
Metric Learning: A Survey

Metric Learning: A Survey

Brian Kulis

Ohio State University

Columbus, OH 43210

USA

kulis@cse.ohio-state.edu

now

the essence of **knowledge**

Boston – Delft

Foundations and Trends[®] in Machine Learning

Published, sold and distributed by:

now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:

now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is B. Kulis, Metric Learning: A Survey, Foundations and Trends[®] in Machine Learning, vol 5, no 4, pp 287–364, 2012.

ISBN: 978-1-60198-696-2
© 2013 B. Kulis

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc. for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1-781-871-0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

**Foundations and Trends[®] in
Machine Learning**
Volume 5 Issue 4, 2012
Editorial Board

Editor-in-Chief:

Michael Jordan

Department of Electrical Engineering and Computer Science

Department of Statistics

University of California, Berkeley

Berkeley, CA 94720-1776

Editors

Peter Bartlett (UC Berkeley)

Yoshua Bengio (Université de Montréal)

Avrim Blum (Carnegie Mellon University)

Craig Boutilier (University of Toronto)

Stephen Boyd (Stanford University)

Carla Brodley (Tufts University)

Inderjit Dhillon (University of Texas at
Austin)

Jerome Friedman (Stanford University)

Kenji Fukumizu (Institute of Statistical
Mathematics)

Zoubin Ghahramani (Cambridge
University)

David Heckerman (Microsoft Research)

Tom Heskes (Radboud University Nijmegen)

Geoffrey Hinton (University of Toronto)

Aapo Hyvarinen (Helsinki Institute for
Information Technology)

Leslie Pack Kaelbling (MIT)

Michael Kearns (University of
Pennsylvania)

Daphne Koller (Stanford University)

John Lafferty (Carnegie Mellon University)

Michael Littman (Rutgers University)

Gabor Lugosi (Pompeu Fabra University)

David Madigan (Columbia University)

Pascal Massart (Université de Paris-Sud)

Andrew McCallum (University of
Massachusetts Amherst)

Marina Meila (University of Washington)

Andrew Moore (Carnegie Mellon
University)

John Platt (Microsoft Research)

Luc de Raedt (Albert-Ludwigs Universitaet
Freiburg)

Christian Robert (Université
Paris-Dauphine)

Sunita Sarawagi (IIT Bombay)

Robert Schapire (Princeton University)

Bernhard Schoelkopf (Max Planck Institute)

Richard Sutton (University of Alberta)

Larry Wasserman (Carnegie Mellon
University)

Bin Yu (UC Berkeley)

Editorial Scope

Foundations and Trends[®] in Machine Learning will publish survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends[®] in Machine Learning, 2012, Volume 5, 4 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Foundations and Trends[®] in
Machine Learning
Vol. 5, No. 4 (2012) 287–364
© 2013 B. Kulis
DOI: 10.1561/22000000019



Metric Learning: A Survey

Brian Kulis

*Ohio State University, CSE Department, Columbus, OH 43210, USA,
kulis@cse.ohio-state.edu*

Abstract

The *metric learning* problem is concerned with learning a distance function tuned to a particular task, and has been shown to be useful when used in conjunction with nearest-neighbor methods and other techniques that rely on distances or similarities. This survey presents an overview of existing research in metric learning, including recent progress on scaling to high-dimensional feature spaces and to data sets with an extremely large number of data points. A goal of the survey is to present as unified as possible a framework under which existing research on metric learning can be cast. The first part of the survey focuses on linear metric learning approaches, mainly concentrating on the class of Mahalanobis distance learning methods. We then discuss nonlinear metric learning approaches, focusing on the connections between the nonlinear and linear approaches. Finally, we discuss extensions of metric learning, as well as applications to a variety of problems in computer vision, text analysis, program analysis, and multimedia.

