

References

- 5739 Agarwal, Arvind, and III, Hal Daumé. 2010. A Geometric View of Conjugate Priors.
5740 *Machine Learning*, **81**(1), 99–113. pages 200
- 5741 Agresti, A. 2002. *Categorical Data Analysis*. Wiley. pages 241
- 5742 Akaike, Hirotugu. 1974. A New Look at the Statistical Model Identification. *IEEE*
5743 *Transactions on Automatic Control*, **19**(6), 716–723. pages 247
- 5744 Akhiezer, N.I., and Glazman, I.M. 1993. *Theory of Linear Operators in Hilbert Space*.
5745 Dover Publications, Inc. pages 341
- 5746 Alpaydin, Ethem. 2010. *Introduction to Machine Learning*. The MIT Press. pages 13
- 5747 Amari, Shun-ichi. 2016. *Information Geometry and Its Applications*. Springer. pages
5748 183
- 5749 Argyriou, Andreas, and Dinuzzo, Francesco. 2014. A Unifying View of Representer
5750 Theorems. Pages 748–756 of: Xing, Eric P., and Jebara, Tony (eds), *Proceedings*
5751 *of the 31st International Conference on Machine Learning*. Proceedings of Machine
5752 Learning Research, vol. 32, no. 2. Beijing, China: PMLR. pages 331
- 5753 Aronszajn, N. 1950. Theory of Reproducing Kernels. *Transactions of the American*
5754 *Mathematical Society*, **68**, 337–404. pages 338, 341
- 5755 Bakir, Gökhan, Hofmann, Thomas, Schölkopf, Bernhard, Smola, Alexander J., Taskar,
5756 Ben, and Vishwanathan, S.V.N (eds). 2007. *Predicting Structured Data*. MIT Press.
5757 pages 238
- 5758 Barber, David. 2012. *Bayesian Reasoning and Machine Learning*. Cambridge University
5759 Press. pages 14, 205, 274, 316
- 5760 Ben-Hur, Asa, Ong, Cheng Soon, Sonnenburg, Sören, Schölkopf, Bernhard, and Rätsch,
5761 Gunnar. 2008. Support Vector Machines and Kernels for Computational Biology.
5762 *PLoS Computational Biology*, **4**(10), e1000173. pages 338, 341
- 5763 Bennett, Kristin P., and Bredensteiner, Erin J. 2000. Duality and Geometry in SVM
5764 Classifiers. In: *Proceedings of the Seventeenth International Conference on Machine*
5765 *Learning*. pages 336
- 5766 Bertsekas, Dimitri P. 1999. *Nonlinear Programming*. Athena Scientific. pages 224
- 5767 Bertsekas, Dimitri P. 2009. *Convex Optimization Theory*. Athena Scientific. pages 224
- 5768 Bickel, Peter J., and Doksum, Kjell. 2006. *Mathematical Statistics, Basic Ideas and*
5769 *Selected Topics, Vol I*. Prentice Hall. pages 204
- 5770 Bickson, Danny, Dolev, Danny, Shental, Ori, Siegel, Paul H., and Wolf, Jack K. 2007.
5771 Linear Detection via Belief Propagation. In: *Proceedings of the Annual Allerton Con-*
5772 *ference on Communication, Control, and Computing*. pages 233
- 5773 Billingsley, Patrick. 1995. *Probability and Measure*. Wiley. pages 171, 204
- 5774 Bishop, Christopher M. 1995. *Neural Networks for Pattern Recognition*. Clarendon
5775 Press. pages 274
- 5776 Bishop, Christopher M. 1999. Bayesian PCA. Pages 382–388 of: *Advances in Neural*
5777 *Information Processing Systems*. pages 296

- 5778 Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. Information
5779 Science and Statistics. Springer-Verlag. pages 13, 87, 163, 168, 180, 200, 201, 204,
5780 205, 233, 238, 259, 274, 287, 315, 316, 341
- 5781 Blei, David M., Kucukelbir, Alp, and McAuliffe, Jon D. 2017. Variational Inference: A
5782 Review for Statisticians. *Journal of the American Statistical Association*, **112**(518),
5783 859–877. pages 204
- 5784 Bonnans, J. Frédéric, Gilbert, J. Charles, Lemaréchal, Claude, and Sagastizábal, Clau-
5785 dia A. 2006. *Numerical Optimization: Theoretical and Practical Aspects*. 2nd edn.
