References

Abarbanel, H., Brown, R., & Kadtke, J., 1989. Prediction and System Identification in Chaotic Time Series with Broadband Fourier Spectra. *Phys. Lett. A*, *138*, 401–408. **Please fill in full name of Journal!**

Abarbanel, H., Brown, R., & Kadtke, J., 1990. Prediction in Chaotic Nonlinear Systems: Methods for Time Series with Broadband Fourier Spectra. *Phys. Rev. A*, *41*, 1782–1807. **Please fill in full name of Journal!**

Abraham-Fuchs, K., Schlang, M., Uebler, J., Hellstrand, E., & H., S., 1992. Map Classification in the Magnetoencephalogram by Neural Networks. In *Proceedings of the 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Piscataway, NJ: IEEE Service Center.

Abu-Mostafa, Y. S., 1989. The Vapnik-Chervonenkis Dimension: Information versus Complexity in Learning. *Neural Computation*, *1*(3), 312–317.

Aha, D., 1990. *A study of instance-based algorithms for supervised learning tasks: Mathematical, empirical, and psychological evaluations*. Ph.D. Dissertation, Department of Information & Computer Science, University of California, Irvine, CA.

Aha, D. & Bankert, R., 1994. Feature selection for case-based classification of cloud types. In *Working Notes of the AAAI94 Workshop on Case-Based Reasoning*, (pp. 106–112). Seattle, WA: AAAI Press.

Aha, D. W., Kibler, D., & Albert, M. K., 1991. Instance-Based Learning Algorithms. *Machine Learning*, *6*, 37–66.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In Petrov & Czaki (Eds.), *Proceedings of the 2nd International Symposium on Information Theory*, (pp. 267–281). Alcademia Kiado, Budapest.

Albano, A., Passamante, A., Hediger, T., & Farell, M. E., 1992. Using Neural Nets to Look for Chaos. *Physica D*, *58*, 1.

Almuallim, H. & Dietterich, T. G., 1991. Learning with many irrelevant features. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, (pp. 547–552). Morgan Kaufmann.

Anderson, J. R. & Matessa, M., 1992. Explorations of an Incremental Bayesian Algorithm for Categorization. *Machine Learning*, *9*, 275–308.

Angluin, D., 1982. Inference of reversible languages. *Journal of the ACM*, *29*(3), 741–765.

Angluin, D. & Laird, P., 1988. Learning from noisy examples. *Machine Learning*, *2*, 343–370.

Anthony, M., 1994. *Probabilistic Analysis of Learning in Artificial Neural Networks: The PAC Model and its Variants.* Mathematics Preprint Series LSE-MPS-67, Dept. of Mathematics, London School of Economics, Houghton St., London WC2A 2AE, UK. Also available as NeuroCOLT technical report NC-TR-94-3, ftp://cscx.cs.rhbnc.ac.uk/pub/neurocolt/tech_reports/.

Anthony, M., Biggs, N., & Shawe-Taylor, J., 1990. The learnability of formal concepts. In *Proceedings of the Third Workshop on Computational Learning Theory*, (pp. 246–257). ACM Press.

Arbib, M. A. (Ed.), 1995. *Handbook of Brain Theory and Neural Networks.* MIT Press.

Arikawa, S., Kuhara, S., Miyano, S., Shinohara, A., & Shinohara, T., 1992. A learning algorithm for elementary formal systems and its experiments on identification of transmembrane domains. In *Proceedings of the 25th Hawaii International Conference on Systems Science*, (pp. 675–684). Los Alamitos, CA: IEEE Computer Society Press.

Atkinson, A. C., 1978. Posterior Probabilities for Choosing a Regression Model. *Biometrika*, *65*, 39–48.

Bacchus, F., 1990. *Representing and Reasoning with Probabilistic Knowledge.* Cambridge, MA: MIT Press.

Bacchus, F., Grove, A., Halpern, J., & Koller, D., 1992. From statistics to beliefs. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, (pp. 602–608). Morgan Kaufmann.

Bartlett, B. & Fyfe, D., 1974. *Handbook of Mathematical Formulae for Engineers and Scientists.* London: Denny Publications Ltd.

Bartlett, P. L., Long, P. M., & Williamson, R. C., 1994. FAT-Shattering and the Learnability of Real-Valued Functions. In *Proceedings of the Seventh Workshop on Computational Learning Theory*, (pp. 299–310). ACM Press.

