References

[1] M. Aizerman, E. Braverman, and L. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, **25** (1964): 821–837.

[2] S. Akaho. A kernel method for canonical correlation analysis. In *Proceedings of the International Meeting of the Psychometric Society (IMPS2001)*. Springer-Verlag, 2001.

[3] J.K. Anlauf and M. Biehl. The adatron: an adaptive perceptron algorithm. *Europhys. Letters*, **10** (1989): 687–692.

[4] M. Anthony and P. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.

[5] M. Anthony and N. Biggs. *Computational Learning Theory*, volume 30 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1992.

[6] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, **68** (1950): 337–404.

[7] F.R. Bach and M.I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, **3** (2002): 1–48.

[8] P.L. Bartlett. The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE Transactions on Information Theory*, **44**(2) (1998): 525–536.

[9] P.L. Bartlett, S. Boucheron, and G. Lugosi. Model selection and error estimation. *Machine Learning*, **48** (2002): 85–113.

[10] P.L. Bartlett and S. Mendelson. Rademacher and Gaussian complexities: risk bounds and structural results. *Journal of Machine Learning Research*, **3** (2002): 463–482.

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

[12] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*. MIT Press, 2002, pp. 585–591.

[13] K.P. Bennett and O.L. Mangasarian. Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, **1** (1992): 23–34.

[14] C.M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.

[15] M. Borga, T. Landelius, and H. Knutsson. A Unified Approach to PCA, PLS, MLR and CCA. Report LiTH-ISY-R-1992, ISY, SE-581 83 Linköping, Sweden, November 1997.

[16] B.E. Boser, I.M. Guyon, and V.N. Vapnik. A training algorithm for optimal margin classifiers. In D. Haussler, editor, *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory (COLT)* ACM Press, July 1992, pp. 144–152.

[17] S. Boucheron, G. Lugosi, and P. Massart. A sharp concentration inequality with applications. *Random Structures and Algorithms*, **16** (2000):277–292, 2000.

[18] O. Bousquet and A. Elisseeff. Algorithmic stability and generalization performance. In T.K. Leen, T.G. Dietterich, and T., Volker editors, *Advances in Neural Information Processing Systems 13* MIT Press, 2001, pp. 196–202.

[19] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth International, 1984.

[20] C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, **2**(2) (1998): 121–167.

[21] C.J.C. Burges and V. Vapnik. A new method for constructing artificial neural networks: Interim technical report, ONR contract N00014-94-C-0186. Technical report, AT&T Bell Laboratories, 1995.

[22] G. Chaitin. *Algorithmic Information Theory*. Cambridge University Press, 1987.

[23] M. Collins and N. Duffy. Convolution kernels for natural language. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*. MIT Press, 2002, pp. 625–632.

[24] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, 2001, 2nd edition.

[25] C. Cortes, P. Haffner, and M. Mohri. Positive definite rational kernels. In *Proceedings of The 16th Annual Conference on Computational Learning Theory (COLT 2003)*, volume 2777 of *Lecture Notes in Computer Science*. Springer-Verlag, 2003, pp. 41–56.

[26] C. Cortes, P. Haffner, and M. Mohri. Rational kernels. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *Advances in Neural Information Processing Systems 15*. MIT Press, 2003.

[27] C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, **20** (1995): 273–297.

[28] T.M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Elect. Comp.*, **14** (1965): 326–334.

[29] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.

[30] K. Crammer and Y. Singer. Pranking with ranking. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*. MIT Press, 2002, pp. 641–647.

[31] N. Cristianini, H. Lodhi, and J. Shawe-Taylor. Latent semantic kernels. *Journal of Intelligent Information Systems*, **18** (2002): 127–152.

[32] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge: Cambridge University Press, 2000.

[33] N. Cristianini, J. Shawe-Taylor, A. Elisseeff, and J. Kandola. On kernel-target alignment. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*. MIT Press, 2002.

[34] N. Cristianini, J. Shawe-Taylor, and J. Kandola. Spectral kernel methods for clustering. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*. MIT Press, 2002, pp. 649–655.