5786 Springer Verlag. pages 224
- 5787 Borwein, Jonathan M., and Lewis, Adrian S. 2006. *Convex Analysis and Nonlinear*
5788 *Optimization*. 2nd edn. Canadian Mathematical Society. pages 224
- 5789 Bottou, Leon. 1998. Online Algorithms and Stochastic Approximations. Pages 1–34
5790 of: *Online Learning and Neural Networks*. Cambridge University Press. pages 212
- 5791 Bottou, Léon, Curtis, Frank E., and Nocedal, Jorge. 2018. *Optimization Methods for*
5792 *Large-Scale Machine Learning*. pages 212, 224
- 5793 Boucheron, Stephane, Lugosi, Gabor, and Massart, Pascal. 2013. Concentration In-
5794 equalities: A Nonasymptotic Theory of Independence. pages 168
- 5795 Boyd, Stephen, and Vandenberghe, Lieven. 2004. *Convex Optimization*. Cambridge
5796 University Press. pages 210, 213, 214, 224
- 5797 Brooks, Steve, Gelman, Andrew, Jones, Galin L., and Meng, Xiao-Li (eds). 2011. *Hand-*
5798 *book of Markov Chain Monte Carlo*. Chapman and Hall/CRC. pages 204
- 5799 Brown, Lawrence D. 1986. *Fundamentals of Statistical Exponential Families: with Ap-*
5800 *plications in Statistical Decision Theory*. Lecture Notes - Monograph Series. Institute
5801 of Mathematical Statistics. pages 202
- 5802 Bryson, Arthur E. 1961. A Gradient Method for Optimizing Multi-stage Allocation
5803 Processes. In: *Proceedings of the Harvard University Symposium on Digital Computers*
5804 *and Their Applications*. pages 152, 154
- 5805 Bubeck, Sébastien. 2015. Convex Optimization: Algorithms and Complexity. *Founda-*
5806 *tions and Trends in Machine Learning*, **8**(3-4), 231–357. pages 224
- 5807 Casella, George, and Berger, Roger L. 2002. *Statistical Inference*. Duxbury. pages 177,
5808 184, 186, 241
- 5809 Chang, Chih-Chung, and Lin, Chih-Jen. 2011. LIBSVM: A library for support vector
5810 machines. *ACM Transactions on Intelligent Systems and Technology*, **2**, 27:1–27:27.
5811 Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. pages 341
- 5812 Chollet, Francois, and Allaire, J. J. 2018. *Deep Learning with R*. Manning Publications.
5813 pages 14
- 5814 Codd, Edgar F. 1990. *The Relational Model for Database Management*. Addison-Wesley
5815 Longman Publishing. pages 10
- 5816 Cunningham, John P., and Ghahramani, Zoubin. 2015. Linear Dimensionality Reduc-
5817 tion: Survey, Insights, and Generalizations. *Journal of Machine Learning Research*,
5818 **16**, 2859–2900. pages 296
- 5819 Dean, Jeffrey, Corrado, Greg S, Monga, Rajat, Chen, Kai, Devin, Matthieu, Le, Quoc V,
5820 Mao, Mark Z, Ranzato, Marc Aurelio, Senior, Andrew, Tucker, Paul, Yang, Ke, and
5821 Ng, Andrew Y. 2012. Large Scale Distributed Deep Networks. Pages 1–11 of: *Ad-*
5822 *vances in Neural Information Processing Systems*. pages 212
- 5823 Deisenroth, Marc P., and Mohamed, Shakir. 2012. Expectation Propagation in Gaus-
5824 sian Process Dynamical Systems. Pages 2618–2626 of: *Advances in Neural Informa-*
5825 *tion Processing Systems*. pages 233
- 5826 Deisenroth, Marc P., and Ohlsson, Henrik. 2011. A General Perspective on Gaussian
5827 Filtering and Smoothing: Explaining Current and Deriving New Algorithms. In:
5828 *Proceedings of the American Control Conference*. pages 192

- 5829 Deisenroth, Marc P., Fox, Dieter, and Rasmussen, Carl E. 2015. Gaussian Processes
5830 for Data-Efficient Learning in Robotics and Control. *IEEE Transactions on Pattern*
5831 *Analysis and Machine Intelligence*, **37**(2), 408–423. pages vi, 84
- 5832 Dempster, A. P., Laird, N. M., and Rubin, D. B. 1977. Maximum Likelihood from
5833 Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*,
5834 **39**(1), 1–38. pages 310
- 5835 Deng, Li, Seltzer, Michael L, Yu, Dong, Acero, Alex, Mohamed, Abdel-rahman, and
5836 Hinton, Geoffrey E. 2010. Binary Coding of Speech Spectrograms using a Deep
5837 Auto-Encoder. Pages 1692–1695 of: *Interspeech*. pages 75
- 5838 Devroye, Luc. 1986. *Non-Uniform Random Variate Generation*. Springer-Verlag. pages
