Chapter 10

Conclusion

Dimensionality reduction is an important preprocessing step in machine learning applications. It involves searching for a more compact representation for the data than their original forms. Most current methods focus on vectorial data with the assumption that the data are residing in a normed vector space. However, there are significant amount of real world data that are non-vectorial, with no convenient vector representations. This kind of data are often more complicated and pose great challenges to the subsequent data processing algorithms. Often, one embeds these data into a low dimensional vector space (usually Euclidean) to facilitate subsequent processing that include classification, visualization, etc. This procedure resembles the task of DR for normal vectorial data. As such, we refer to this as the DR for non-vectorial data in this thesis.

This thesis addresses the DR problem for non-vectorial data by resorting to measurements matching which is a powerful approach adopted by many DR algorithms. To avoid the excessive loss of information due to the vectorization procedure, we take advantage of the fact that a kernel can be a suitable similarity estimator for non-vectorial data. In fact, there are a considerable number of kernels for data from different fields such as bioinformatics, image processing, text categorization, data mining, etc. Apparently, kernels can be utilized to describe the similarities among the input data. The questions remained as to how to represent the relationships among the embeddings in the latent space, and how to implement match the similarity measures, one for the input and another for the embeddings.

After proposed the SCIGV kernel, a specially designed kernel for shapes and images, we presented the Kernel Laplacian Eigenmaps where in latent space the Euclidean distance is used as the measurement. KLE shares similar structure to that of Laplacian Eigenmaps in that the matching is done by minimizing the trace of the product of two matrices and the solution is obtained by referring to graph Laplacian and spectral decomposition. The kernel Gram matrix in KLE in the input space replaces the proximity matrix in LE whence non-vectorial data can be handled by this algorithm.

However, KLE has its limitation. This comes from the choice of the measure for the embeddings. Kernels is highly nonlinear, but it has been simply matched to the Euclidean distance which is not ideal in capturing the nonlinearity. This idea inspired the Twin Kernel Embedding (TKE) whose relational descriptor for embeddings is captured by another kernel. This kernel is a homogeneous kernel in order to bridge the relation between distance and kernel. Typically, it is an RBF kernel. The virtue of the homogeneous kernel is that it is constructed on distance (normally Euclidean) however expressed as a nonlinear kernel. It is this dual property that ensures both the geometrical interpretation and the desired nonlinearity. Matching the kernel on the input data with the kernel on the embeddings preserves the symmetry on the structural level from which nonlinearity in the input can be more faithfully maintained. This choice increases the complexity of the algorithm. The matching was still accomplished by trace of the product of two matrices in this case they are kernel Gram matrices, and the regularization terms on the RBF kernel and the embeddings were involved to make the objective function well defined. Due to high nonlinearity and non-convexity in the results, the gradient descent based searching technique is applied in the optimization procedure.

Behind the design of these two algorithms lies the general design method for DR algorithms that is based upon the measurement matching framework. The measures for both data and their corresponding embeddings and the matching functional are the three major parts of the DR which can be devised separately to obtain different algorithms. The versatility of this framework can be witnessed from quite a few of existing DR algorithms that can be interpreted by it. These algorithms include both deterministic and probabilistic models. They can handle either vectorial or non-vectorial data or both. Some of them originate from manifold estimation, but others from

graph theory. Although they look different, most end up with the common structure that falls into the measurement matching framework.

We can further enrich the content of the framework by introducing other elements that are common across DR methods including the backward/forward mapping functions between data and embeddings, (semi)supervised setting, etc. BCTKE and RCTKE are practical examples of applying backward mapping to enhance the host DR algorithms. In order not to compromise the non-vectorial applicability, we presented the mapping function derived from kernel machine and LS-SVM core part and incorporated them into the original TKE by substitution (BCTKE) or via the regularization terms (RCTKE). The backward mapping function constructs a relation between input data and their embeddings, thus solving the out-of-sample problem of TKE, enabling this algorithm to be applied to a much wider range of applications such as classification, information retrieval, patter recognition, etc. Intuitively, the mapping functions can be embedded in other host algorithms to enable them to deal with novel samples. The only requirement is that the optimization problem is still solvable after the reconstruction. As far as (semi)supervised learning is concerned, we also present solution on TKE. The revised TKE objective function reflected the consideration that samples from the same class or cluster should be close and far away otherwise. This can be further adopted in BCTKE and RCTKE accordingly and the learnt mapping function would consequently be more meaningful in favor of better separability. Interestingly, the method used in supervised TKE generated new method like [114] for vectorial data which is is confirmed by experiments to be more robust and powerful than LDA.