Baum, E., 1989. When are k-nearest neighbor and back propagation accurate for feasible sized sets of examples? Manuscript.

Baum, E. & Haussler, D., 1989. What Size Net Gives Valid Generalization? *Neural Computation*, *1*(1), 151–160.

Beck, C. & Schlögl, F., 1993. *Thermodynamics of Chaotic Systems*. Cambridge Nonlinear Science Series. Cambridge University Press.

Benedek, G. & Itai, A., 1988. Learnability by fixed distributions. In *Proceedings of the First Workshop on Computational Learning Theory*, (pp. 80–90). ACM Press.

Benedek, G. & Itai, A., 1991. Learnability with repsect to fixed distributions. *Theoretical Computer Science*, *86*, 377–389.

Berger, J. O., 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag.

Bhansali, S., 1991. *Domain-based program synthesis using planning and derivational analogy*. PhD thesis, Department of Computer Science, University of Illinois at Urbana-Champaign. Technical report UIUCDCS R-91-1701.

Bhansali, S. & Harandi, M. T., 1993. Synthesis of UNIX Programs using Derivational Analogy. *Machine Learning*, *10*(2), 7–55.

Bilbro, G. & Van den Bout, D., 1992. Maximum entropy and learning theory. *Neural Computation*, *4*, 839–853.

Binder, K., 1988. Monte Carlo simulation in statistical physics: an introduction. *Springer series in solid-state-sciences*, *80*.

Bishop, C. M., 1994. *Mixture Density Networks*. Technical Report NCRG/4288, Aston University, Birmingham, U.K.

Biswas, G., Weinberg, J., Yang, Q., & Koller, G., 1991. Conceptual Clustering and Exploratory Data Analysis. In *Proceedings of the Eighth International Conference on Machine Learning*, (pp. 591–595). Evanston, IL: Morgan Kaufmann.

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K., 1989a. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, *36*(4), 929–965.

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K., 1989b. Learnability and the Vapnik-Chervonenkis Dimension. *Journal of the ACM*, *36*, 929–965.

Boland, P. J., Proschan, F., & Tong, Y. L., 1989. Modelling dependence in simple and indirect majority systems. *Journal of Applied Probability*, *26*, 81–88.

Bollobás, B., 1985. *Random Graphs*. Academic Press.

Borgida, A. & Etherington, D., 1989. Hierarchical knowledge bases and efficient disjunctive reasoning. In *Proceedings of the First International Conference on Knowledge Repesentation and Reasoning*, (pp. 33–43). Toronto: Morgan Kaufmann.

Brazma, A., 1993a. Efficient identification of regular expressions from representative examples. In *Proceedings of the Sixth Workshop on Computational Learning Theory*, (pp. 236–242). ACM Press.

Brazma, A., 1993b. Learning a subclass of regular expressions by recognizing periodic repetitions. In *Proceedings of the Fourth Scandinavian Conference on AI (SCAI'93)*, (pp. 137–146). IOS Press.

Brazma, A., 1995. Learning of regular expressions by pattern matching. In *Proceedings of the Second European Conference on Computational Learning – Lecture Notes in Artificial Intelligence, 904*, (pp. 392–403). Springer.

Brazma, A. & Cerans, K., 1994. Efficient Learning of Regular Expressions from Good Examples. In *Proceedings of the International Workshop on Analogical and Inductive Inference (AII'94), Lecture Notes in AI, 872*, (pp. 76–90). Springer.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J., 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.

Brent, R. P., 1991. Fast Training Algorithms for Neural Nets. *IEEE Transactions on Neural Networks*, *2*.

Bridle, J., 1989. Probabilistic interpretation of feedforward classification network outputs, with relationship to statistical pattern recognition. In F. Fougelman-Soulie & J. Herault (Eds.), *Neuro-computing: algorithms, architectures and applications*. Springer-Verlag.

Broomhead, D. & Lowe, D., 1988. Multivariable functional interpolation and adaptive networks. *Complex Systems*, *2*(3), 321–355.

Brown, T. H., Zador, A. M., Mainen, Z. F., & Claiborne, B. J., 1992. Hebbian Computations in Hippocampal Dendrites and Spines. In (McKenna et al., 1992), (pp. 81–116).