5839 196
- 5840 Domingos, Pedro. 2012. A few useful things to know about machine learning. *Com-*
5841 *munications of the ACM*, **55**, 78–87. pages 13
- 5842 Dostál, Zdeněk. 2009. *Optimal Quadratic Programming Algorithms: With Applications*
5843 *to Variational Inequalities*. Springer-Verlag. pages 332
- 5844 Downey, Allen B. 2014. *Think Stats: Exploratory Data Analysis*. 2nd edn. O'Reilly
5845 Media. pages 204
- 5846 Dreyfus, Stuart. 1962. The Numerical Solution of Variational Problems. *Journal of*
5847 *Mathematical Analysis and Applications*, **5**(1), 30–45. pages 152, 154
- 5848 Drumm, Volker, and Weil, Wolfgang. 2001. *Lineare Algebra und Analytische Geometrie*.
5849 Lecture Notes, Universität Karlsruhe. pages 16
- 5850 Dudley, R. M. 2002. *Real Analysis and Probability*. Cambridge University Press. pages
5851 204
- 5852 Efron, Bradley, and Hastie, Trevor. 2016. *Computer Age Statistical Inference: Algorithms,*
5853 *Evidence and Data Science*. Cambridge University Press. pages 196
- 5854 Evgeniou, Theodoros, Pontil, Massimiliano, and Poggio, Tomaso. 2000. Statistical
5855 Learning Theory: A Primer. *International Journal of Computer Vision*, **38**(1), 9–13.
5856 pages 241
- 5857 Gal, Yarin, van der Wilk, Mark, and Rasmussen, Carl E. 2014. Distributed Variational
5858 Inference in Sparse Gaussian Process Regression and Latent Variable Models. In:
5859 *Advances in Neural Information Processing Systems*. pages 212
- 5860 Gärtner, Thomas. 2008. *Kernels for Structured Data*. World Scientific. pages 338
- 5861 Ghahramani, Zoubin. 2015. Probabilistic Machine Learning and Artificial Intelligence.
5862 *Nature*, **521**, 452–459. pages 174, 204
- 5863 Ghahramani, Zoubin, and Roweis, Sam T. 1999. Learning Nonlinear Dynamical Sys-
5864 tems using an EM Algorithm. In: Kearns, M. S., Solla, S. A., and Cohn, D. A. (eds),
5865 *Advances in Neural Information Processing Systems*, vol. 11. The MIT Press. pages
5866 316
- 5867 Gneiting, Tilmann, and Raftery, Adrian E. 2007. Strictly proper scoring rules, pre-
5868 diction, and estimation. *Journal of the American Statistical Association*, **102**(477),
5869 359–378. pages 341
- 5870 Goh, Gabriel. 2017. Why Momentum Really Works. *Distill*. pages 210
- 5871 Gonçalves, Hugo. 2014. *Legendre and Legendre-Fenchel transforms*. Accessed on 3
5872 March 2018. pages 224
- 5873 Goodfellow, Ian, Bengio, Yoshua, and Courville, Aaron. 2016. *Deep Learning*. MIT
5874 Press. <http://www.deeplearningbook.org>. pages 204, 238, 274
- 5875 Grinstead, Charles M., and Snell, J. Laurie. 1997. *Introduction to Probability*. American
5876 Mathematical Society. pages 166, 190, 204
- 5877 Hacking, Ian. 2001. *Probability and Inductive Logic*. Cambridge University Press. pages
5878 204
- 5879 Hasselblatt, Boris, and Katok, Anatole. 2003. *A first course in dynamics with a*
5880 *Panorama of Recent Developments*. Cambridge University Press. pages 166

- 5881 Hazan, Elad. 2015. Introduction to Online Convex Optimization. *Foundations and*
5882 *Trends*, **2**(3-4), 157–325. pages 224
- 5883 Hensman, James, Fusi, Nicolò, and Lawrence, Neil D. 2013. Gaussian Processes for
5884 Big Data. In: Nicholson, A., and Smyth, P. (eds), *Proceedings of the Conference on*
5885 *Uncertainty in Artificial Intelligence*. AUAI Press. pages 212
- 5886 Herbrich, Ralf, Minka, Tom, and Graepel, Thore. 2007. TrueSkill(TM): A Bayesian
5887 Skill Rating System. Pages 569–576 of: *Advances in Neural Information Processing*
5888 *Systems*. MIT Press. pages 233
- 5889 Hiriart-Urruty, Jean-Baptiste, and Lemaréchal, Claude. 2001. *Fundamentals of Convex*
5890 *Analysis*. Springer Verlag. pages 224, 332
- 5891 Hoffman, Matthew D., Blei, David M., Wang, Chong, and Paisley, John. 2013. Stochas-
5892 tic Variational Inference. *Journal of Machine Learning Research*, **14**(1), 1303–1347.