[35] N. Cristianini, J. Shawe-Taylor, and H. Lodhi. Latent semantic kernels. In Carla Brodley and Andrea Dany In *Proceedings of the 18th International Conference on Machine Learning (ICML 2001)*. Morgan Kaufmann, 2001, pp. 66–73.

[36] M. Cuturi and J.-P. Vert. A covariance kernel for proteins. Technical report, 2003. http://arxiv.org/abs/q-bio.GN/0310022.

[37] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, **41** (1990): 391–407

[38] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Number 31 in Applications of mathematics. Springer-Verlag, 1996.

[39] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.

[40] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley-Interscience, 2000, 2nd edition.

[41] S.T. Dumais, J. Platt, D. Heckerman, and M. Sahami. Inductive learning algorithms and representations for text categorisation. In *Proceedings of the Seventh International Conference on Information and Knowledge Management*, 1998, pp. 148–155 (see www.csee.umbc.edu/cikm/1998#announce).

[42] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

[43] T. Evgeniou, M. Pontil, and T. Poggio. Regularization networks and support vector machines. *Advances in Computational Mathematics*, **13**(1) (2000): 1–50.

[44] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**(2) (1986): 179–188.

[45] K.S. Fu. *Syntactic Methods in Pattern Recognition*. Academic Press, 1974.

[46] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990, 2nd edition.

[47] F. Girosi. An equivalence between sparse approximation and support vector machines. *Neural Computation*, **10**(6) (1998): 1455–1480.

[48] F. Girosi, M. Jones, and T. Poggio. Regularization theory and neural networks architectures. *Neural Computation*, **7**(2) 1995: 219–269.

[49] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 1996.

[50] D. Gusfield. *Algorithms on Strings, Trees, and Sequences*. Cambridge University Press, 1997.

[51] D.R. Hardoon, and J. Shawe-Taylor. KCCA for different level precision in content-based image retrieval. In *Proceedings of Third International Workshop on Content-Based Multimedia Indexing*, 2003, IRISA, Rennes, France.

[52] D.R. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis; an overview with application to learning methods. Technical Report CSD-TR-03-02, Royal Holloway University of London, 2003.

[53] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning – Data Mining, Inference and Prediction*. Springer-Verlag, 2001.

[54] D. Haussler. Convolution kernels on discrete structures. Technical Report UCSC-CRL-99-10, University of California in Santa Cruz, Computer Science Department, July 1999.

[55] R. Herbrich. *Learning Kernel Classifiers: Theory and Algorithms*. MIT Press, 2001.

[56] R. Herbrich, T. Graepel, and K. Obermayer. Large margin rank boundaries for ordinal regression. In *Advances in Large Margin Classifiers*. MIT Press, 2000, pp. 115–132.

[57] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, 1991.

[58] A.E. Hoerl and R.W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**(1) (1970): 55–67.

[59] T. Hofmann. Learning the similarity of documents: an information-geometric approach to document retrieval and categorization. In *Proceedings of Advances in Neural Information Processing Systems (NIPS'99)*, volume 12. MIT Press, 2000.

[60] A. Höskuldsson. Pls regression methods. *Journal of Chemometrics*, **2** (1998): 211–228.

[61] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, **24** (1933): 417–441 and 498–520.

[62] H. Hotelling. Relations between two sets of variables. *Biometrika*, **28** (1936): 321–377.

[63] T.S. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In *Advances in Neural Information Processing Systems 11, 1998*. MIT Press, 1998.

[64] T.S. Jaakkola and D. Haussler. Probabilistic kernel regression models. In *Proceedings of the 1999 Conference on AI and Statistics*. Morgan Kauffman, 1999.

[65] T. Jebara and Risi Kondor. Bhattacharyya and expected likelihood kernels. In *Proceedings of the 16th Conference on Learning Theory (COLT)*. Association for computing Machinery (ACM), 2003.

[66] F. Jiang and M. Littman. Approximate dimension equalization in vector-based information retrieval. In *Proceedings of XXVII International Conference on Machine Learning*. Morgan Kauffman, 2000.

