
References

- [1] M. Aizerman, E. Braverman, and L. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.
- [2] N. Alon, S. Ben-David, N. Cesa-Bianchi, and D. Haussler. Scale-sensitive dimensions, uniform convergence, and learnability. *Journal of the ACM*, 44(4):615–631, 1997.
- [3] S. Amari and S. Wu. Improving support vector machine classifiers by modifying kernel functions. *Neural Networks*, 1999. to appear.
- [4] J. K. Anlauf and M. Biehl. The adatron: an adaptive perceptron algorithm. *Europhysics Letters*, 10:687–692, 1989.
- [5] M. Anthony and P. Bartlett. *Learning in Neural Networks : Theoretical Foundations*. Cambridge University Press, 1999.
- [6] M. Anthony and N. Biggs. *Computational Learning Theory*, volume 30 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1992.
- [7] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.
- [8] S. Arora, L. Babai, J. Stern, and Z. Sweedyk. Hardness of approximate optima in lattices, codes and linear systems. *Journal of Computer and System Sciences*, 54(2):317–331, 1997.
- [9] P. Bartlett and J. Shawe-Taylor. Generalization performance of support vector machines and other pattern classifiers. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in Kernel Methods – Support Vector Learning*, pages 43–54. MIT Press, 1999.
- [10] 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):525–536, 1998.

- [11] M. Bazaraa, D. Sherali, and C. Shetty. *Nonlinear Programming : Theory and Algorithms*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, 1992.
- [12] K. Bennett, N. Cristianini, J. Shawe-Taylor, and D. Wu. Enlarging the margin in perceptron decision trees. *Machine Learning*. to appear.
- [13] K. Bennett and A. Demiriz. Semi-supervised support vector machines. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Advances in Neural Information Processing Systems, 12*, pages 368–374. MIT Press, 1998.
- [14] K. P. Bennett and E. J. Bredensteiner. Geometry in learning. In C. Gorini, E. Hart, W. Meyer, and T. Phillips, editors, *Geometry at Work*. Mathematical Association of America, 1998.
- [15] K. P. Bennett and O. L. Mangasarian. Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, 1:23–34, 1992.
- [16] C. M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.
- [17] V. Blanz, B. Schölkopf, H. Bülthoff, C. Burges, V. Vapnik, and T. Vetter. Comparison of view-based object recognition algorithms using realistic 3D models. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff, editors, *Artificial Neural Networks – ICANN'96*, pages 251 – 256. Springer Lecture Notes in Computer Science, Vol. 1112, 1996.
- [18] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, 1989.
- [19] 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*, pages 144–152. ACM Press, 1992.
- [20] P. S. Bradley, O. L. Mangasarian, and D. R. Musicant. Optimization methods in massive datasets. Technical Report Data Mining Institute TR-99-01, University of Wisconsin in Madison, 1999.
- [21] M. Brown, W. Grundy, D. Lin, N. Cristianini, C. Sugnet, T. Furey, M. Ares, and D. Haussler. Knowledge-based analysis of microarray gene expression data using support vector machines. Technical report, University of California in Santa Cruz, 1999. (submitted for publication).
- [22] M. Brown, W. Grundy, D. Lin, N. Cristianini, C. Sugnet, T. Furey, M. Ares, and D. Haussler. Knowledge-based analysis of microarray gene expression data using support vector machines, 1999. [<http://www.cse.ucsc.edu/research/compbio/genex/genex.html>].

Santa Cruz, University of California, Department of Computer Science and Engineering.