5893 pages 212
- 5894 Hofmann, Thomas, Schölkopf, Bernhard, and Smola, Alexander J. 2008. Kernel meth-
5895 ods in machine learning. *Ann. Statist.*, **36**(3), 1171–1220. pages 331
- 5896 Hogben, L. (ed). 2013. *Handbook of Linear Algebra*. 2nd edn. Discrete Mathematics
5897 and Its Applications. Chapman and Hall. pages 16
- 5898 Hotelling, Harold. 1933. Analysis of a Complex of Statistical Variables into Principal
5899 Components. *Journal of Educational Psychology*, **24**, 417–441. pages 75, 87, 276
- 5900 Imbens, Guido W., and Rubin, Donald B. 2015. *Causal Inference for statistics, social*
5901 *and biomedical sciences*. Cambridge University Press. pages 238
- 5902 Jacod, Jean, and Protter, Philip. 2004. *Probability Essentials*. 2nd edn. Springer-Verlag.
5903 pages 165, 204
- 5904 Jaynes, Edwin T. 2003. *Probability Theory: The Logic of Science*. Cambridge University
5905 Press. pages 166, 167, 174, 204
- 5906 Jefferys, William H., and Berger, James O. 1992. Ockham’s Razor and Bayesian Analy-
5907 sis. *American Scientist*, **80**, 64–72. pages 245
- 5908 Joachims, Thorsten. 1999. Making large-Scale SVM Learning Practical. In: Schölkopf,
5909 B., Burges, C., and Smola, A. (eds), *Advances in Kernel Methods - Support Vector*
5910 *Learning*. MIT Press. pages 341
- 5911 Julier, Simon J., and Uhlmann, Jeffrey K. 1997. A New Extension of the Kalman Filter
5912 to Nonlinear Systems. Pages 182–193 of: *Proceedings of AeroSense: 11th Symposium*
5913 *on Aerospace/Defense Sensing, Simulation and Controls*. pages 163
- 5914 Kalman, Dan. 1996. A singularly valuable decomposition: the SVD of a matrix. *The*
5915 *college mathematics journal*, **27**(1), 2–23. pages 115, 123
- 5916 Kalman, Rudolf E. 1960. A New Approach to Linear Filtering and Prediction Problems.
5917 *Transactions of the ASME—Journal of Basic Engineering*, **82**(Series D), 35–45. pages
5918 192
- 5919 Kelley, Henry J. 1960. Gradient Theory of Optimal Flight Paths. *Ars Journal*, **30**(10),
5920 947–954. pages 152, 154
- 5921 Kimeldorf, George S., and Wahba, Grace. 1970. A correspondence between Bayesian
5922 estimation on stochastic processes and smoothing by splines. *The Annals of Mathe-*
5923 *matical Statistics*, **41**(2), 495–502. pages 331
- 5924 Kingma, Diederik P, and Welling, Max. 2013. Auto-encoding variational bayes. *arXiv*
5925 *preprint arXiv:1312.6114*. pages 108
- 5926 Kittler, J., and Föglein, J. 1984. Contextual Classification of Multispectral Pixel Data.