Besides presenting the design of DR algorithms and the TKE family (including BCTKE, RCTKE, we simply call them TKE algorithms), we also studied several important aspects of algorithms such as computational complexity, convergence, model selection, initialization (required by gradient based optimization) mainly empirically by experimental results.

TKE algorithms cost more than other algorithms solved by eigen-decompositions from the nature of the non-convexity and iterative searching. However, as shown in the experiments, the convergence is very fast at the beginning and level off gradually afterwards. As a result, we could stop the optimization in several hundreds of iterations or less. In each iteration, the cost of the TKE algorithms is comparable to those of spectral methods. Similar to these other algorithms,

the complexity of TKE algorithms grows with the number of the data obtained as a result of the pairwise similarity matching. It becomes extremely high when N approaching the order of thousands. In some applications such as web mining, millions of pages will prohibit these algorithms. It is necessary to tackle this problem in future research.

It has been shown in this thesis that the TKE algorithms are quite insensitive to the selection of the regularization parameters. However, the size of the neighborhood, i.e. n in nearest neighbor filtering will influence the result greatly. In practice, n is better to be less than the number of samples in each class. Non-parametric version of the TKE algorithms would be beneficial by freeing them from the need for choosing n. This will be investigated in near future.

Currently TKE algorithms are initialized by KPCA, KLE if the targets are non-vectorial objects. We have studied the cases which started from random initialization. It indicated that the uniformly distributed initial embeddings can be possibly adjusted to the state that is similar to the results initialized by KPCA or KLE. This behavior is worth further study since it may suggest the power of TKE algorithms in preserving the relational structure of the data and be helpful in understanding the distribution of the optimal solutions.

For constraints in DR, we have seen some typical mapping functions: linear transformation, RBF neural network, MLP and kernel machine. The first one is widely used in linear methods because it is simple and fast. The last one corresponds to kernel subspace projection designed for non-vectorial data. A natural question is which is better, or which one truly reveals the relation between original data and their embeddings. Even the existence of such relation is doubtful. It is reasonable to assume that the data are parameterized by a few degree of freedoms in a much lower dimensional space, or in other words, there is a function mapping low dimensional coordinates to high dimensional space. This is the basic assumption of the DR methods. But it is not clear whether the corresponding inverse function exists or even if it is there, whether the listed mapping functions before are good approximations. We want to know the form of the inverse functions and obtain all the coefficients that is to recover the function explicitly. The understanding of this problem touches the core of the DR itself. This can only be fulfilled based on the accumulating previous research and exploring new knowledge.

With regard to the research of the DR algorithms, we cannot deny there are still many other

design methods and tools can be borrowed such as neural networks, regression, etc. We can look at DR problem from different points of view. For example, we can focus on the classification performance and consider the combination of DR with classifiers which leads to hybrids. We can consider the hierarchical structure by stacking basic DR algorithms to reduce the complexity or to realize different treatment to different division of the data. We are also be able to take the mixture model since the data may lie on several separate manifold instead of one. Other than these, the evaluation of the performance of different DR algorithms is of significance in DR research. The existing evaluation methods can only provide limited assessment in several aspects, such as the quality of the clusters, classification performance, etc. A universal overall evaluation standard is required. Finally, we also hope to extend the application of TKE algorithms to real world applications and develop new algorithms which is closer to the ground truth of the intrinsic dimensionality and that can be computed efficiently.

Bibliography

[1] Dimitris K. Agrafiotis. A new method for analyzing protein sequence relationships based on sammon maps. *Protein Science*, 6(2):287–293, 1997.

[2] Dimitris K. Agrafiotis. Stochastic proximity embedding. *Journal of Computational Chemistry*, 24(10):1211–1217, 2002.

[3] Izydor Apostol and Wojciech Szpankowski. Indexing and mapping of proteins using a modified nonlinear sammon projection. *Journal of Computational Chemistry*, 20(10):1049–1059, 1999.

[4] Tom M. Apostol. *Mathematical Analysis*. Addison-Wesley, second edition, 1974.

[5] Mukund Balasubramanian and Eric L. Schwartz. The isomap algorithm and topological stability. *Science*, 295(4):7, 2002.