[67] T. Joachims. Text categorization with support vector machines. In *European Conference on Machine Learning (ECML)*. Springer-Verlag, 1998.

[68] T. Joachims. *Learning to Classify Text using Support Vector Machines*. Kluwer, 2002.

[69] T. Joachims, N. Cristianini, and J. Shawe-Taylor. Composite kernels for hypertext categorisation. In *Proceedings of the International Conference on Machine Learning, ICML'01*. Morgan Kauffman, 2001.

[70] M.I. Jordan. *An Introduction to Probabilistic Graphical Models*. Forthcoming, 2004.

[71] J. Kandola, J. Shawe-Taylor, and N. Cristianini. Learning semantic similarity. In *Neural Information Processing Systems 15 (NIPS 15)*. MIT Press, 2002.

[72] J. Kandola, J. Shawe-Taylor, and N. Cristianini. On the extensions of kernel alignment. NeuroCOLT Technical Report NC-TR-2002-120, http://www.neurocolt.org, 2002.

[73] J. Kandola, J. Shawe-Taylor, and N. Cristianini. Optimizing kernel alignment over combinations of kernel. NeuroCOLT Technical Report NC-TR-2002-121, http://www.neurocolt.org, 2002.

[74] R. Kannan, S. Vempala, and A. Vetta. On clusterings: good, bad and spectral. In *Proc. of the 41st Foundations of Computer Science (FOCS '00)*. Morgan Kauffman, 2000.

[75] H. Kashima, K. Tsuda, and A. Inokuchi. Marginalized kernels between labeled graphs. In *Proceedings of the 20th International Conference on Machine Learning*. Morgan Kauffman, 2003.

[76] M. Kearns and U. Vazirani. *An Introduction to Computational Learning Theory*. MIT Press, 1994.

[77] T. Kin, K. Tsuda, and K. Asai. Marginalized kernels for RNA sequence data analysis. In R. H. Lathtop, K. Nakai, S. Miyano, T. Takagi, and M. Kanehisa, editors, *Genome Informatics 2002*. Universal Academic Press, 2002, pp. 112–132.

[78] A. Koestler. *The Sleepwalkers*. Penguin, 1979.

[79] V. Koltchinskii. Rademacher penalties and structural risk minimization. *IEEE Transactions on Information Theory*, **47**(5) (July 2001): 1902–1914.

[80] V. Koltchinskii and D. Panchenko. Rademacher processes and bounding the risk of function learning. *High Dimensional Probability II*, 2000, pp. 443–459.

[81] Vladimir I. Koltchinskii and Dmitry Panchenko. Empirical margin distributions and bounding the generalization error of combined classifiers. *Annals of Statistics*, **30**(1), (2002).

[82] R.I. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete structures. In *Proceedings of Intenational Conference on Machine Learning (ICML 2002)*. Morgan, Kauffman 2002.

[83] P.L. Lai and C. Fyfe. Kernel and nonlinear canonical correlation analysis. *International Journal of Neural Systems*, **10**(5) (2000): 365–377.

[84] J. Langford and J. Shawe-Taylor. PAC Bayes and margins. In *Advances in Neural Information Processing Systems 15*. MIT Press, 2003.

[85] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer-Verlag, 1991.

[86] E. Leopold and J. Kinderman. Text categorization with support vector machines. How to represent texts in input space? *Machine Learning: Special Issue on Support Vector and Kernel Methods*, **46** (1–3) (2002): 423–444.

[87] C. Leslie, E. Eskin, A. Cohen, J. Weston, and William Stafford Noble. Mismatch string kernels for discriminative protein classification. *Bioinformatics*, 2003. to appear.

[88] C. Leslie, E. Eskin, and W.S. Noble. The spectrum kernel: A string kernel for SVM protein classification. In *Proceedings of the Pacific Symposium on Biocomputing (PSB-2002)*. World Scientific Publishing, 2002.