Contents

1	Introduction	1
2	Distance Learning via Linear Transformations	5
2.1	A Simple Motivating Example	5
2.2	Basic Techniques and Notation	6
2.3	Regularized Transformation Learning	8
2.4	Representative Special Cases	13
2.5	Optimization Techniques	23
2.6	Summary	33
3	Nonlinear Models for Metric Learning	35
3.1	Kernelization of Linear Methods	36
3.2	Other Nonlinear Methods	47
4	Extensions	53
4.1	Metric Learning for Kernel Regression	53
4.2	Metric Learning for Ranking	54
4.3	Dimensionality Reduction and Data Visualization	55
4.4	Database Indexing	56
4.5	Domain Adaptation	57

5 Applications	59
5.1 Computer Vision	59
5.2 Text Analysis	62
5.3 Other Applications	64
6 Conclusions	67
A Representer Theorem Proof	69
Acknowledgments	73
References	75

1

Introduction

Consider the images in Figure 1.1, and imagine a scenario in which we must compute similarity or distances over pairs of images (for example, for clustering or nearest neighbor classification). A basic question that arises is precisely *how* to assess the similarity or distance between the pairs of images. For instance, if our goal is to find matching faces based on identity, then we should choose a distance function that emphasizes appropriate features (hair color, ratios of distances between facial keypoints, etc). But we may also have an application where we want to determine the pose of an individual, and therefore require a distance function that captures pose similarity. Clearly other features are more applicable in this scenario. To handle multiple similarity or distance metrics, we could attempt to determine by hand an appropriate distance function for each task, by an appropriate choice of features and the combination of those features. However, this approach may require significant effort and may not be robust to changes in the data. A desirable alternative — and the focus of this survey — is to apply *metric learning*, which aims to automate this process and learn task-specific distance functions in a supervised manner.

2 Introduction



Fig. 1.1 Example face data set. In one application, our notion of “distance” between faces may depend on the pose, whereas in another application it may depend on the identity.

A possible informal formulation of the metric learning problem could be given as follows: given an input distance function $d(\mathbf{x}, \mathbf{y})$ between objects \mathbf{x} and \mathbf{y} (for example, the Euclidean distance), along with supervised information regarding an ideal distance, construct a new distance function $\tilde{d}(\mathbf{x}, \mathbf{y})$ which is “better” than the original distance function (we could also easily replace “distance” with “similarity,” and d with s for some similarity function $s(\mathbf{x}, \mathbf{y})$). This survey will focus, for the most part, on learning distance functions $\tilde{d}(\mathbf{x}, \mathbf{y})$ of the form $d(f(\mathbf{x}), f(\mathbf{y}))$ for some function f — that is, we learn some mapping f and utilize the original distance function over the mapped data. We will denote this approach as *global metric learning* methods, since they learn a single mapping f to be applied to all the data.

One possible drawback to the above definition of metric learning is that it assumes that we have at least some supervision available to learn the new distance; the fact that we assume supervision seems somewhat arbitrary. Take, for instance, dimensionality reduction: linear methods such as principal components analysis can be viewed as constructing a linear transformation P to be applied globally to the data,

in an unsupervised manner. The resulting distance between objects is therefore $d(P\mathbf{x}, P\mathbf{y})$, and one may claim that this is also a form of metric learning. In contrast, the methods we study typically have supervised information regarding the structure of the desired distance function. For example, one popular form of supervision — relative distance constraints — assumes we may not know the target distance between pairs of instances, but does assume we know that object \mathbf{x} is more similar to \mathbf{y} than it is to \mathbf{z} . The fact that the supervised information is a function of the ideal distance (or similarity) is key to distinguishing the methods we study in this survey from other existing techniques such as dimensionality reduction methods or classification techniques. Furthermore, incorporating such supervision lends itself to interesting algorithmic and analysis challenges, as we will see. Thus, in this survey we will mainly focus on metric learning as a supervised learning problem.

We will break down global metric learning into two subclasses — linear and nonlinear. For both cases, we will mainly focus on the case where the input distance function is the Euclidean distance, i.e., $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$. In the linear case, we aim to learn a linear mapping based on supervision, which we can encode as a matrix G such that the learned distance is $\|G\mathbf{x} - G\mathbf{y}\|_2$. This paradigm is by far the most prevalent in the metric learning community due to the fact that many of the resulting formulations are tractable (at the least, local solutions can be found easily). To achieve convexity, many methods assume that G is square and full-rank, leading to convex optimization problems with positive semi-definiteness constraints. We will discuss such methods in Section 2.