[6] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[7] Mikhail Belkin and Partha Niyogi. Semi-supervised learning on riemannian manifolds. *Machine Learning*, 56:209–239, 2004. Invited, Special Issue on Clustering.

[8] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(24):509–522, April 2002.

[9] Yoshua Bengio, Olivier Delalleau, Nicolas Le Roux, Jean-Francçois Paiement, Pascal Vincent, and Marie Ouimet. Learning eigenfunctions links spectral embedding and kernel PCA. *Neural Computation*, 16:2197–2219, 2004.

[10] Yoshua Bengio, Olivier Delalleau, Nicolas Le Roux, Jean-Francois Paiement, Pascal Vincent, and Marie Ouimet. Spectral dimensionality reduction. *Studies in Fuzziness and Soft Computing*, 207:519–550, 2006.

[11] Yoshua Bengio, Jean-François Paiement, Pascal Vincent, Olivier Delalleau, Nicolas Le Roux, and Marie Ouimet. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. In *Advances in Neural Information Processing Systems 16*.

[12] Ella Bingham and Heikki Mannila. Random projection in dimensionality reduction: applications to image and text data. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 245–250, 2001.

[13] Christopher M. Bishop. Variational principal components. In *Ninth International Conference on Artificial Neural Networks*, volume 1, pages 509–514, 1999.

[14] Christopher M. Bishop, Markus Svensén, and Christopher K. I.Williams. Gtm: The generative topographic mapping. *Neural Computation*, 10:215–234, April 1998.

[15] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.

[16] Michael Bowling, Ali Ghodsi, and Dana Wilkinson. Action respecting embedding. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 65–72, New York, NY, USA, 2005. ACM Press.

[17] Matthew Brand. Charting a manifold. In *Advances in Neural Information Processing Systems 15*. MIT Press.

[18] Samarasena Buchala, Neil Davey, Ray J. Frank, and Tim M. Gale. Dimensionality reduction of face images for gender classification. In *Proceedings of 2nd International IEEE Conference on Intelligent Systems*, volume 1, pages 88–93, June 2004.

[19] Hong Chang and Dit-YanYeung. Robust locally linear embedding. *Pattern Recognition*, 39:1053–1065, 2006.

[20] Yizong Cheng and George M. Church. Biclustering of expression data. In *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology (ISMB)*, pages 93–103, 2000.

[21] Heeyoul Choi and Seungjin Choi. Robust kernel isomap. *Pattern Recognition*, 40(3):853–862, 2007.

[22] Fan R. K. Chung. *Spectral Graph Theory*. Number 92 in Regional Conference Series in Mathmatics. AMS, 1997.

[23] D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions. *Journal of symbolic computation*, 9(3):251–280, 1990.

[24] Renato Fernandes Corrêa and Teresa Bernarda Ludermir. Improving self-organization of document collections by semantic mapping. *Neurocomputing*, 70:60–69, August 2006.

[25] Trevor F. Cox and Michael A. A. Cox. *Multidimensional Scaling*. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, 2 edition, 2001.

[26] Marco Cuturi, Kenji Fukumizu, and Jean-Philippe Vert. Semigroup kernels on measures. *Journal of Machine Learning Research*, 6:1169–1198, 7 2005.

[27] Marco Cuturi and Jean-Philippe Vert. A mutual information kernel for sequences. *Proceedings of the International Joint Conference on Neural Networks (IJCNN'04)*, pages 1905–1910, 2004.

[28] Payel Das, Mark Moll, Hernán Stamati, Lydia E. Kavraki, and Cecilia Clementi. Low-dimensional free energy landscapes of protein folding reactions by nonlinear dimensionality reduction. In *Proceedings of the National Academy of Sciences*, volume 103, pages 9885–9890, USA, 2006.

[29] Mark L. Davison. *Multidimensional Scaling*. Wiley series in probability and mathematical statistics, Applied probability and statistics. Wiley, New York, 1983.

[30] Vin de Silva and Joshua B. Tenenbaum. Global versus local methods in nonlinear dimensionality reduction. In S. Thrun S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*. MIT Press.

[31] Vin de Silva and Joshua B. Tenenbaum. Global versus local methods in nonlinear dimensionality reduction. In *Advances in Neural Information Processing Systems 15*, pages 705–712. MIT Press, Cambridge, MA, 2003.