- [23] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
- [24] C. Campbell and N. Cristianini. Simple training algorithms for support vector machines. Technical Report CIG-TR-KA, University of Bristol, Engineering Mathematics, Computational Intelligence Group, 1999.
- [25] O. Chapelle, P. Haffner, and V. Vapnik. SVMs for histogram-based image classification. *IEEE Transaction on Neural Networks*, 1999.
- [26] C. Cortes. *Prediction of Generalization Ability in Learning Machines*. PhD thesis, Department of Computer Science, University of Rochester, USA, 1995.
- [27] C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1995.
- [28] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, 14:326–334, 1965.
- [29] N. Cristianini, C. Campbell, and J. Shawe-Taylor. A multiplicative updating algorithm for training support vector machine. In *Proceedings of the 6th European Symposium on Artificial Neural Networks (ESANN)*, 1999.
- [30] N. Cristianini and J. Shawe-Taylor. An Introduction to Support Vector Machines: the web-site associated with the book, 2000.
- [31] N. Cristianini, J. Shawe-Taylor, and C. Campbell. Dynamically adapting kernels in support vector machines. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Advances in Neural Information Processing Systems*, 11. MIT Press, 1998.
- [32] N. Cristianini, J. Shawe-Taylor, and P. Sykacek. Bayesian classifiers are large margin hyperplanes in a Hilbert space. In J. Shavlik, editor, *Machine Learning: Proceedings of the Fifteenth International Conference*, pages 109–117. Morgan Kaufmann, 1998.
- [33] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Number 31 in Applications of mathematics. Springer, 1996.
- [34] R. Dietrich, M. Opper, and H. Sompolinsky. Statistical mechanics of support vector networks. *Physics Review Letters*, 82:2975, 1999.
- [35] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. Wiley, 1973.

- [36] S. Dumais, J. Platt, D. Heckerman, and M. Sahami. Inductive learning algorithms and representations for text categorization. In *7th International Conference on Information and Knowledge Management*, 1998.
- [37] T. Evgeniou and M. Pontil. On the V_γ dimension for regression in reproducing kernel Hilbert spaces. In *Algorithmic Learning Theory: ALT-99*. Springer-Verlag, 1999.
- [38] T. Evgeniou, M. Pontil, and T. Poggio. Regularization networks and support vector machines. In A.J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [39] T. Evgeniou, M. Pontil, and T. Poggio. A unified framework for regularization networks and support vector machines. Technical Report CBCL Paper #171/AI Memo #1654, Massachusetts Institute of Technology, 1999.
- [40] R. Fisher. *Contributions to Mathematical Statistics*. Wiley, 1952.
- [41] R. Fletcher. *Practical methods of Optimization*. Wiley, 1988.
- [42] S. Floyd and M. Warmuth. Sample compression, learnability, and the Vapnik-Chervonenkis dimension. *Machine Learning*, 21(3):269–304, 1995.
- [43] Y. Freund and R.E. Schapire. Large margin classification using the perceptron algorithm. In J. Shavlik, editor, *Machine Learning: Proceedings of the Fifteenth International Conference*. Morgan Kaufmann, 1998.
- [44] T. Friess, N. Cristianini, and C. Campbell. The kernel-Adatron: a fast and simple training procedure for support vector machines. In J. Shavlik, editor, *Machine Learning: Proceedings of the Fifteenth International Conference*. Morgan Kaufmann, 1998.
- [45] T. Friess and R. Harrison. Support vector neural networks: the kernel adatron with bias and soft margin. Technical Report ACSE-TR-752, University of Sheffield, Department of ACSE, 1998.
- [46] T. Friess and R. Harrison. A kernel based adaline. In *Proceedings of the 6th European Symposium on Artificial Neural Networks (ESANN)*, 1999.
- [47] S. I. Gallant. Perceptron based learning algorithms. *IEEE Transactions on Neural Networks*, 1:179–191, 1990.
- [48] C. Gentile and M. K. Warmuth. Linear hinge loss and average margin. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Advances in Neural Information Processing Systems 11*. MIT Press, 1999.
- [49] M. Gibbs and D. MacKay. Efficient implementation of Gaussian processes. Technical report, Department of Physics, Cavendish Laboratory, Cambridge University, UK, 1997.