Buchanan, B. G., Mitchell, T. M., Smith, R. G., & Johnson, Jr., C. R., 1978. Models of Learning Systems. In *Encyclopedia of Computer Science and Technology*, volume 11. Dekker.

Buntine, W. & Weigend, A., 1991. Bayesian back-propagation. *Complex Systems*, *5*(6), 603–643.

Buzug, T. & Pfister, G., 1992. Optimal Delay Time and Embedding Dimension for Delay-Time Coordinates by Analysis of the Global Static and Local Dynamical Behavior of Strange Attractors. *Phys. Rev. A*, *45*, 7073–7084. **Please fill in full name of Journal!**

Carbonell, J. G., 1983. Derivational Analogy and its Role in Problem Solving. In *Proceedings of the Third National Conference on Artificial Intelligence*, (pp. 64–69). Washington, D.C.: Morgan Kaufmann.

Cardie, C., 1993. Using decision trees to improve case-based learning. In *Proceedings of the Tenth International Conference on Machine Learning*, (pp. 25–32). Amherst, MA: Morgan Kaufmann.

Caruana, R. & Freitag, D., 1994. Greedy Attribute Selection. In *Proceedings of the Eleventh International Conference on Machine Learning*, (pp. 28–36). New Brunswick: Morgan Kaufmann.

Casdagli, M., 1989. Nonlinear Prediction of Chaotic Time Series. *Physica D*, *35*, 335–356.

Cheeseman, P., Kelly, J., Self, M., Stutz, J., Taylor, W., & Freeman, D., 1988. AutoClass: A Bayesian Classification System. In *Proceedings of the Fifth International Conference on Machine Learning*. Ann Arbor: Morgan Kaufmann.

Chen, S., Cowan, C., & Grant, P., 1991a. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks*, *2*(2), 302–309.

Chen, S., Gibson, G., Cowan, C., & Grant, P., 1991b. Reconstruction of Binary Signals using an Adaptive Radial-Basis-Function Equalizer. *Signal Processing*, *22*(1).

Chernoff, H., 1952. A Measure of Asymptotic Efficiency for Tests of a Hypothesis based on the Sums of Observations. *Annals of Mathematical Statistics*, *23*, 493–507.

Chow, C. K., 1965. Statistical independence and threshold functions. *IEEE Transactions on Electronic Computers*, *EC-16*, 66–68.

Chow, Y. S., Robbins, H., & Siegmund, D., 1991. *The Theory of Optimal Stopping*. New York: Dover. (Originally published as "Great Expectations: The Theory of Optimal Stopping" by Houghton Mifflin in 1971.).

Cios, K. J. & Liu, N., 1992. A Machine Learning Method for Generation of a Neural Network Architecture: A Continuous ID3 Algorithm. *IEEE Transactions on Neural Networks*, *3*.

Clancey, W., 1985. Heuristic classification. *Artificial Intelligence*, *27*, 289–350.

Cook, R. D. & Weisberg, S., 1980. Characterization of an empirical influence function for detecting influential cases in regression. *Technometrics*, *22*, 495–508.

Cornette, J. L., Cease, K. B., Margalit, H., Spouge, J. L., Berzofsky, J., A., & DeLisi, C., 1987. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *Journal of Molecular Biology*, *195*, 659–685. **Please specify all authors!**

Cost, S. & Salzberg, S., 1993. A Weighted Nearest Neighbor Algorithm for Learning With Symbolic Features. *Machine Learning*, *10*, 57–78.

Cover, T. & Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*, 21–27.

Cover, T. M., 1965. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, *14*, 326–334.

Craven, P. & Wahba, G., 1979. Smoothing Noisy Data with Spline Functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerical Mathematics*, *31*, 377–403.

Crutchfield, J. & McNamara, 1987. Equations of Motion from Data Series. *Complex Systems*, *1*, 417–452.

de Sa, V., 1994. Learning Classification with Unlabeled Data. In Cowan, J. D., Tesauro, G., & Alspector, J. (Eds.), *Advances in neural information processing systems 6*, (pp. 112–119). Denver, CO: Morgan Kaufmann.

DeLisi, C., 1988. Computers in molecular biology: current applications and emerging trends. *Science*, *240*, 47–51.

Dempster, A. P., Laird, N. M., & Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society Series B*, *39*, 1–38.