[32] Piotr Dollár, Vincent Rabaud, and Serge Belongie. Non-isometric manifold learning: Analysis and an algorithm. In *Proceedings of the 24th International Conference on Machine Learning*, 2007.

[33] David L. Donoho and Carrie Grimes. Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. In *Proceedings of the National Academy of Arts and Sciences*, number 10, pages 5591–5596, 30 April 2003.

[34] David L. Donoho and Carrie Grimes. Local isomap perfectly recovers the underlying parametrization of occluded/lacunary libraries of articulated images. In *IEEE Computer Vision and Pattern Recognition*, 2003.

[35] David L. Donoho and Carrie Grimes. When does Isomap recover the natural parameterization of families of articulated images. Technical report, Department of Statistics, Stanford University, 2003.

[36] Michael A. Farnum, Huafeng Xu, and Dimitris K. Agrafiotis. Exploring the nonlinear geometry of protein homology. *Protein Science*, 12(1604-1612), 2003.

[37] Ronald A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

[38] Imola K. Fodor. A survey of dimension reduction techniques. Technical report, University of California, May 2002.

[39] Ernest Fokoué. Mixtures of factor analyzers: an extension with covariates. *Journal of Multivariate Analysis*, 95(2):370–384, 2005.

[40] Brendan J. Frey and Delbert Dueck. Clustering by passing messages between data points. *Science*, 315:972–976, 2007.

[41] Baudat G. and Anouar F. Generalized discriminant analysis using a kernel approach. *Neural Computation*, 12(10):2385–2404, 2000.

[42] Thomas Gärtner. A survey of kernels for structured data. *ACM SIGKDD Explorations Newsletter*, 5(1):49–58, 2003.

[43] Thomas Gärtner, John W. Lloyd, and Peter A. Flach. Kernels for structured data. In *Proceedings of the 12th International Conference on Inductive Logic Programming*, 2002.

[44] Zoubin Ghahramani and Geoffrey E. Hinton. The EM algorithm for mixtures of factor analyzers. Technical Report CRG-TR-96-1, 1996.

[45] Amir Globerson, Gal Chechik, Fernando Pereira, and Naftali Tishby. Euclidean embedding of co-occurrence data. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 497–504. MIT Press, Cambridge, MA, 2005.

[46] Yi Guo, Junbin Gao, and Paul W. Kwan. Kernel Laplacian eigenmaps for visualization of non-vectorial data. In *Lecture Notes on Artificial Intelligent*, volume 4304, pages 1179–1183, 2006.

[47] Yi Guo, Junbin Gao, and Paul W. Kwan. Visualization of non-vectorial data using twin kernel embedding. In *First International Workshop on Integrating AI and Data Mining*, volume 0, pages 11–17, 2006.

[48] Yi Guo, Junbin Gao, and Paul W. Kwan. Learning out-of-sample mapping in non-vectorial data dimensionality reduction using constrained twin kernel embedding. In *International Conference on Machine Learning and Cybernetics*, 2007.

[49] Jihun Ham, Daniel D. Lee, Sebastian Mika, and Bernhard Schölkopf. A kernel view of the dimensionality reduction of manifolds. In *Proceedings of the International Conference in Machine Learning*, pages 47–54, 2004.

[50] Jihun Ham, Yuanqing Lin, and Daniel. D. Lee. Learning nonlinear appearance manifolds for robot localization. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2971– 2976, August 2005.

[51] Jens Hanke and Jens G. Reich. Kohonen map as a visualization tool for the analysis of protein sequences: multiple alignments, domains and segments of secondary structures. *Bioinformatics*, 12(6):447–454, 1996.

[52] David Haussler. Convolution kernels on discrete structures. Technical Report UCSC-CRL-99-10, University of California at Santa Cruz, 1999.

[53] Xiaofei He, Deng Cai, Shuicheng Yan, and Hong-Jiang Zhang. Neighborhood preserving embedding. *ICCV*, 2:1208–1213, 2005.

[54] Xiaofei He, O. King, Weiying Ma, Mingjing Li, and Hongjiang Zhang. Learning a semantic space from user's relevance feedback for image retrieval. *IEEE Transactions on Circuits and System for Video Technology*, 13(1):39–48, 2003.

[55] Xiaofei He, Weiying Ma, and Hongjiang Zhang. Learning an image manifold for retrieval. In *ACM conference on Multimedia 2004*, pages 17–23, New York, Oct. 2004.