- [50] M. N. Gibbs. *Bayesian Gaussian Processes for Regression and Classification*. PhD thesis, University of Cambridge, 1997.
- [51] F. Girosi. An equivalence between sparse approximation and support vector machines. *Neural Computation*, 10(6):1455–1480, 1998.
- [52] F. Girosi, M. Jones, and T. Poggio. Regularization theory and neural networks architectures. *Neural Computation*, 7(2):219–269, 1995.
- [53] GMD-FIRST. GMD-FIRST web site on Support Vector Machines. <http://svm.first.gmd.de>.
- [54] L. Gurvits. A note on a scale-sensitive dimension of linear bounded functionals in Banach spaces. In *Proceedings of Algorithmic Learning Theory, ALT-97*, 1997.
- [55] I. Guyon, N. Matić, and V. Vapnik. Discovering informative patterns and data cleaning. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smythand, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, pages 181–203. MIT Press, 1996.
- [56] I. Guyon and D. Stork. Linear discriminant and support vector classifiers. In A.J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [57] M. H. Hassoun. *Optical Threshold Gates and Logical Signal Processing*. PhD thesis, Department of ECE, Wayne State University, Detroit, USA, 1986.
- [58] D. Haussler. Convolution kernels on discrete structures. Technical Report UCSC-CRL-99-10, University of California in Santa Cruz, Computer Science Department, July 1999.
- [59] R. Herbrich, T. Graepel, and C. Campbell. Bayes point machines: Estimating the Bayes point in kernel space. In *Proceedings of IJCAI Workshop Support Vector Machines*, 1999.
- [60] R. Herbrich, K. Obermayer, and T. Graepel. Large margin rank boundaries for ordinal regression. In A.J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [61] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, 1952.
- [62] C. Hildreth. A quadratic programming procedure. *Naval Research Logistics Quarterly*, 4:79–85, 1957.
- [63] A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

- [64] K. U. Höffgen, K. S. van Horn, and H. U. Simon. Robust trainability of single neurons. *Journal of Computer and System Sciences*, 50(1):114–125, 1995.
- [65] T. S. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Advances in Neural Information Processing Systems, 11*. MIT Press, 1998.
- [66] T. S. Jaakkola and D. Haussler. Probabilistic kernel regression models. In *Proceedings of the 1999 Conference on AI and Statistics*, 1999.
- [67] T. Joachims. Text categorization with support vector machines. In *Proceedings of European Conference on Machine Learning (ECML)*, 1998.
- [68] T. Joachims. Making large-scale SVM learning practical. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in Kernel Methods – Support Vector Learning*, pages 169–184. MIT Press, 1999.
- [69] W. Karush. *Minima of Functions of Several Variables with Inequalities as Side Constraints*. Department of Mathematics, University of Chicago, 1939. MSc Thesis.
- [70] L. Kaufmann. Solving the quadratic programming problem arising in support vector classification. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in Kernel Methods – Support Vector Learning*, pages 147–168. MIT Press, 1999.
- [71] M. Kearns and U. Vazirani. *An Introduction to Computational Learning Theory*. MIT Press, 1994.
- [72] M. J. Kearns and R. E. Schapire. Efficient distribution-free learning of probabilistic concepts. *Journal of Computer and System Science*, 48(3):464–497, 1994. Earlier version appeared in FOCS90.
- [73] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy. A fast iterative nearest point algorithm for support vector machine classifier design. Technical report, Department of CSA, IISc, Bangalore, India, 1999. Technical Report No. TR-ISL-99-03.
- [74] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy. Improvements to Platt’s SMO algorithm for SVM classifier design. Technical report, Control Division, Department of Mechanical and Production Engineering, National University of Singapore, 1999. Technical Report No. CD-99-14.
- [75] J. Kivinen and M. K. Warmuth. Additive versus exponentiated gradient for linear prediction. *Information and Computation*, 132:1–64, 1997.
- [76] A. Kowalczyk. Maximal margin perceptron. In A.J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 1999.