5927 *Image and Vision Computing*, **2**(1), 13–29. pages 233
- 5928 Koller, Daphne, and Friedman, Nir. 2009. *Probabilistic Graphical Models*. MIT Press.
5929 pages 233, 238
- 5930 Lawrence, Neil. 2005. Probabilistic Non-linear Principal Component Analysis with
5931 Gaussian Process Latent Variable Models. *Journal of Machine Learning Research*,
5932 **6**(Nov.), 1783–1816. pages 296

- 5933 Leemis, Lawrence M., and McQueston, Jacquelyn T. 2008. Univariate Distribution
5934 Relationships. *The American Statistician*, **62**(1), 45–53. pages 196, 199
- 5935 Lehmann, Erich L., and Romano, Joseph P. 2005. *Testing Statistical Hypotheses*.
5936 Springer. pages 187
- 5937 Lehmann, Erich Leo, and Casella, George. 1998. *Theory of Point Estimation*. Springer.
5938 pages 202, 204
- 5939 Liesen, Jörg, and Mehrmann, Volker. 2015. *Linear Algebra*. Springer Undergraduate
5940 Mathematics Series. Springer. pages 16, 106
- 5941 Luenberger, David G. 1969. *Optimization by Vector Space Methods*. John Wiley and
5942 Sons. pages 224
- 5943 MacKay, David J. C. 2003. *Information Theory, Inference, and Learning Algorithms*. The
5944 Edinburgh Building, Cambridge CB2 2RU, UK: Cambridge University Press. pages
5945 245, 316
- 5946 MacKay, David J. C. 1992. Bayesian Interpolation. *Neural Computation*, **4**, 415–447.
5947 pages 245
- 5948 MacKay, David J. C. 1998. Introduction to Gaussian Processes. Pages 133–165 of:
5949 Bishop, C. M. (ed), *Neural Networks and Machine Learning*, vol. 168. Berlin, Ger-
5950 many: Springer. pages 275
- 5951 Mackay, David J.C. 2003. *Information Theory, Inference and Learning Algorithms*.
5952 Cambridge University Press. pages 13, 204, 205
- 5953 Manton, Jonathan H., and Amblard, Pierre-Olivier. 2015. A Primer on Reproducing
5954 Kernel Hilbert Spaces. *Foundations and Trends in Signal Processing*, **8**(12), 1–126.
5955 pages 341
- 5956 Maybeck, Peter S. 1979. *Stochastic Models, Estimation, and Control*. Mathematics in
5957 Science and Engineering, vol. 141. Academic Press, Inc. pages 163
- 5958 McEliece, Robert J., MacKay, David J. C., and Cheng, Jung-Fu. 1998. Turbo Decoding
5959 as an Instance of Pearl’s “Belief Propagation” Algorithm. *IEEE Journal on Selected*
5960 *Areas in Communications*, **16**(2), 140–152. pages 233
- 5961 Mnih, Volodymyr, Kavukcuoglu, Koray, Silver, David, Rusu, Andrei A., Veness, Joel,
5962 Bellemare, Marc G., Graves, Alex, Riedmiller, Martin, Fidjeland, Andreas K., Os-
5963 trovski, Georg, Petersen, Stig, Beattie, Charles, Sadik, Amir, Antonoglou, Ioan-
5964 nis, King, Helen, Kumaran, Dhharshan, Wierstra, Daan, Legg, Shane, and Hassabis,
5965 Demis. 2015. Human-Level Control through Deep Reinforcement Learning. *Nature*,
5966 **518**(Feb.), 529–533. pages 212
- 5967 Müller, Andreas C., and Guido, Sarah. 2016. *Introduction to Machine Learning with*
5968 *Python: A Guide for Data Scientists*. O’Reilly Publishing. pages 14
- 5969 Murphy, Kevin P. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge, MA,
5970 USA: MIT Press. pages 13, 192, 199, 204, 205, 244, 247, 274, 316, 341
- 5971 Neal, Radford M. 1996. *Bayesian Learning for Neural Networks*. Ph.D. thesis, Depart-
5972 ment of Computer Science, University of Toronto. pages 275
- 5973 Neal, Radford M., and Hinton, Geoffrey E. 1999. A View of the EM Algorithm that
5974 Justifies Incremental, Sparse, and Other Variants. Pages 355–368 of: Jordan, M. I.
5975 (ed), *Learning in Graphical Models*. MIT Press. pages 310
- 5976 Nelsen, Roger. 2006. *An Introduction to Copulas*. Springer Verlag. pages 187
- 5977 Nocedal, Jorge, and Wright, Stephen J. 2006. *Numerical Optimization*. Springer Series
5978 in Operations Research. Springer. pages 157, 224, 341
- 5979 Nowozin, Sebastian, Gehler, Peter V., Jancsary, Jeremy, and Lampert, Christoph H.