[56] Xiaofei He and Partha Niyogi. Locality preserving projections. In Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[57] Geoffrey E. Hinton and P. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 28 July 2006.

[58] S. Huang, M. Ward, and E. Rundensteiner. Exploration of dimensionality reduction for text visualization. Technical Report TR-03-14, Worcester Polytechnic Institute, Computer Science Department, 2003.

[59] A. Jain, S. Prabhakar, L. Hong, and S. Pankanti. Filterbank-based fingerprint matching. *IEEE Transactions on Image Processing*, 9(5):846–859, 2000.

[60] Gaurav Jain, Abbasali Ginwala, and Y. Alp Aslandogan. An approach to text classification using dimensionality reduction and combination of classifiers. In *Proceedings of the 2004 IEEE International Conference on Information Reuse and Integration*, pages 564–569, 2004.

[61] Odest Chadwicke Jenkins and Maja J Matarić. A spatio-temporal extension to isomap nonlinear dimension reduction. In *Proceedings of the* 21^{st} *International Conference on Machine Learning*, 2004.

[62] Richard Jensen and Qiang Shen. Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches. *IEEE Transactions on Knowledge and Data Engineering*, 16(12):1457–1471, December 2004.

[63] M. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 1986.

[64] Roy Jonker and A. Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Journal of Computing*, 38(4):325–340, 1987.

[65] Nandakishore Kambhatla and Todd K. Leen. Dimension reduction by local principal component analysis. *Neural Comput.*, 9(7):1493–1516, 1997.

[66] George Karypis and Eui-Hong Han. Concept indexing: A fast dimensionality reduction algorithm with applications to document retrieval and categorization. Technical Report Technical Report TR-00-0016, University of Minnesota, 2000.

[67] Hyunsoo Kim, Peg Howland, and Haesun Park. Dimension reduction in text classification with support vector machines. *Journal of Machine Learning Research*, 6:37–53, 2005.

[68] Philip M. Kim and Bruce Tidor. Subsystem identification through dimensionality reduction of large-scale gene expression data. *Genome Research*, 13:1706–1718, 2003.

[69] Effrosyni Kokiopoulou and Yousef Saad. Orthogonal neighborhood preserving projections. In *IEEE International Conference on Data Mining*, 2005.

[70] Effrosyni Kokiopoulou and Yousef Saad. Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2143–2156, 2007.

[71] Risi Kondor and Tony Jebara. A kernel between sets of vectors. In *The Proceedings of International Conference of Machine Learning, (ICML 2003)*, 2003.

[72] Olga Kouropteva, Oleg Okun, and Matti Pietikäinen. Incremental locally linear embedding. *Pattern Recgonition*, 38:1764–1767, 2005.

[73] J. B. Kruskal. Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrika*, 29:1–27, 1964.

[74] J. B. Kruskal. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29:115–129, 1964.

[75] Paul W. Kwan and Junbin Gao. A multi-step strategy for approximate similarity search in image databases. In *Proceedings of The Seventeenth Australasian Database Conference (ADC)*, pages 139–147, Hobart, TAS, Australia, January 2006.

[76] Paul W. Kwan, Junbin Gao, and Yi Guo. Fingerprint matching using enhanced shape context. In *Proceedings of the Image and Vision Computing New Zealand*, pages 115–120, 2006.

[77] John Lafferty and Guy Lebanon. Information diffusion kernels. In *Advances in Neural Information Processing Systems 14*. MIT Press, Cambridge, MA, 2002.

[78] Stéphane Lafon and Ann B. Lee. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1393–1403, 2006.

[79] Savio L. Y. Lam and Dik Lun Lee. Feature reduction for neural network based text categorization. In *Proceedings of 6th International Conference on Database Systems for Advanced Applications*, pages 195–202, 1999.

[80] Gert R. G. Lanckriet, Gert R. G., Minghua Deng, Nello Cristianini, Michael I. Jordan, and William Stafford Noble. Kernel-based data fusion and its application to protein function prediction in yeast. In *Proceedings of the Pacific Symposium on Biocomputing*, pages 300–311, January 3-8 2004.

[81] Martin H. C. Law, Nan Zhang, and Anil K. Jain. Nonlinear manifold learning for data stream. In *Proceedings of SIAM Data Mining 2004*, pages 33–44, 2004.