- [77] H. Kuhn and A. Tucker. Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium on Mathematical Statistics and Probabilistics*, pages 481–492. University of California Press, 1951.
- [78] Y. LeCun, L. D. Jackel, L. Bottou, A. Brunot, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. A. Müller, E. Säkingen, P. Simard, and V. Vapnik. Comparison of learning algorithms for handwritten digit recognition. In F. Fogelman-Soulié and P. Gallinari, editors, *Proceedings ICANN'95 – International Conference on Artificial Neural Networks*, volume II, pages 53–60. EC2, 1995.
- [79] N. Littlestone and M. Warmuth. Relating data compression and learnability. Technical report, University of California, Santa Cruz, 1986.
- [80] D. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, 1984.
- [81] D. MacKay. Introduction to Gaussian processes. In *Neural Networks and Machine Learning (NATO Asi Series)*; Ed. by Chris Bishop, 1999.
- [82] D. J. C. MacKay. A practical Bayesian framework for backprop networks. *Neural Computation*, 4:448–472, 1992.
- [83] O. Mangasarian. Generalized support vector machines. In A.J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [84] O. L. Mangasarian. Linear and nonlinear separation of patterns by linear programming. *Operations Research*, 13:444–452, 1965.
- [85] O. L. Mangasarian. Multi-surface method of pattern separation. *IEEE Transactions on Information Theory*, IT-14:801–807, 1968.
- [86] O. L. Mangasarian. *Nonlinear Programming*. SIAM, 1994.
- [87] O. L. Mangasarian. Mathematical programming in data mining. *Data Mining and Knowledge Discovery*, 42(1):183–201, 1997.
- [88] O. L. Mangasarian. Generalized support vector machines. Technical Report Mathematical Programming TR 98-14, University of Wisconsin in Madison, 1998.
- [89] O. L. Mangasarian and D. R. Musicant. Successive overrelaxation for support vector machines. Technical Report Mathematical Programming TR 98-18, University of Wisconsin in Madison, 1998.
- [90] O. L. Mangasarian and D. R. Musicant. Data discrimination via nonlinear generalized support vector machines. Technical Report Mathematical Programming TR 99-03, University of Wisconsin in Madison, 1999.

- [91] O. L. Mangasarian and D. R. Musicant. Massive support vector regression. Technical Report Data Mining Institute TR-99-02, University of Wisconsin in Madison, 1999.
- [92] MATLAB. *User's Guide*. The MathWorks, Inc., 1992.
- [93] David A. McAllester. Some PAC-Bayesian theorems. In *Proceedings of the 11th Annual Conference on Computational Learning Theory*, pages 230–234. ACM Press, 1998.
- [94] David A. McAllester. PAC-Bayesian model averaging. In *Proceedings of the 12th Annual Conference on Computational Learning Theory*, pages 164–170. ACM Press, 1999.
- [95] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philos. Trans. Roy. Soc. London, A* 209:415–446, 1909.
- [96] C. J. Merz and P. M. Murphy. UCI repository of machine learning databases, 1998. [<http://www.ics.uci.edu/~mlearn/MLRepository.html>]. Irvine, CA: University of California, Department of Information and Computer Science.
- [97] C. A. Micchelli. Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constructive Approximation*, 2:11–22, 1986.
- [98] M.L. Minsky and S.A. Papert. *Perceptrons*. MIT Press, 1969. Expanded Edition 1990.
- [99] T. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- [100] J. J. More and S. J. Wright. *Optimization Software Guide*. Frontiers in Applied Mathematics, Volume 14. Society for Industrial and Applied Mathematics (SIAM), 1993.
- [101] B. A. Murtagh and M. A. Saunders. MINOS 5.4 user's guide. Technical Report SOL 83.20, Stanford University, 1993.
- [102] R. Neal. *Bayesian Learning in Neural Networks*. Springer Verlag, 1996.
- [103] R. M. Neal. Monte carlo implementation of Gaussian process models for Bayesian regression and classification. Technical Report TR 9702, Department of Statistics, University of Toronto, 1997.
- [104] A. B. Novikoff. On convergence proofs on perceptrons. In *Symposium on the Mathematical Theory of Automata*, volume 12, pages 615–622. Polytechnic Institute of Brooklyn, 1962.