Devroye, L., 1982. Bounds for the uniform deviation of empirical measures. *Journal of Multivariate Analysis*, *12*, 72–79.

Dietterich, T. G., Jain, A. N., Lathrop, R. H., & Lozano-Perez, T., 1994. A Comparison of Dynamic Reposing and Tangent Distance for Drug Activity Prediction. In Cowan, Tesauro, & Alspector (Eds.), *Advances in neural information processing systems 6*. San Francisco: Morgan Kaufmann.

Doorenbos, R. B., 1993. Matching 100,000 learned rules. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, (pp. 290–296). Washington D.C.: Morgan Kaufmann.

Duda, R. O. & Hart, P. E., 1973. *Pattern Classification and Scene Analysis*. New York: Wiley.

Durfy, S. & Willard, H., 1989. Patterns of intra- and interarray sequence variation in alpha satellite from the human X chromosome: evidence for shortrange homogenization of tandemly repeated DNA sequences. *Genomics*, *5*, 810–821.

Eckmann,
J. & Ruelle, D., 1985. Ergodic Theory of Chaos and Strange Attractors. *Rev. Mod. Phys.*, *57*, 617–656. **Please fill in full name of Journal!**

Efron, B. & Tibshirani, R., 1991. Statistical data analysis in the computer age. *Science*, *253*, 390–395.

Efron, B. F. & Morris, C. N., 1973. Stein's Estimation Rule and Its Competitors — An Empirical Bayes Approach. *Journal of the American Statistics Association*, *68*, 117–130.

Ehrenfeucht, A., Haussler, D., Kearns, M., & Valiant, L., 1988. A general lower bound on the number of examples needed for learning. In *Proceedings of the First Workshop on Computational Learning Theory*, (pp. 139–154). Morgan Kaufmann.

Ehrenfeucht, A., Haussler, D., Kearns, M., & Valiant, L., 1989. A general lower bound on the number of examples needed for learning. *Information* *and Computation*, *82*, 247–261.

Fahlman, S. & Lebiere, C., 1990. The Cascade-Correlation Architecture. In *Advances in Neural Information Processing Systems 2*, (pp. 524–532). Morgan Kaufmann.

Farmer, J. & Sidorowich, J., 1987. Predicting Chaotic Time Series. *Phys. Rev. Letters*, *59*, 845. **Please fill in full name of Journal!**

Farrell, J. & Baker, W., 1992. Learning Control Systems. In P. Antsaklis & K. Passino (Eds.), *An Introduction Intelligent and Autonomous Control Systems*. Boston, MA: Klumwer Academic.

Fisher, D., 1987. Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, *2*, 139–172.

Fisher, D., Xu, L., & Zard, N., 1992. Ordering Effects in Clustering. In *Proceedings of the Ninth International Conference on Machine Learning*, (pp. 163–168). San Mateo, CA: Morgan Kaufmann.

Fisher, R., 1936. The use of Multiple Measurements in Taxonomic Problems. *Annual Eugenics*, *7, Part II*, 179–188.

Frean, M., 1990. The Upstart Algorithm: A Method for Constructing and Training Feedforward Neural Networks. *Neural Computation*, *2*, 198–209.

Freund, Y., 1990. Boosting a weak learning algorithm by majority. In *Proceedings of Workshop on Computational Learning Theory*, (pp. 202–216). Morgan Kaufmann.

Freund, Y., Seung, S., Shamir, E., & Tishby, N., 1992. Information, Prediction and Query by Committee. In *Advances in Neural Information Processing System 5*, (pp. 483–490). Morgan Kaufmann.

Friedman, J. H., 1991. Multivariate adaptive regression splines. *The Annals of Statistics*, *19*, 1–141.

Froehling, H., Crutchfield, J., Farmer, D., Packard, N., & Shaw, R., 1981. On Determining the Dimension of Chaotic Flow. *Physica (Utrecht)*, *3D*, 605.

Gelb, A., 1974. *Applied Optimal Estimation*. MIT Press.

Genest, C. & Zidek, J. V., 1986. Combining probability distributions:A critique and an annotated bibliography. *Statistical Science*, *1*(1), 114–148.

Gennari, J., Langley, P., & Fisher, D., 1989. Models of incremental concept formation. *Artificial Intelligence*, *40*, 11–61.