[89] C. Leslie, E. Eskin, J. Weston, and W.S. Noble. Mismatch string kernels for SVM protein classification. In *Advances in Neural Information Processing Systems 15*. MIT Press, 2003.

[90] C. Leslie and Rui Kuang. Fast kernels for inexact string matching. In *Proceedings of the 16th Conference on Learning Theory (COLT)*. ACM, 2003.

[91] M. Li and P. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, 1997.

[92] M.L. Littman, S.T. Dumais, and T.K. Landauer. Automatic cross-language information retrieval using latent semantic indexing. In G. Grefenstette, editor, *Cross Language Information Retrieval*. Kluwer, 1998.

[93] H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins. Text classification using string kernels. *Journal of Machine Learning Research*, **2** (2002): 419–444.

[94] O.L. Mangasarian. *Nonlinear Programming*. SIAM, 1994.

[95] O.L. Mangasarian. Linear and nonlinear separation of patterns by linear programming. *Operations Research*, **13** (1965): 444–452.

[96] C. McDiarmid. On the method of bounded differences. In 141 London Mathematical Society Lecture Notes Series, editor, *Surveys in Combinatorics 1989*, Cambridge University Press, 1989, pp. 148–188.

[97] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London*, Series A **209** (1909): 415–446.

[98] C. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.

[99] C.A. Micchelli. Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constructive Approximation*, 2 (1986): 11–22.

[100] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller. Fisher discriminant analysis with kernels. In Y.-H. Hu, J. Larsen. E. Wilson, and S. Douglas, editors, *Neural Networks for Signal Processing IX*. IEEE, 1999, pp. 41–48.

[101] M. Minsky and S. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, 1969.

[102] T. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[103] A. Ng, M.I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*. MIT Press, 2002.

[104] C.S. Ong, A. Smola, and Robert C. Williamson. Superkernels. In *Advances in Neural Information Processing Systems (NIPS 02)*. MIT Press, 2003.

[105] T. Pisanski and J. Shawe-Taylor. Characterising graph drawing with eigenvectors. *Journal of Chemical Information and Computer Sciences*, **40**(3) (2000): 567–571.

[106] T. Poggio. On optimal nonlinear associative recall. *Biological Cybernetics*, **19** (1975): 201–209.

[107] T. Poggio and F. Girosi. Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, **247** (1990): 978–982.

[108] T. Poggio and S. Smale. The mathematics of learning: dealing with data. *American Mathematical Society*, **50**(5) (2003): 537–544.

[109] J.R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kauffmann, 1993.

[110] B.D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, 1996.

[111] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, **65**(6) (1958): 386–408.

[112] R. Rosipal and L.J. Trejo. Kernel partial least squares regression in reproducing kernel Hilbert space. *Journal of Machine Learning Research*, **2** (2001): 97–123.

[113] S. Saitoh. *Theory of Reproducing Kernels and its Applications*. Longman Scientific & Technical, 1988.

[114] G. Salton, A. Wang, and C. Yang. A vector space model for information retrieval. *Journal of the American Society for Information Science*, **18** (1975): 613–620.

[115] C. Saunders, A. Gammermann and V. Vovk. Ridge regression learning algorithm in dual variables. In J. Shavlik, editor, *Machine Learning: Proceedings of the Fifteenth International Conference(ICML '98)*. Morgan Kaufmann, 1998.

[116] C. Saunders, J. Shawe-Taylor, and A. Vinokourov. String kernels, Fisher kernels and finite state automata. In *Advances of Neural Information Processing Systems 15*. 2002.

[117] B. Schölkopf, C. Burges, and V. Vapnik. Extracting support data for a given task. In U.M. Fayyad and R. Uthurusamy, editors, *Proceedings, First International Conference on Knowledge Discovery & Data Mining*. AAAI Press, 1995.

[118] B. Schölkopf. *Support Vector Learning*. R. Oldenbourg Verlag, 1997.

[119] B. Schölkopf, J.C. Platt, J.S. Shawe-Taylor, A.J. Smola, and R.C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, **13**(7) (2001): 1443–1471.

[120] B. Schölkopf and A. Smola. *Learning with Kernels*. MIT Press, 2002.