- [105] M. Oppor and W. Kinzel. Physics of generalization. In E. Domany J. L. van Hemmen and K. Schulten, editors, *Physics of Neural Networks III*. Springer Verlag, 1996.
- [106] M. Oppor and F. Vivarelli. General bounds on Bayes errors for regression with Gaussian processes. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Advances in Neural Information Processing Systems, 11*. MIT Press, 1998.
- [107] M. Oppor and O. Winther. Gaussian processes and SVM: Mean field and leave-one-out. In A.J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [108] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. In *Proceedings Computer Vision and Pattern Recognition*, pages 193–199, 1997.
- [109] E. Osuna, R. Freund, and F. Girosi. An improved training algorithm for support vector machines. In J. Principe, L. Gile, N. Morgan, and E. Wilson, editors, *Neural Networks for Signal Processing VII — Proceedings of the 1997 IEEE Workshop*, pages 276–285. IEEE, 1997.
- [110] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: An application to face detection. In *Proceedings of Computer Vision and Pattern Recognition*, pages 130–136, 1997.
- [111] E. Osuna and F. Girosi. Reducing run-time complexity in SVMs. In *Proceedings of the 14th International Conference on Pattern Recognition, Brisbane, Australia*, 1998. To appear.
- [112] J. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in Kernel Methods – Support Vector Learning*, pages 185–208. MIT Press, 1999.
- [113] J. Platt, N. Cristianini, and J. Shawe-Taylor. Large margin DAGs for multiclass classification. In *Neural Information Processing Systems (NIPS 99)*, 1999. to appear.
- [114] J. C. Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. Technical Report MSR-TR-98-14, Microsoft Research, 1998.
- [115] T. Poggio. On optimal nonlinear associative recall. *Biological Cybernetics*, 19:201–209, 1975.
- [116] T. Poggio and F. Girosi. Networks for approximation and learning. *Proceedings of the IEEE*, 78(9), September 1990.
- [117] D. Pollard. *Convergence of Stochastic Processes*. Springer, 1984.

- [118] M. Pontil and A. Verri. Object recognition with support vector machines. *IEEE Trans. on PAMI*, 20:637–646, 1998.
- [119] K. Popper. *The Logic of Scientific Discovery*. Springer, 1934. First English Edition by Hutchinson, 1959.
- [120] C. Rasmussen. *Evaluation of Gaussian Processes and Other Methods for Non-Linear Regression*. PhD thesis, Department of Computer Science, University of Toronto, 1996. <ftp://ftp.cs.toronto.edu/pub/carl/thesis.ps.gz>.
- [121] R. Rifkin, M. Pontil, and A. Verri. A note on support vector machines degeneracy. Department of Mathematical Sciences CBCL Paper #177/AI Memo #1661, Massachusetts Institute of Technology, June 1999.
- [122] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–408, 1959.
- [123] S. Saitoh. *Theory of Reproducing Kernels and its Applications*. Longman Scientific & Technical, 1988.
- [124] A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal on Research and Development*, 49:210–229, 1959.
- [125] 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*. Morgan Kaufmann, 1998.
- [126] C. Saunders, M. O. Stitson, J. Weston, L. Bottou, B. Schölkopf, and A. Smola. Support vector machine – reference manual. Technical Report CSD-TR-98-03, Department of Computer Science, Royal Holloway, University of London, Egham, TW20 0EX, UK, 1998. TR available as http://www.dcs.rhnc.ac.uk/research/compint/areas/comp_learn/sv/pub/report98-03.ps; SVM available at <http://svm.dcs.rhnc.ac.uk/>.
- [127] R. Schapire, Y. Freund, P. Bartlett, and W. Sun Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. *Annals of Statistics*, 26(5):1651–1686, 1998.
- [128] 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.
- [129] B. Schölkopf. *Support Vector Learning*. R. Oldenbourg Verlag, 1997.
- [130] B. Schölkopf, P. Bartlett, A. Smola, and R. Williamson. Shrinking the tube: a new support vector regression algorithm. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Advances in Neural Information Processing Systems, 11*. MIT Press, 1998.