Gentner, D., 1983. Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, *7*(2), 155–170.

Ghahramani, Z., 1994. Solving inverse problems using an EM approach to density estimation. In *Proceedings of the 1993 Connectionist Models Summer School*. Hillsdale, NJ: Erlbaum.

Ghahramani, Z. & Jordan, M. I., 1994. Supervised Learning from Incomplete Data via an EM approach. In Cowan, J. D., Tesauro, G., & Alspector, J. (Eds.), *Advances in neural information processing systems 6*, (pp. 120–127). Denver, CO: Morgan Kaufmann.

Ghahramani, Z. & Jordan, M. I., 1996. Mixture Models for Learning from Incomplete Data. In *Proceedings of the Fourth Annual Workshop on Computational Learning Theory and Natural Learning Systems*. MIT Press. (This volume).

Ginsberg, M., 1987. *Readings in Nonmonotonic Reasoning*. Los Altos, CA: Morgan Kaufmann.

Gluck, M. & Corter, J., 1985. Information, Uncertainty, and the Utility of Categories. In *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, (pp. 283–287). Irvine, CA: Lawrence Erlbaum.

Gold, E., 1967. Language identification in the limit. *Journal of Information and Control*, *10*, 447–474.

Goldman, S., Kearns, M., & Schapire, R., 1990. Exact identification of circuits using fixed points of amplification functions. In *Proceedings of the 31st Foundations of Computer Science*. IEEE Computer Society Press.

Goldman, S. A. & Sloan, R. A., 1995. Can PAC Learning Algorithms Tolerate Random Attribute Noise? *Algorithmica*, *14*(1).

Golea, M., 1993. *On Efficient Learning Algorithms for Neural Networks*. Technical Report PhD Thesis, University of Ottawa.

Golea, M., 1995. Average Case Analysis of a Learning Algorithm for μ-DNF Expressions. In *Proceedings of the Second European Conference on Computational Learning Theory*, (pp. 342–356). Springer-Verlag. Lecture Notes in Artificial Intelligence, Vol. 904.

Golea, M. & Marchand, M., 1993a. Learning Curves of the Clipped Hebb Rule for Networks with Binary Weights. *Journal of Physics A: Mathematical and General*, *26*, 5751–5766.

Golea, M. & Marchand, M., 1993b. On Learning Perceptrons with binary weights. *Neural Computation*, *5*, 767–782.

Golea, M., Marchand, M., & Hancock, T., 1992. On Learning μ-Perceptron Networks with Binary Weights. In *Advances in Neural Information Processing Systems 4*, (pp. 591–598). Morgan Kaufmann.

Golub, G. & van Loan, C., 1990. *Matrix Computations (2nd ed.)*. London: John Hopkins University Press.

Grassberger, P. & Procaccia, I., 1983. Characterization of Strange Attractors. *Phys. Rev. Lett.*, *50*, 346. **Please fill in full name of Journal!**

Greiner, R. & Elkan, C., 1991. Measuring and Improving the Effectiveness of Representations. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, (pp. 518–524). Sydney: Morgan Kaufmann.

Greiner, R., Hancock, T., & Rao, R. B., 1994. Knowing What Doesn't Matter. In *Proceedings of the "Relevance" Symposium*. New Orleans: AAAI Press.

Greiner, R. & Schuurmans, D., 1992. Learning an Optimally Accurate Representational System. In *ECAI Workshop on Theoretical Foundations of Knowledge Representation and Reasoning*. Vienna: Morgan Kaufmann.

Greiner, R. & Subramanian, D., 1995. *Proceedings of the "Relevance" Symposium*. Technical Report FS-94-02, AAAI, Menlo Park.

Gyorgyi, G., 1990. First-order transition to perfect generalization in a neural network with binary synapses. *Physical Review A*, *41*, 7097–7100.

Hampel, F., Rouchetti, E., Rousseeuw, P., & Stahel, W., 1986. *Robust Statistics: The approach based on influence functions*. Wiley.

Hanson, S. & Bauer, M., 1989. Conceptual Clustering, Categorization, and Polymorphy. *Machine Learning*, *3*, 343–372.

Hanson, S. J., 1990. A Stochastic Version of the Delta Rule. *Physica D*, *42*, 265–272.

Hartigan, J. A., 1975. *Clustering Algorithms*. New York: Wiley.