[121] B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, **10** (1998): 1299–1319.

[122] B. Schölkopf, A. Smola, R. Williamson, and P. Bartlett. New support vector algorithms. *Neural Computation*, **12**(5) (2000): 1207–1245.

[123] A. Shashua. On the relationship between the support vector machine for classification and sparsified fisher's linear discriminant. *Neural Processing Letters*, **9**(2) (April 1999): 129-139.

[124] J. Shawe-Taylor, P.L. Bartlett, R.C. Williamson, and M. Anthony. Structural risk minimization over data-dependent hierarchies. *IEEE Transactions on Information Theory*, **44**(5) (1998): 1926–1940.

[125] J. Shawe-Taylor and N. Cristianini. On the generalisation of soft margin algorithms. *IEEE Transactions on Information Theory*, **48**(10) (2002): 2721–2735.
[126] J. Shawe-Taylor, N. Cristianini, and J. Kandola. On the concentration of spectral properties. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*. MIT Press, 2002, pp. 511–517.
[127] J. Shawe-Taylor, C. Williams, N. Cristianini, and J.S. Kandola. On the eigenspectrum of the Gram matrix and its relationship to the operator eigenspectrum. In *Proceedings of the 13th International Conference on Algorithmic Learning Theory (ALT2002)*, volume 2533, Springer-Verlag, 2002, pp. 23–40.
[128] G. Siolas and F. d'Alché Buc. Support vector machines based on a semantic kernel for text categorization. In *International Joint Conference on Neural Networks (1JCNN)*. Springer-Verlag, 2002.
[129] F.W. Smith. Pattern classifier design by linear programming. *IEEE Transactions on Computers*, **C-17** (1968): 367–372.
[130] A. Smola, B. Schölkopf, and K.-R. Müller. The connection between regularization operators and support vector kernels. *Neural Networks*, **11** (1998): 637–649.
[131] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific Publishing Co., 2002.
[132] E. Takimoto and M. Warmuth. Path kernels and multiplicative updates. In *Proceedings of the 15th Conference on Learning Theory (COLT)*. ACM, 2002.
[133] E. Takimoto and M. Warmuth. Predicting nearly as well as the best pruning of a planar decision graph. *Theoretical Computer Science*, **288**(2) (2002): 217–235.
[134] M. Talagrand. Sharper bounds for Gaussian and empirical processes. *Annals of Probability*, **22** (1994): 28–76.
[135] M. Talagrand. The Glivenko–Cantelli problem, ten years later. *Journal of Theoretical Probability*, **9**(2) (1996): 371–384.
[136] D.M.J. Tax and R.P.W. Duin. Support vector domain description. *Pattern Recognition Letters*, pages **20**(11–13) (December 1999): 1191–1199.
[137] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 2003.
[138] A.N. Tikhonov and V.Y. Arsenin. *Solutions of Ill-posed Problems*. W.H. Winston, 1977.
[139] K. Tsuda, T. Kin, and K. Asai. Marginalized kernels for biological sequences. *Bioinformatics*, **18** (2002): 268–275.
[140] V. Vapnik. *Estimation of Dependences Based on Empirical Data [in Russian]*. Nauka, Moscow, 1979. (English translation, Springer Verlag, 1982).
[141] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.
[142] V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
[143] V. Vapnik and A. Chervonenkis. A note on one class of perceptrons. *Automation and Remote Control*, **25** (1964).
[144] V. Vapnik and A. Chervonenkis. Uniform convergence of frequencies of occurence of events to their probabilities. *Doklady Akademii Nauk SSSR*, **181** (1968): 915–918.

[145] V. Vapnik and A. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, **16**(2) (1971): 264–280.

[146] V. Vapnik and A. Lerner. Pattern recognition using generalized portrait method. *Automation and Remote Control*, **24** (1963).

[147] J.-P. Vert. Support vector machine prediction of signal peptide cleavage site using a new class of kernels for strings. In R.B. Altman, A.K. Dunker, L. Hunter, K. Lauerdale, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing 2002*. World Scientific, 2002, pp. 649–660.