[82] Martin H.C. Law and Anil K. Jain. Incremental nonlinear dimensionality reduction by manifold learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):377–391, 2006.

[83] Neil Lawrence. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *Journal of Machine Learning Research*, 6:1783–1816, 2005.

[84] Neil D. Lawrence and Joaquin Quiñonero-Candela. Local distance preservation in the GP-LVM through back constraints. In W. Cohen and A. Moore, editors, *Proceedings of the International Conference in Machine Learning*, San Francisco, CA, 2006. Morgan Kauffman.

[85] Hongyu Li, Li Teng, Wenbin Chen, and I-Fan Shen. Supervised learning on local tangent space. In *Second International Symposium on Neural Networks*, volume 3496, pages 546–551, 2005.

[86] Li Liao and William Stafford Noble. Combining pairwise sequence similarity and support vector machines for remote protein homology detection. In *Proceedings of the sixth International Conference on Computational Molecular Biology*, pages 225–232, April 2002.

[87] Huan Liu and Rudy Setiono. Dimensionality reduction via discretization. *Knowledge-Based Systems*, 9:67–72, 1996.

[88] Xiaoming Liu, Jianwei Yin, Zhilin Feng, and Jinxiang Dong. Incremental manifold learning via tangent space alignment. In *Artificial Neural Networks in Pattern Recognition, Second IAPR Workshop*, pages 107–121, 2006.

[89] Zhiyong Liu and Lei Xu. Topological local principal component analysis. *Neutocomputing*, 55:739–745, 2003.

[90] Huma Lodhi, Craig Saunders, John Shawe-Taylor, Nello Cristianini, and Christopher J. C. H. Watkins. Text classification using string kernels. *Journal of Machine Learning Research*, 2:419–444, 2002.

[91] Huma Lodhi, John Shawe-Taylor, Nello Cristianini, and Christopher J. C. H. Watkins. Text classification using string kernels. In *Advances in Neural Information Processing Systems 13*, pages 563–569, 2000.

[92] Rohit Lotlikar and Ravi Kothari. Bayes-optimality motivated linear and multilayered perceptron-based dimensionality reduction. *IEEE Transactions on Neural Networks*, 11(2):452–463, March 2000.

[93] David Lowe and Michael Tipping. Feed-forward neural networks and topographic mappings for exploratory data analysis. *Neural Computing and Applications*, 4(2):83–95, June 1996.

[94] D. Maio, D. Maltoni, R. Cappelli, J. Wayman, and A. Jain. Fvc2002: Second fingerprint verification competition. In *The 16th International Conference on Pattern Recognition (ICPR'02)*, volume 3, pages 811–814.

[95] Davide Maltoni, Dario Maio, and Anil K. Jain. *Handbook of Fingerprint Recognition.* Springer, 2003.

[96] Peter Meinicke, Stefan Klanke, Roland Memisevic, and Helge Ritter. Principal surfaces from unsupervised kernel regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(9):1379–1391, 2005.

[97] Nathan Mekuz, Christian Bauckhage, and John K. Tsotsos. Face recognition with weighted locally linear embedding. In *Proceedings of the 2nd Canadian Conference on Computer and Robot Vision*, pages 290–296, May 2005.

[98] Nathan Mekuz and John K. Tsotsos. Parameterless isomap with adaptive neighborhood selection. In *Proceedings of 28th DAGM Symposium*, volume 4174, pages 364–373, Berlin, September 2006.

[99] Roland Memisevic and Geoffrey E. Hinton. Improving dimensionality reduction with spectral gradient descent. *Neural Networks*, 18:702–710, 2005.

[100] Martin F. Møller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525–533, 1993.

[101] Donald M. Monro, Soumyadip Rakshit, and Dexin Zhang. Dct-based iris recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):586–595, 2007.

[102] Ian T. Nabney. *NETLAB: Algorihtms for Pattern Recognition.* Advances in Pattern Recognition. Springer, London, 2004.

[103] Miguel Á. Carreira-Perpi nán. A review of dimension reduction techniques. Technical report, University of Sheffield, 1997.

[104] Michel Neuhaus and Horst Bunke. Edit distance-based kernel functions for structural pattern classification. *Pattern Recognition*, 39(10):1852–1863, October 2006.