We study nonlinear methods for global metric learning in Section 3. In this case, the distance function is the more general $d(\mathbf{x}, \mathbf{y}) = \|f(\mathbf{x}) - f(\mathbf{y})\|_2$. One of the most well-understood and effective techniques for learning such nonlinear mappings is to extend linear methods via kernelization. The basic idea is to learn a linear mapping in the feature space of some potentially nonlinear function ϕ ; that is, the distance function may be written $d(\mathbf{x}, \mathbf{y}) = \|G\phi(\mathbf{x}) - G\phi(\mathbf{y})\|_2$, where ϕ may be a nonlinear function. While it may not appear that we have gained anything by this, if we further assume that we can compute

4 Introduction

the *kernel* function $\kappa(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^T \phi(\mathbf{y})$, then it turns out that we may efficiently learn G in the input space using extensions of linear techniques. Crucially, the resulting algorithms scale independently of the dimensionality of the feature space of ϕ , allowing us to utilize kernel functions whose embedding functions may be extremely high-dimensional (or even infinite-dimensional, as in the Gaussian kernel). A core result that we discuss is a representer theorem that demonstrates when such metrics may be learned. Beyond kernelization, we discuss some other proposed methods for nonlinear metric learning, including methods based on neural networks.

The goal of the survey is to provide an overview of recent advances in metric learning. For the sake of clarity, we will attempt to present as much of the literature as possible under a unified framework. Of course, given the broad scope of the metric learning problem, and the fact that not all material fits neatly into such a unified presentation, we will have to divert from the main presentation from time to time. In addition to presenting the main metric learning models and algorithms that have been studied, we also focus on several recent applications, including applications from computer vision, multimedia, and text analysis. It is our hope that this survey will synthesize much of the recent work on metric learning, and inspire new algorithms and applications.

References

- [1] A. Beck and M. Teboulle, “Mirror descent and nonlinear projected subgradient methods for convex optimization,” *Operations Research Letters*, vol. 31, pp. 167–175, 2003.
- [2] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.
- [3] M. Bilenko, S. Basu, and R. Mooney, “Integrating constraints and metric learning in semi-supervised clustering,” in *Proceedings of International Conference on Machine Learning (ICML)*, 2004.
- [4] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [5] L. Bottou, “Online algorithms and stochastic approximations,” in *Online Learning and Neural Networks*, (D. Saad, ed.), Cambridge University Press, 1998.
- [6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [7] J. P. Boyle and R. L. Dykstra, “A method for finding projections onto the intersection of convex sets in Hilbert spaces,” *Lecture Notes in Statistics*, vol. 37, pp. 28–47, 1986.
- [8] L. M. Bregman, “The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming,” *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 3, pp. 200–217, 1967.
- [9] Q. Cao, Z. C. Guo, and Y. Ying, “Generalization bounds for metric and similarity learning,” arXiv:1207.5437, 2012.
- [10] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge University Press, 2006.

76 References

- [11] R. Chatpatanasiri, T. Korsrilabutr, P. Tangchanachaianan, and B. Kijirikul, "A new kernelization framework for Mahalanobis distance learning algorithms," *Neurocomputing*, vol. 73, no. 10–12, pp. 1570–1579, 2010.
- [12] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "An online algorithm for large scale image similarity learning," in *Advances in Neural Information Processing Systems (NIPS)*, 2009.
- [13] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.
- [14] P. Ciaccia, M. Patella, and P. Zezula, "M-Tree: An efficient access method for similarity search in metric spaces," in *Proceedings of International Conference on Very Large Data Bases (VLDB)*, 1997.
- [15] K. Crammer, O. Dekel, S. Shalev-Shwartz, and Y. Singer, "Online passive-aggressive algorithms," in *Advances in Neural Information Processing Systems*, 2004.
- [16] O. G. Cula and K. J. Dana, "3D texture recognition using bidirectional feature histograms," *International Journal of Computer Vision (IJCV)*, vol. 59, no. 1, pp. 33–60, 2004.
- [17] Curious Labs, Inc., *Poser 5 — Reference Manual*. Santa Cruz, CA, 2002.
- [18] H. Daume, "Frustratingly easy domain adaptation," in *Conference of the Association for Computational Linguistics (ACL)*, 2007.
- [19] J. Davis and I. S. Dhillon, "Structured metric learning for high-dimensional problems," in *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008.
- [20] J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon, "Information-theoretic metric learning," in *Proceedings of International Conference on Machine Learning (ICML)*, 2007.
- [21] R. Fletcher, "A new variational result for quasi-Newton formulae," *SIAM Journal on Optimization*, vol. 1, no. 1, 1991.
- [22] J. Friedman, J. Bentley, and A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Transactions on Mathematics Software*, vol. 3, no. 3, pp. 209–226, 1977.
- [23] A. Frome, Y. Singer, F. Sha, and J. Malik, "Learning globally consistent local distance functions for shape-based image retrieval and classification," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2007.
- [24] T. Gaertner, "A survey of kernels for structured data," *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 1, 2003.
- [25] A. Globerson and S. Roweis, "Metric learning by collapsing classes," in *Advances in Neural Information Processing Systems (NIPS)*, 2005.
- [26] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood components analysis," in *Advances in Neural Information Processing Systems (NIPS)*, 2004.
- [27] A. A. Goldstein, "Convex programming in Hilbert space," *Bulletin of the American Mathematical Society*, vol. 70, pp. 709–710, 1964.
- [28] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 1996.