[148] J.-P. Vert. A tree kernel to analyze phylogenetic profiles. *Bioinformatics*, **18** (2002): 276–284.

[149] J.-P. Vert and M. Kanehisa. Graph-driven features extraction from microarray data using diffusion kernels and kernel CCA. In *Advances in Neural Information Processing Systems 15*. MIT Press, 2003.

[150] J.-P. Vert, H. Saigo, and T. Akutsu. Convolution and local alignment kernels. In B. Schölkopf, K. Tsuda, and J.-P. Vert, editors, *Kernel Methods in Computational Biology*. MIT Press, 2004.

[151] H.D. Vinod. Canonical ridge and econometrics of joint production. *J. Econometrics*, **4** (1976): 147–166.

[152] A. Vinokourov, D. Hardoon, J. Shawe-Taylor. Learning the semantics of multimedia content with application to web image retrieval and classification. In *Proceedings of Fourth International Symposium on Independent Component Analysis and Blind Source Separation*, Nara, 2003.

[153] A. Vinokourov, J. Shawe-Taylor, and N. Cristianini. Inferring a semantic representation of text via cross-language correlation analysis. In *Advances of Neural Information Processing Systems 15*. MIT Press, 2002.

[154] S.V.N. Vishwanathan and A.J. Smola. Fast kernels on strings and trees. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*. MIT Press, 2002.

[155] G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, 1990.

[156] G. Wahba. Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV. In B. Schölkopf, C.J.C. Burges, and A.J. Smola, editors, *Advances in Kernel Methods – Support Vector Learning*. MIT Press, pp. 69–88.

[157] C. Watkins. Kernels from matching operations. Technical Report CSD-TR-98-1?, Royal Holloway College, Computer Science Department, July 1999.

[158] C. Watkins. Dynamic alignment kernels. In A.J. Smola, P.L. Bartlett, B. Schölkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 2002, pp. 39–50.

[159] D.W. Wichern and R.A. Johnson. *Applied Multivariate Statistical Analysis*. Prentice Hall, 5th edition, 2002.

[160] B. Widrow and M. Hoff. Adaptive switching circuits. In *IRE WESCON Convention Record*, (1960), pp. 96–104.

[161] C.K.I. Williams. Prediction with Gaussian processes: From linear regression to linear prediction and beyond. In M.I. Jordan, editor, *Learning and Inference in Graphical Models*. Kluwer, 1998.

[162] H. Wold. Estimation of principal components and related models by iterative least squares. In P.R. Krishnaiah, editor, *Multivariate Analysis*. Academic Press, 1966, pp. 391–420.

[163] H. Wold. Path models with latent variables: the NIPALS approach. In H.M. Blalock *et al.*, editors, *Quantitative Sociology: International Perspectives on Mathematical and Statistical Model Building*. Academic Press, 1975, pp. 307–357.

[164] H. Wold. Soft modeling by latent variables; the nonlinear iterative partial least squares NIPALS approach. *Perspectives in Probability and Statistics*, **9** (1975): 117–142.

[165] H. Wold. Partial least squares. In S. Kotz and N.L. Johnson, editors, *Encyclopedia of the Statistical Sciences*, volume 6. John Wiley & Sons, 1985, pp. 581–591.

[166] H. Wolkowicz, R. Saigal, and editors L. Vandenberghe. *Handbook of Semidefinite Programming*. Kluwer, 2001.

[167] S.K.M. Wong, W. Ziarko, and P.C.N. Wong. Generalized vector space model in information retrieval. In *ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'85)*. 1985, pp. 18–25.

[168] Y. Yamanishi, J.-P. Vert, A. Nakaya, and M. Kanehisa. Extraction of correlated gene clusters from multiple genomic data by generalized kernel canonical correlation analysis. *Bioinformatics*, 2003. to appear.

[169] H. Zha, X. He, C. Ding, M. Gu, and H. Simon. Spectral relaxation for k-means clustering. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*. MIT Press, 2002, pp. 1057–1064.