5980 (eds). 2014. *Advanced Structured Prediction*. MIT Press. pages 238
- 5981 O’Hagan, Anthony. 1991. Bayes-Hermite Quadrature. *Journal of Statistical Planning*
5982 *and Inference*, **29**, 245–260. pages 247
- 5983 Parzen, Emanuel. 1962. On Estimation of a Probability Density Function and Mode.
5984 *The Annals of Mathematical Statistics*, **33**(3), 1065–1076. pages 317

- 5985 Pearl, Judea. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible*
5986 *Inference*. Morgan Kaufmann. pages 167, 237
- 5987 Pearl, Judea. 2009. *Causality: Models, Reasoning and Inference*. 2nd edn. Cambridge
5988 University Press. pages 238
- 5989 Pearson, Karl. 1895. Contributions to the Mathematical Theory of Evolution. II. Skew
5990 Variation in Homogeneous Material. *Philosophical Transactions of the Royal Society*
5991 *A: Mathematical, Physical and Engineering Sciences*, **186**, 343–414. pages 317
- 5992 Pearson, Karl. 1901. On Lines and Planes of Closest Fit to Systems of Points in Space.
5993 *Philosophical Magazine*, **2**(11), 559–572. pages 276
- 5994 Peters, Jonas, Janzing, Dominik, and Schölkopf, Bernhard. 2017. *Elements of Causal*
5995 *Inference: Foundations and Learning Algorithms*. The MIT Press. pages 238
- 5996 Petersen, K. B., and Pedersen, M. S. 2012 (nov). *The Matrix Cookbook*. Version
5997 20121115. pages 151
- 5998 Pollard, David. 2002. *A User's Guide to Measure Theoretic Probability*. Cambridge
5999 University Press. pages 204
- 6000 Polyak, Roman A. 2016. The Legendre Transformation in Modern Optimization. Pages
6001 437–507 of: *Optimization and Its Applications in Control and Data Sciences*. Springer
6002 *Optimization and Its Applications*. pages 224
- 6003 Raschka, Sebastian, and Mirjalili, Wahid. 2017. *Python Machine Learning: Machine*
6004 *Learning and Deep Learning with Python, scikit-learn, and TensorFlow*. Packt Publish-
6005 ing. pages 14
- 6006 Rasmussen, Carl E., and Ghahramani, Zoubin. 2003. Bayesian Monte Carlo. Pages
6007 489–496 of: Becker, S., Thrun, S., and Obermayer, K. (eds), *Advances in Neural*
6008 *Information Processing Systems 15*. Cambridge, MA, USA: The MIT Press. pages 247
- 6009 Rasmussen, Carl E., and Williams, Christopher K. I. 2006. *Gaussian Processes for Ma-*
6010 *chine Learning*. Adaptive Computation and Machine Learning. Cambridge, MA,
6011 USA: The MIT Press. pages 87, 192, 205, 274, 275, 334
- 6012 Rasmussen, Carl Edward, and Ghahramani, Zoubin. 2001. Occam's Razor. Pages 294–
6013 300 of: *Advances in Neural Information Processing Systems 13*. The MIT Press. pages
6014 247
- 6015 Reid, Mark, and Williamson, Robert C. 2011. Information, Divergence and Risk for
6016 Binary Experiments. *Journal of Machine Learning Research*, **12**, 731–817. pages 341
- 6017 Rezende, Danilo Jimenez, and Mohamed, Shakir. 2015. Variational Inference with
6018 Normalizing Flows. In: *ICML*. pages 204
- 6019 Rifkin, Ryan M., and Lippert, Ross A. 2007. Value Regularization and Fenchel Duality.