- [131] B. Schölkopf, P. Bartlett, A. Smola, and R. Williamson. Support vector regression with automatic accuracy control. In L. Niklasson, M. Bodén, and T. Ziemke, editors, *Proceedings of the 8th International Conference on Artificial Neural Networks*, Perspectives in Neural Computing, pages 147 – 152. Springer Verlag, 1998.
- [132] B. Schölkopf, C. J. C. Burges, and A. J. Smola. *Advances in Kernel Methods – Support Vector Learning*. MIT Press, 1999.
- [133] B. Schölkopf, J. Shawe-Taylor, A. Smola, and R. Williamson. Generalization bounds via the eigenvalues of the gram matrix. Technical Report NC-TR-1999-035, NeuroCOLT Working Group, <http://www.neurocolt.com>, 1999.
- [134] B. Schölkopf, A. Smola, and K.-R. Müller. Kernel principal component analysis. In W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud, editors, *Artificial Neural Networks – ICANN'97*, pages 583–588. Springer Lecture Notes in Computer Science, Volume 1327, 1997.
- [135] B. Schölkopf, A. Smola, R. Williamson, and P. Bartlett. New support vector algorithms. Technical Report NC-TR-98-031, NeuroCOLT Working Group, <http://www.neurocolt.com>, 1998.
- [136] B. Schölkopf, A. J. Smola, and K. Müller. Kernel principal component analysis. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in Kernel Methods – Support Vector Learning*, pages 327–352. MIT Press, 1999.
- [137] J. Shawe-Taylor. Classification accuracy based on observed margin. *Algorithmica*, 22:157–172, 1998.
- [138] 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):1926–1940, 1998.
- [139] J. Shawe-Taylor and N. Cristianini. Robust bounds on generalization from the margin distribution. Technical Report NC-TR-1998-020, NeuroCOLT Working Group, <http://www.neurocolt.com>, 1998.
- [140] J. Shawe-Taylor and N. Cristianini. Further results on the margin distribution. In *Proceedings of the Conference on Computational Learning Theory, COLT 99*, pages 278–285, 1999.
- [141] J. Shawe-Taylor and N. Cristianini. Margin distribution and soft margin. In A.J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [142] J. Shawe-Taylor and N. Cristianini. Margin distribution bounds on generalization. In *Proceedings of the European Conference on Computational Learning Theory, EuroCOLT'99*, pages 263–273, 1999.