Hassibi, B. & Stork, D., 1992. Second Order Derivatives for Network Pruning: Optimal Brain Surgeon. In *Advances in Neural Information Processing Systems 4*, (pp. 164–171). Morgan Kaufmann.

Hastie, T. & Tibshirani, R., 1994. *Discriminant Analysis by Gaussian mixtures.* Technical report, AT&T Bell Laboratories, Murray Hill, NJ.

Hastie, T. & Tibshirani, R., 1996. Discriminate Adaptive Nearest Neighbor Classification and Regression. In D, T., Mozer, M., & Hasselmo, M. (Eds.), *Advances in Neural Information Processing Systems*, volume 8. Denver, CO: MIT Press.

Haussler, D., 1987. Learning conjunctive concepts in structural domains. In *Proceedings of the Sixth National Conference on Artificial Intelligence*, (pp. 466–470). Morgan Kaufmann.

Haussler, D., 1988. Quantifying inductive bias: AI learning algorithms and Valiant's learning framework. *Artificial Intelligence*, *36*, 117–221.

Haussler, D., 1990. Probably approximately correct learning. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, (pp. 1101–1108). Morgan Kaufmann.

Haussler, D., 1992. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, *100*, 78–150.

Haussler, D., Kearns, M., Opper, M., & Schapire, R., 1992. Estimating average-case learning curves using Bayesian, statistical physics and VC dimension methods. *Advances in Neural Information Processing Systems 4*, (pp. 855–862).

Haussler, D., Kearns, M., Seung, H. S., & Tishby, N., 1994. Rigorous Learning Curve Bounds from Statistical Mechanics. In *Proceedings of the Seventh Workshop on Computational Learning Theory*, (pp. 67–75). New Brunswick, NJ: ACM Press.

Hebb, D., 1949. *The Organization of Behavior.* New-York: Wiley.

Hénon, M., 1976. A Two-Dimensional Mapping with a Strange Attractor. *Comm. Math. Phys.*, *50*, 69. **Please fill in full name of Journal!**

Hertz, J., Krogh, A., & Palmer, R., 1991a. *Introduction to the Theory of Neural Computation.* Addison-Wesley.

Hertz, J., Krogh, A., & Palmer, R. G., 1991b. *Introduction to the theory of neural computation.* Redwood City, California: Addison-Wesley.

Hinton, G., 1989. Connectionist learning procedures. *Artificial Intelligence*, *40*, 185–234.

Hinton, G. E. & Sejnowski, T. J., 1986. Learning and relearning in Boltzmann machines. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press.

Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, *58*(301), 13–30.

Hogg, R. & Tanis, E., 1988. *Probability and Statistical Inference*. New-York: Macmillan.

Holden, S. & Rayner, P., 1992. Generalization and learning in Volterra basis function networks: a theoretical analysis. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, (pp. II: 273–276). Morgan Kaufmann.

Holland, P. & Welsch, R., 1977. Robust regression using iteratively reweighted least-squares. *Communcations in Statistics*, *6*, 813–88.

Hollatz, J. & Tresp, V., 1992. A Rule-based Network Architecture. In I. Aleksander & J. Taylor (Eds.), *Artifical Neural Network II*. Amsterdam: Elsevier.

Huber, P., 1981. *Robust Statistics*. Wiley.

Iba, W. & Langley, P., 1993. Average-Case Analysis of a Nearest Neighbor Algorithm. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, (pp. 889–894). San Francisco: Morgan Kaufmann.

Ikehara, S., Shirai, S., Yokoo, A., & Nakaiwa, H., 1990. Toward an MT System without Pre-Editing—Effects of New Methods in ALT-J/E. In *MT Summit-3 Proceedings*. MT Summit III Organizing Committee, C/O of Center for Machine Translation at Carnegie Mellon University.

Jacobs, R., Jordan, M., Nowlan, S., & Hinton, G., 1991. Adaptive mixture of local experts. *Neural Computation*, *3*, 79–87.

Jazwinski, A., 1970. *Stochastic Processes and Filter Theory*. New York, NY: Academic Press.

John, G. H., Kohavi, R., & Pfleger, K., 1994. Irrelevant Features and the Subset Selection Problem. In *Proceedings of the Eleventh International Conference on Machine Learning*, (pp. 121–129). New Brunswick: Morgan Kaufmann.