6020 *Journal of Machine Learning Research*, **8**, 441–479. pages 332, 334
- 6021 Rockafellar, R. Tyrrell. 1970. *Convex Analysis*. Princeton University Press. pages 224
- 6022 Rogers, Simon, and Girolami, Mark. 2016. *A First Course in Machine Learning*. Second
6023 edn. Chapman and Hall/CRC. pages 13, 315
- 6024 Rosenbaum, Paul R. 2017. *Observation & Experiment: An introduction to causal infer-*
6025 *ence*. Harvard University Press. pages 238
- 6026 Rosenblatt, Murray. 1956. Remarks on Some Nonparametric Estimates of a Density
6027 Function. *The Annals of Mathematical Statistics*, **27**(3), 832–837. pages 317
- 6028 Roweis, Sam, and Ghahramani, Zoubin. 1999. A Unifying Review of Linear Gaussian
6029 Models. *Neural Computation*, **11**(2), 305–345. pages 192, 316
- 6030 Roweis, Sam T. 1998. EM algorithms for PCA and SPCA. Pages 626–632 of: *Advances*
6031 *in Neural Information Processing Systems*. pages 295
- 6032 Rumelhart, David E., Hinton, Geoffrey E., and Williams, Ronald J. 1986. Learning
6033 Representations by Back-propagating Errors. *Nature*, **323**(6088), 533–536. pages
6034 152, 154, 210
- 6035 Saitoh, Saburo. 1988. *Theory of Reproducing Kernels and its Applications*. Longman
6036 Scientific & Technical. pages 338, 341

- 6037 Schölkopf, Bernhard, and Smola, Alexander J. 2002. *Learning with Kernels—Support*
6038 *Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation
6039 and Machine Learning. Cambridge, MA, USA: The MIT Press. pages 87, 275, 296,
6040 323, 337, 341
- 6041 Schölkopf, Bernhard, Smola, Alexander J., and Müller, Klaus-Robert. 1998. Nonlinear
6042 Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, **10**(5),
6043 1299–1319. pages 87
- 6044 Schwartz, Laurent. 1964. Sous espaces Hilbertiens d’espaces vectoriels topologiques
6045 et noyaux associés. *Journal d’Analyse Mathématique*, **13**, 115–256. in French. pages
6046 341
- 6047 Schwarz, Gideon E. 1978. Estimating the Dimension of a Model. *Annals of Statistics*,
6048 **6**(2), 461–464. pages 248
- 6049 Shalev-Shwartz, Shai, and Ben-David, Shai. 2014. *Understanding Machine Learning:*
6050 *From Theory to Algorithms*. Cambridge University Press. pages 14, 168, 324
- 6051 Shawe-Taylor, John, and Cristianini, Nello. 2004. *Kernel Methods for Pattern Analysis*.
6052 Cambridge University Press. pages 337, 341
- 6053 Shawe-Taylor, John, and Sun, Shiliang. 2011. A review of optimization methodologies
6054 in support vector machines. *Neurocomputing*, **74**(17), 3609–3618. pages 341
- 6055 Shental, O., Bickson, D., P. H. Siegel and, J. K. Wolf, and Dolev, D. 2008. Gaussian
6056 Belief Propagation Solver for Systems of Linear Equations. In: *IEEE International*
6057 *Symposium on Information Theory*. pages 233
- 6058 Shewchuk, Jonathan Richard. 1994 (August). *An Introduction to the Conjugate Gradient*
6059 *Method Without the Agonizing Pain*. Carnegie Mellon University, Edition 1 1/4.
6060 pages 224
- 6061 Shi, Qinfeng, Petterson, James, Dror, Gideon, Langford, John, Smola, Alex, and Vish-
6062 wanathan, S.V.N. 2009. Hash Kernels for Structured Data. *Journal of Machine*
6063 *Learning Research*, 2615–2637. pages 338
- 6064 Shiryaev, A. N. 1984. *Probability*. Springer-Verlag. pages 204
- 6065 Shor, Naum Z. 1985. *Minimization Methods for Non-differentiable Functions*. Springer.
6066 pages 224
- 6067 Shotton, Jamie, Winn, John, Rother, Carsten, and Criminisi, Antonio. 2006. Texton-
6068 Boost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recog-
6069 nition and Segmentation. In: *Proceedings of the European Conference on Computer*
6070 *Vision*. pages 233
- 6071 Spiegelhalter, David, and Smith, A. F. M. 1980. Bayes Factors and Choice Criteria for
6072 Linear Models. *Journal of the Royal Statistical Society B*, **42**(2), 213–220. pages 245
- 6073 Steinwart, Ingo, and Christmann, Andreas. 2008. *Support Vector Machines*. Springer-
6074 Verlag. pages 332
- 6075 Stoer, Josef, and Burlirsch, Roland. 2002. *Introduction to Numerical Analysis*. Springer.