- [143] F. W. Smith. Pattern classifier design by linear programming. *IEEE Transactions on Computers*, C-17:367–372, 1968.
- [144] A. Smola and B. Schölkopf. On a kernel-based method for pattern recognition, regression, approximation and operator inversion. *Algorithmica*, 22:211–231, 1998.
- [145] A. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 1998. Invited paper, in press.
- [146] A. Smola, B. Schölkopf, and K.-R. Müller. The connection between regularization operators and support vector kernels. *Neural Networks*, 11:637–649, 1998.
- [147] A. Smola, B. Schölkopf, and K.-R. Müller. General cost functions for support vector regression. In T. Downs, M. Freaan, and M. Gallagher, editors, *Proc. of the Ninth Australian Conf. on Neural Networks*, pages 79–83. University of Queensland, 1998.
- [148] A. J. Smola. *Learning with Kernels*. PhD thesis, Technische Universität Berlin, 1998.
- [149] A. J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans. *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [150] P. Sollich. Learning curves for Gaussian processes. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Advances in Neural Information Processing Systems, 11*. MIT Press, 1998.
- [151] R. J. Solomonoff. A formal theory of inductive inference: Part 1. *Inform. Control*, 7:1–22, 1964.
- [152] R. J. Solomonoff. A formal theory of inductive inference: Part 2. *Inform. Control*, 7:224–254, 1964.
- [153] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-posed Problems*. W. H. Winston, 1977.
- [154] A. M. Turing. Computing machinery and intelligence. *Mind*, 49:433–460, 1950.
- [155] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, Nov 1984.
- [156] R. J. Vanderbei. LOQO user's manual – version 3.10. Technical Report SOR-97-08, Princeton University, Statistics and Operations Research, 1997. Code available at <http://www.princeton.edu/rvdb/>.
- [157] V. Vapnik. *Estimation of Dependences Based on Empirical Data [in Russian]*. Nauka, 1979. (English translation Springer Verlag, 1982).

- [158] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.
- [159] V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
- [160] V. Vapnik and O. Chapelle. Bounds on error expectation for SVM. In A.J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [161] V. Vapnik and A. Chervonenkis. A note on one class of perceptrons. *Automation and Remote Control*, 25, 1964.
- [162] 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):264–280, 1971.
- [163] V. Vapnik and A. Chervonenkis. *Theory of Pattern Recognition [in Russian]*. Nauka, 1974. (German Translation: W. Vapnik & A. Tscherwonkis, *Theorie der Zeichenerkennung*, Akademie-Verlag, Berlin, 1979).
- [164] V. Vapnik and A. Chervonenkis. Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theory of Probability and its Applications*, 26(3):532–553, 1981.
- [165] V. Vapnik and A. Chervonenkis. The necessary and sufficient conditions for consistency in the empirical risk minimization method. *Pattern Recognition and Image Analysis*, 1(3):283–305, 1991.
- [166] V. Vapnik and A. Lerner. Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24, 1963.
- [167] V. Vapnik and S. Mukherjee. Support vector method for multivariate density estimation. In *Neural Information Processing Systems (NIPS 99)*, 1999. to appear.
- [168] K. Veropoulos, C. Campbell, and N. Cristianini. Controlling the sensitivity of support vector machines. In *Proceedings of IJCAI Workshop Support Vector Machines*, 1999.
- [169] M. Vidyasagar. *A Theory of Learning and Generalization*. Springer, 1997.
- [170] S. Vijayakumar and S. Wu. Sequential support vector classifiers and regression. In *Proceedings of the International Conference on Soft Computing (SOCO'99)*, pages 610–619, 1999.
- [171] G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, 1990.
- [172] 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*, pages 69–88. MIT Press, 1999.

- [173] G. Wahba, Y. Lin, and H. Zhang. GACV for support vector machines. In A.J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [174] C. Watkins. Dynamic alignment kernels. Technical Report CSD-TR-98-11, Royal Holloway, University of London, Computer Science department, January 1999.
- [175] C. Watkins. Dynamic alignment kernels. In A.J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [176] C. Watkins. Kernels from matching operations. Technical Report CSD-TR-98-07, Royal Holloway, University of London, Computer Science Department, July 1999.
- [177] J. Weston and R. Herbrich. Adaptive margin support vector machines. In A.J. Smola, P. Bartlett, B. Schölkopf, and C. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [178] J. Weston and C. Watkins. Support vector machines for multi-class pattern recognition. In *Proceedings of the 6th European Symposium on Artificial Neural Networks (ESANN)*, 1999.
- [179] B. Widrow and M. Hoff. Adaptive switching circuits. *IRE WESCON Convention record*, 4:96–104, 1960.
- [180] 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.