- [29] K. Grauman and T. Darrell, “The pyramid match kernel: Efficient learning with sets of features,” *Journal of Machine Learning Research*, vol. 8, pp. 725–760, 2007.
- [30] M. Guillaumin, J. Verbeek, and C. Schmid, “Is that you? Metric learning approaches for face identification,” in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2009.
- [31] J. Ha, C. Rossbach, J. Davis, I. Roy, D. Chen, H. Ramadan, and E. Witchel, “Improved error reporting for software that uses black box components,” in *Proceedings of Programming Language Design and Implementation (PLDI)*, 2007.
- [32] S. C. H. Hoi, W. Liu, M. R. Lyu, and W. Y. Ma, “Learning distance metrics with contextual constraints for image retrieval,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.
- [33] G. Hua, M. Brown, and S. Winder, “Discriminant embedding for local image descriptors,” in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2007.
- [34] P. Indyk and R. Motwani, “Approximate nearest neighbors: Towards removing the curse of dimensionality,” in *Proceedings of Symposium on Theory of Computing (STOC)*, 1998.
- [35] P. Jain, B. Kulis, J. Davis, and I. Dhillon, “Metric and kernel learning using a linear transformation,” *Journal of Machine Learning Research*, vol. 13, pp. 519–547, 2012.
- [36] P. Jain, B. Kulis, and I. Dhillon, “Inductive regularized learning of kernel functions,” in *Advances in Neural Information Processing Systems (NIPS)*, 2010.
- [37] P. Jain, B. Kulis, I. Dhillon, and K. Grauman, “Online metric learning and fast similarity search,” in *Advances in Neural Information Processing Systems (NIPS)*, 2008.
- [38] P. Jain, B. Kulis, and K. Grauman, “Fast image search for learned metrics,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.
- [39] W. James and C. Stein, “Estimation with quadratic loss,” *Proceedings of Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 361–379, 1961.
- [40] R. Jin, S. Wang, and Y. Zhou, “Regularized distance metric learning: Theory and algorithm,” in *Advances in Neural Information Processing Systems (NIPS)*, 2009.
- [41] S. Kaski and J. Peltonen, “Informative discriminant analysis,” in *Proceedings of International Conference on Machine Learning (ICML)*, 2003.
- [42] D. Kedem, S. Tyree, K. Q. Weinberger, F. Sha, and G. Lanckriet, “Non-linear metric learning,” in *Advances in Neural Information Processing Systems (NIPS)*, 2012.
- [43] B. Kulis, P. Jain, and K. Grauman, “Fast similarity search for learned metrics,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2143–2157, 2009.
- [44] B. Kulis, K. Saenko, and T. Darrell, “What you saw is not what you get: Domain adaptation using asymmetric kernel transforms,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