[105] Oleg Okun, Helen Priisalu, and Alexessander Alves. Fast non-negative dimensionality reduction for protein fold recognition. In *ECML*, pages 665–672, 2005.

[106] Zhengjun Pan, Rod Adams, and Hamid Bolouri. Dimensionality reduction of face images using discrete cosine transforms for recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.

[107] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall, 1982.

[108] Terry R. Payne and Peter Edwards. Dimensionality reduction through sub-space mapping for nearest neighbour algorithms. In *Proceedings of European conference on Machine learning*, pages 331–343, 2000.

[109] Mihail Popescu, James M. Keller, and Joyce A. Mitchell. Fuzzy measures on the gene ontology for gene product similarity. *IEEE Transactions on Computational Biology and Bioinformatics*, 3(3):263–274, 2006.

[110] Jian Qiu, Martial Hue, Asa Ben-Hur, Jean-Philippe Vert, and William Stafford Noble. An alignment kernel for protein structures. In *Bioinformatics*, to appear in 2007.

[111] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive computation and machine learning. The MIT Press, Cambridge, Massachusetts, 2006.

[112] Dmitrii N. Rassokhin and Dimitris K. Agrafiotis. A modified update rule for stochastic proximity embedding. *Journal of Molecular Graphics and Modelling 22 (2003) 133C140*, 22:133–140, 2003.

[113] Bisser Raytchev, Ikushi Yoda, and Katsuhiko Sakaue. Multi-view face recognition by non-linear dimensionality reduction and generalized linear models. In *FGR '06: Proceedings*

of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06), pages 625–630, Washington, DC, USA, 2006.

[114] Matthew Robards, Junbin Gao, and Philip Charlton. A discriminant analysis for under-sampled data. In *Second International Workshop on Integrating AI and Data Mining*, volume 1, 2007.

[115] Sam Roweis, Lawrence K. Saul, and Geoffrey E. Hinton. Global coordination of local linear models. In *Advances in Neural Information Processing Systems 14*. MIT Press, Cambridge, MA, 2002.

[116] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(22):2323–2326, Dec. 2000.

[117] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Information Retieval and Language Processing*, 18(11):613–620, November 1975.

[118] Gerard Salton. *Automatic Text Processing: The Transformation Analysis and Retrieval of Information by Computer*. Addison-Wesley, 1989.

[119] Lawrence K. Saul, Kilian Q. Weinberger, Fei Sha, Jihun Ham, and Daniel D. Lee. Spectral methods for dimensionality reduction. In Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien, editors, *Semi-Supervised Learning*. MIT Press, MA, 2006.

[120] B. Schölkopf and A.J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, Cambridge, MA, 2002.

[121] Bernhard Schölkopf, Alexander J. Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319, 1998.

[122] Bernhard Schölkopf, Koji Tsudo, and Jean-Philippe Vert. *Kernel Methods in Computational Biology*. Computational Molecular Biology. The MIT Press, 2004.

[123] H.S. Seung and D. Lee. The manifold ways of perception. *Science*, 290(22):2268–2269, Dec. 2000.

[124] L. Sirovich and M. Kirby. Low-dimension procedure for the charactoerinzation of human faces. *Journal of the Optical Society of America*, 4(3):519–524, 1987.

[125] Masashi Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of Machine Learning Research*, 8:1027–1061, 2007.

[126] Johan A.K. Suykens. Data visualization and dimensionality reduction using kernel maps with a reference point. Technical Report 07-22, K.U. Leuven, ESAT-SCD/SISTA, 2007.

[127] Feng Tang and Hai Tao. Binary principal component analysis. In *Proceeding of British Machine Vision Conference*, pages 377–386, 2006.

[128] Yee Whye Teh and Sam Roweis. Automatic alignment of local representations. In S. Thrun S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*. MIT Press.

[129] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(22):2319–2323, Dec. 2000.

[130] M. Teodoro, G. Phillips, and L. Kavraki. A dimensionality reduction approach to modeling protein flexibility. In *International Conference on Computational Molecular Biology (RECOMB)*, 2002.

[131] Miguel L. Teodoro, George N. Phillips Jr, and Lydia E. Kavraki. A dimensionality reduction approach to modeling protein flexibility. In *International Conference on Computational Molecular Biology (RECOMB)*, pages 299–308, April 2002.

[132] Miguel L. Teodoro, George N. Phillips Jr, and Lydia E. Kavraki. Understanding protein flexibility through dimensionality reduction. *Journal of Computational Biology*, 10(3-4):617–634, June 2003.

