000.
[45] F. Chazal and A. Lieutier. Topology guaranteeing manifold reconstruction using distance function to noisy data. In *Proc. Symp. on Computational Geometry*, 2006.
[46] J.D. Boissonnat, L. Guibas, and S. Oudot. Manifold reconstruction in arbitrary dimensions using witness complexes. In *Proc. Symp. on Computational Geometry*, pages 194–203, 2007.
[47] J.D. Boissonnat, F. Nielsen, and R. Nock. On Bregman Voronoi diagrams. In *Proc. 18th ACM-SIAM Symp. on Discrete Algorithms*, 2007.
[48] M. Aupetit. Learning topology with the generative gaussian graph and the em algorithm. In *Proc. of NIPS*, 2005.
[49] T. Kohonen. The self-organizing map. *Proc. IEEE*, 78(9):1464–1480, 1990.
[50] T. Villmann, R. Der, M. Herrmann, and T. Martinetz. Topology preservation in self-organizing feature maps: Exact definition and measurement. *IEEE Transactions on Neural Networks*, 8(2):256–266, 1997.
[51] H.-U. Bauer and K.R. Pawelzik. Quantifying the neighborhood preservation of Self-Organizing Feature Maps. *IEEE Trans. on Neural Networks*, 3(4):570–579, 1992.
[52] H. U. Bauer, R. Der, and M. Herrmann. Controlling the magnification factor of self-organizing feature maps. *Neural Computation*, 8(4):757–71, 1996.
[53] J. Sinkkonen and S. Kaski. Clustering based on conditional distributions in an auxiliary space. *Neural Computation*, 14:217–239, 2002.
[54] J. Lampinen and T. Kostiainen. Generative probability density model in the self-organizing map. In U. Seiffert and L.C. Jain, editors, *Self-Organizing Neural Networks*, Studies in Fuzzi-

ness and Soft Computing, pages 75–92. Physica-Verlag, Heidelberg, New York, 2001.
[55] M. Van Hulle. *Faithful Representations and Topographic Maps*. Wiley Series and Adaptive Learning Systems for Signal Processing, Communications, and Control. Wiley & Sons, New York, 2000.
[56] J.A. Lee, C. Archambeau, and M. Verleysen. Locally linear embedding versus Isotop. In *Proc. of the 11th Europ. Symp. on Art. Neur. Netw. (ESANN)*, pages 527–534, Bruges, Belgium, 2003. d-side publishers.
[57] A. Wismüller, F. Vietze, D.R. Dersch, J. Behrends, K. Hahn, and H. Ritter. The deformable feature map – a novel neurocomputing algorithm for adaptive plasticity in pattern analysis. *Neurocomputing*, 48:107–139, 2002.
[58] A. Wismüller, F. Vietze, J. Behrends, A. Meyer-Baese, M.F. Reiser, and H. Ritter. Fully-automated biomedical image segmentation by self-organized model adaptation. *Neural Networks*, 17:1327–1344, 2004.
[59] T. Kohonen. *Self-Organizing Maps*. Springer, 3rd edition, 2001.
[60] A. Wismüller. Exploratory Morphogenesis (XOM): A Novel Computational Framework for Self-Organization. Ph.D. thesis, Technical University of Munich, Department of Electrical and Computer Engineering, 2006.
[61] A. Wismüller. A computational framework for nonlinear dimensionality reduction and clustering. In *Lecture Notes in Computer Science. Advances in Self-Organizing Maps*, pages 334–343, 2009.
[62] A. Wismüller. Exploration-organized morphogenesis (XOM) - a general framework for learning by self-organization. In *Human and Machine Perception. Research Reports of the Institute for Phonetics and Speech Communication (FIPKM)*, volume 37, pages 205–239. University of Munich, 2001.
[63] A. Wismüller. A computational framework for exploratory data analysis. In M. Verleysen, editor, *European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning*. d-side Publishers, 2009.
[64] A. Wismüller. The exploration machine - a novel method for structure-preserving dimensionality reduction. In M. Verleysen, editor, *European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning*. d-side Publishers, 2009.
[65] A. Wismüller. The exploration machine: a novel method for analyzing high-dimensional data in computer-aided diagnosis. In N. Karssemeijer and M. Giger, editors, *Medical Imaging 2009: Computer-Aided Diagnosis. Proc. SPIE*, volume 7260, pages 72600G–72600G–7, 2009.
[66] K. Bunte, B. Hammer, A. Wismüller, and M. Biehl. Adaptive local dissimilarity measures for discriminative dimension reduction of labeled data. *Neurocomputing*, In Press:–, 2010.
[67] K. Bunte, B. Hammer, T. Villmann, M. Biehl, and A. Wismüller. Exploratory Observation Machine (XOM) with Kullback-Leibler divergence for dimensionality reduction and visualization. In *Proc. of the 18th Europ. Symp. on Art. Neur. Netw. (ESANN)*, Bruges, Belgium, 2010. d-side publishers.
[68] G. Hinton and S. Roweis. Stochastic neighbor embedding. In *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2003.
[69] M. Aupetit. Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing*, 70(7-9):1304–1330, 2007.
[70] S. Lespinats, M. Verleysen, A. Giron, and B. Fertil. DD-HDS: A method for visualization and exploration of high-dimensional data. *IEEE Transactions on Neural Networks*, 18(5):1265–1279, 2007.
[71] J. Vesanto. SOM-based data visualization methods. *Intelligent Data Analysis*, 3(2):111–126, 1999.
[72] H.P. Siemon and A. Ultsch. Kohonen networks on transputers: Implementation and animation. In *International Neural Networks*, pages 643 – 646. Kluwer Academic Press, Paris, 1990.
[73] S. Kaski, J. Venna, and T. Kohonen. Coloring that reveals cluster structures in multivariate data. *Australian Journal of Intelligent Information Processing ,Systems*, 6:82–88, 2000.
[74] P. Rousset and C. Guinot. Distance between kohonen classes, visualization tool to use som in data set analysis and representation. In J. Mira and A. Prieto, editors, *Proc. Intl. Workshop on Artificial Neural Networks (IWANN'01), LNCS 2085*, pages 119–126. Springer-Verlag, 2001.
[75] G. Pölzlbauer, A. Rauber, and M. Dittenbach. Graph projection techniques for self-organizing maps. In Michel Verleysen, editor, *Proc. European Symposium on Artificial Neural Networks (ESANN'05)*, pages 533–538, Bruges, Belgium, April 27-29 2005. d-side publications.
[76] K. Tasdemir and E. Merényi. Exploiting data topology in visualization and clustering of self-organizing maps. *IEEE Trans. on Neural Networks*, 20(4):549–562, 2009.
[77] Th. Martinetz and K. Schulten. Topology representing networks. *Neural Networks*, 7(3):507–522, 1994.
[78] P. Gaillard, M. Aupetit, and G. Govaert. Learning topology of a labeled data set with the supervised generative gaussian graph. *Neurocomputing*, 71(7-9):1283–1299, 2008.