78 References

- [45] B. Kulis, M. Sustik, and I. Dhillon, "Learning low-rank kernel matrices," in *Proceedings of International Conference on Machine Learning (ICML)*, 2006.
- [46] B. Kulis, M. Sustik, and I. Dhillon, "Low-rank kernel learning with Bregman matrix divergences," *Journal of Machine Learning Research*, vol. 10, pp. 341–376, 2009.
- [47] J. Kwok and I. Tsang, "Learning with idealized kernels," in *Proceedings of International Conference on Machine Learning (ICML)*, 2003.
- [48] G. Lebanon, "Metric learning for text documents," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 497–508, 2006.
- [49] E. S. Levitin and B. T. Polyak, "Constrained minimization problems," *USSR Computational Mathematics and Mathematical Physics*, vol. 6, pp. 1–50, 1966.
- [50] H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins, "Text classification using string kernels," *Journal of Machine Learning Research*, vol. 2, pp. 419–444, 2002.
- [51] P. C. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1, pp. 49–55, 1936.
- [52] B. McFee and G. Lanckriet, "Metric learning to rank," in *Proceedings of International Conference on Machine Learning ICML*, 2010.
- [53] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Interscience, 2004.
- [54] S. Parameswaran and K. Q. Weinberger, "Large margin multi-task metric learning," in *Advances in Neural Information Processing Systems (NIPS)*, 2010.
- [55] R. Rosales and G. Fung, "Learning sparse metric via linear programming," in *Proceedings of SIGKDD Conference*, 2006.
- [56] S. Roweis, "EM algorithms for PCA and SPCA," in *Advances in Neural Information Processing Systems*, 1998.
- [57] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2010.
- [58] R. Salakhutdinov and G. Hinton, "Learning a nonlinear embedding by preserving class neighbourhood structure," in *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2007.
- [59] B. Schoelkopf and A. Smola, *Learning with Kernels*. MIT Press, 2002.
- [60] B. Schoelkopf, A. Smola, and K.-R. Mueller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [61] M. Schultz and T. Joachims, "Learning a distance metric from relative comparisons," in *Advances in Neural Information Processing Systems (NIPS)*, 2003.
- [62] G. Shakhnarovich, P. Viola, and T. Darrell, "Fast pose estimation with parameter-sensitive hashing," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2003.
- [63] S. Shalev-Shwartz, Y. Singer, and A. Y. Ng, "Online learning of pseudo-metrics," in *Proceedings of International Conference on Machine Learning (ICML)*, 2004.
- [64] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

- [65] M. Slaney, K. Q. Weinberger, and W. White, "Learning a metric for music similarity," in *International Symposium on Music Information Retrieval (ISMIR)*, 2008.
- [66] N. Snavely, S. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D," in *Proceedings of ACM SIGGRAPH*, 2006.
- [67] I. Takeuchi, M. Nakagawa, and M. Seto, "Metric learning for DNA microarray data analysis," in *Proceedings of International Workshop on Statistical-Mechanical Informatics (IW-SMI)*, 2009.
- [68] M. Taylor, B. Kulis, and F. Sha, "Metric learning for reinforcement learning agents," in *Proceedings of International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2011.
- [69] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, no. 1, pp. 267–288, 1996.
- [70] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of Royal Statistical Society, Series B*, vol. 21, no. 3, pp. 611–622, 1999.
- [71] L. Torresani and K. Lee, "Large margin component analysis," in *Advances in Neural Information Processing Systems (NIPS)*, 2007.
- [72] D. Tran and A. Sorokin, "Human activity recognition with metric learning," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2008.
- [73] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.
- [74] J. Uhlmann, "Satisfying general proximity/similarity queries with metric trees," *Information Processing Letters*, vol. 40, pp. 175–179, 1991.
- [75] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2032–2047, 2009.
- [76] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," in *Advances in Neural Information Processing Systems (NIPS)*, 2005.
- [77] K. Q. Weinberger and L. K. Saul, "Fast solvers and efficient implementations for distance metric learning," in *Proceedings of International Conference on Machine Learning (ICML)*, 2008.
- [78] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.
- [79] K. Q. Weinberger and G. Tesauro, "Metric learning for kernel regression," in *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2007.
- [80] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning, with application to clustering with side-information," in *Advances in Neural Information Processing Systems (NIPS)*, 2002.
- [81] H. Xiong and X. Chen, "Kernel-based distance metric learning for microarray data classification," *BMC Bioinformatics*, vol. 7, p. 299, 2006.
- [82] Y. Ying, K. Huang, and C. Campbell, "Sparse metric learning via smooth optimization," in *Advances in Neural Information Processing Systems (NIPS)*, 2009.

80 *References*

- [83] Y. Ying and P. Li, “Distance metric learning with eigenvalue optimization,” *Journal of Machine Learning Research*, vol. 13, pp. 1–26, 2012.
- [84] M. Zinkevich, “Online convex programming and generalized infinitesimal gradient ascent,” in *Proceedings of International Conference on Machine Learning (ICML)*, 2003.