6076 pages 87, 247
- 6077 Strang, Gilbert. 1993. The fundamental theorem of linear algebra. *The American*
6078 *Mathematical Monthly*, **100**(9), 848–855. pages 115, 123
- 6079 Strang, Gilbert. 2003. *Introduction to Linear Algebra*. 3rd edn. Wellesley-Cambridge
6080 Press. pages 16, 83
- 6081 Strang, Gilbert, Strang, Gilbert, Strang, Gilbert, and Strang, Gilbert. 1993. *Introduction*
6082 *to linear algebra*. Vol. 3. Wellesley-Cambridge Press Wellesley, MA. pages 123
- 6083 Sucar, Luis E., and Gillies, Duncan F. 1994. Probabilistic Reasoning in High-Level
6084 Vision. *Image and Vision Computing*, **12**(1), 42–60. pages 233
- 6085 Szeliski, Richard, Zabih, Ramin, Scharstein, Daniel, Veksler, Olga, Kolmogorov,
6086 Vladimir, Agarwala, Aseem, Tappen, Marshall, and Rother, Carsten. 2008. A Com-
6087 parative Study of Energy Minimization Methods for Markov Random Fields with
6088 Smoothness-based Priors. *IEEE Transactions on Pattern Analysis and Machine Intelli-*
6089 *gence*, **30**(6), 1068–1080. pages 233

- 6090 Tibshirani, Robert. 1996. Regression Selection and Shrinkage via the Lasso. *Journal*
6091 *of the Royal Statistical Society B*, **58**(1), 267–288. pages 259, 275
- 6092 Tipping, Michael E., and Bishop, Christopher M. 1999. Probabilistic Principal Com-
6093 ponent Analysis. *Journal of the Royal Statistical Society: Series B*, **61**(3), 611–622.
6094 pages 192, 295
- 6095 Titsias, Michalis K., and Lawrence, Neil D. 2010. Bayesian Gaussian Process Latent
6096 Variable Model. Pages 844–851 of: Teh, Y. W., and Titterton, D. M. (eds), *Pro-*
6097 *ceedings of the International Conference on Artificial Intelligence and Statistics*. JMLR
6098 W&CP, vol. 9. pages 296
- 6099 Toussaint, Marc. 2012. *Some Notes on Gradient Descent*. pages 209
- 6100 Vapnik, Vladimir. 2000. *The Nature of Statistical Learning Theory*. Springer Verlag.
6101 pages 323
- 6102 Vapnik, Vladimir N. 1998. *Statistical Learning Theory*. Wiley. pages 241
- 6103 Vapnik, Vladimir N. 1999. An Overview of Statistical Learning Theory. *IEEE Transac-*
6104 *tions on Neural Networks*, **10**(5), 988–999. pages 241
- 6105 Vishwanathan, S.V.N., Schraudolph, Nicol N., Kondor, Risi, and Borgwardt, Karsten M.
6106 2010. Graph Kernels. **11**, 1201–1242. pages 338
- 6107 von Luxburg, Ulrike, and Schölkopf, Bernhard. 2011. *Statistical Learning Theory: Mod-*
6108 *els, Concepts, and Results*. Vol. 10. Amsterdam, Netherlands: Elsevier North Holland.
6109 Pages 651–706. pages 241
- 6110 Wahba, Grace. 1990. *Spline Models for Observational Data*. Society for Industrial and
6111 Applied Mathematics. pages 341
- 6112 Wasserman, Larry. 2004. *All of Statistics*. Springer. pages 185
- 6113 Wasserman, Larry. 2007. *All of Nonparametric Statistics*. Springer Verlag. pages 201
- 6114 Wickham, Hadley. 2014. Tidy Data. *Journal of Statistical Software*, **59**. pages 10
- 6115 Williams, Christopher K. I. 1997. Computing with Infinite Networks. In: *Neural Infor-*
6116 *mation Processing Systems*. pages 275
- 6117 Zia, Royce K. P., Redish, Edward F., and McKay, Susan R. 2009. Making sense of the
6118 Legendre transform. *American Journal of Physics*, **77**(614). pages 224