[133] Michael E. Tipping and Christopher M. Bishop. Mixtures of probabilistic principal component analysers. *Neural Computation*, 11(2):443–482, 1999.

[134] Michael E. Tipping and Christopher M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 61(3):611–622, 1999.

[135] David G. Underhill, Luke K. McDowell, David J. Marchette, and Jeffrey L. Solka. Enhancing text analysis via dimensionality reduction. In *Proceedings of IEEE International Conference on Information Reuse and Integration*, pages 384–353, 2007.

[136] L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik. *Dimensionality Reduction: A Comparative Review*, http://www.cs.unimaas.nl/l.vandermaaten/dr/DR_draft.pdf, 2007.

[137] Jakob Verbeek. Learning nonlinear image manifolds by global alignment of local linear models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1236–1250, 2006.

[138] Vishwa Vinay, Ingemar J. Cox, Ken Wood, and Natasa Milic-Frayling. A comparison of dimensionality reduction techniques for text retrieval. In *Proceedings of the Fourth International Conference on Machine Learning and Applications*, 2005.

[139] Peng Wang, Qiang Ji, and James L. Wayman. Modeling and predicting face recognition system performance based on analysis of similarity scores. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):665–670, 2007.

[140] Chris Watkins. Dynamic alignment kernels. Technical Report CSD-TR-98-11, University of London, 1999.

[141] J. Wayman, A. Jain, D. Maltoni, and D. Maio, editors. *Biometric Systems Technology, Design and Performance Evaluation*. Springer-Verlag, London, 2005.

[142] Kilian Q. Weinberger, Fei Sha, and Lawrence K. Saul. Learning a kernel matrix for nonlinear dimensionality reduction. In *Proceedings of the International Conference in Machine Learning*, pages 106–113, 2004.

[143] James A. Wise, James J. Thomas, Kelly Pennock, David Lantrip, Marc Pottier, Anne Schur, and Vern Crow. Visualizing the non-visual: Spatial analysis and interaction with

information from text documents. In *Proceedings of the IEEE Information Visualization Symposium*, pages 51–58, 1995.

[144] Lior Wolf and Ammon Shashua. Learning over sets using kernel principal angles. *Journal of Machine Learning Research*, 4(10):913–931, 2003.

[145] Mingrui Wu and Jason Farquhar. A subspace kernel for nonlinear feature extraction. In *International Joint Conference on Artificial Intelligence*, pages 1125–1130, 2007.

[146] Mingrui Wu, Kai Yu, Shipeng Yu, and Bernhard Schölkopf. Local learning projections. In *Proceedings of the 24th International Conference on Machine Learning*, 2007.

[147] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, and Stephen Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007.

[148] Jian Yang, Fuxin Li, and Jue Wang. A better scaled local tangent space alignment algorithm. In *Proceedings of IEEE International Joint Conference on Neural Networks*, volume 2, pages 1006–1011, 2005.

[149] Jian Yang, David Zhang, Jing yu Yang, and Ben Niu. Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):650–664, 2007.

[150] Li Yang. Builidng k-connected neighborhood graphs for isometric data embedding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):827–831, May 2006.

[151] Xin Yang, Haoying Fu, Hongyuan Zha, and Jesse Barlow. Semi-supervised nonlinear dimensionality reduction. In *Proceedings of the International Conference in Machine Learning*, 2006.

[152] Yiming Yang and Jan O. Pedersen. A comparative study on feature selection in text categorization. In *Proceedings of the International Conference in Machine Learning*, pages 412–420, 1997.

[153] Hujun Yin. Nonlinear dimensionality reduction and data visualization: A review. *International Journal of Automation and Computing*, 4(3):294–303, July 2007.

[154] Hongyuan Zha and Zhenyue Zhang. Isometric embedding and continuum isomap. In *Proceedings of the Twentieth International Conference on Machine Learning*, 2003.

[155] Tianhao Zhang, Jie Yanga, Deli Zhao, and Xinliang Ge. Linear local tangent space alignment and application to face recognition. *Neurocomputing*, 70:1547–1553, 2007.

[156] Zhenyue Zhang and Hongyuan Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space. *SIAM Journal on Scientific Computing*, 26(1):313–338, 2005.

[157] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. Technical report, statistics department, Stanford University, 2004.