Johnson, D. S. & Niemi, K. A., 1983. On knapsacks, partitions and a new dynamic programming technique for trees. *Mathematics of Operations Research*, *8*(1).

Jordan, M. & Jacobs, R., 1994. Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, *6*, 181–214.

Judd, J. S. & Munro, P. W., 1993. Nets with Unreliable Hidden Units Learn Error-Correcting Codes. In Hanson, S. J., Cowan, J. D., & Giles, C. L. (Eds.), *Advances in neural information processing systems 5*, (pp. 483–490). Denver, CO: Morgan Kaufmann.

Kalil, R., 1989. Synapse Formation in the Developing Brain. *Scientific American*, *261*, 38–45.

Kambhampati, S. & Hendler, J., 1992. A validation-structure based theory of plan modification and reuse. *Artificial Intelligence*, *55*(2–3), 193–258.

Kearns, M., Li, M., Pitt, L., & Valiant, L., 1987. On the Learnability of Boolean Formulae. In *Proceedings of the 19th Symposium on the Theory of Computations*, (pp. 285–295). New York: ACM Press.

Kearns, M. & Seung, H. S., 1993. Learning from a population of hypotheses. In *Proceedings of Annual ACM Conference on Computational Learning Theory*, (pp. 101–110). ACM Press.

Kearns, M. J. & Li, M., 1988. Learning in the presence of malicious errors. In *Proceedings of the 20th Symposium on the Theory of Computation*, (pp. 267–280). ACM Press.

Kearns, M. J. & Schapire, R. E., 1990. Efficient distribution-free learning of probabilistic concepts. In *Proceedings of the 31st Foundations of Computer Science*. IEEE Computer Society Press.

Kearns, M. J., Schapire, R. E., & Sellie, L. M., 1992. Toward efficient agnostic leaning. In *Proceedings of the Fifth Workshop on Computational Learning Theory*. ACM Press.

Kira, K. & Rendell, L., 1992. A practical approach to feature selection. In *Proceedings of the Ninth International Conference on Machine Learning*, (pp. 249–256). Aberdeen, Scotland: Morgan Kaufmann.

Koch, C., Bernander, Ö., & Douglas, R. J., 1995. Do neurons have a voltage or a current threshold for action potential initiation? *Journal of Computational Neuroscience*, *2*, 63–82.

Koch, C., Poggio, T., & Torre, V., 1982. Retinal Ganglion Cells: a Functional Interpretation of Dendritic Morphology. *Proceedings of the Royal Society of London B*, *298*, 227–264.

Kocher, I. & Monasson, R., 1992. On the Capacity of Neural Networks with Binary Weights. *Journal of Physics A: Mathematical and General*, *25*, 367–380.

Kohavi, R., 1994. Bottom-up induction of oblivious read-once decision graphs: Strengths and limitations. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, (pp. 613–618). Seattle, WA: AAAI Press.

Köhler, H., Diederich, S., Kinzel, W., & Opper, M., 1990. Learning Algorithm for a Neural Network with Binary Synapses. *Zeitschrift fur Physik B*, *78*, 333–342.

Konagaya, A., 1992. A Stochastic Approach to Genetic Information. In *Proceedings of the Third Workshop on Algorithmic Learning Theory (ALT'92)*, (pp. 25–36). Japanese Society for Artificial Intelligence.

Koton, P. A., 1989. Evaluating case-based problem solving. In *Proceedings of a Workshop on Case-based Reasoning*, (pp. 173–175). Pensacola Beach, FL: Morgan Kaufmann.

Kressel, U., Franke, J., & Schurmann, J., 1991. *Polynomial classifier versus multilayer perceptron.* Technical report, Daimler-Benz AG, Research Center Ulm, Wilhelm-Runge-Str. 11, 7900 Ulm, F.R. Germany.

Krogh, A. & Hertz, J., 1992. Generalization in a linear perceptron in the presence of noise. *Journal of Physics*, *A25*, 1135–1148.

Kubat, M., Flotzinger, D., & Pfurtscheller, G., 1993. Discovering patterns in EEG signals: Comparative study of a few methods. In *Proceedings of the 1993 European Conference on Machine Learning*, (pp. 367–371). Vienna, Austria: Springer-Verlag.

Kulkarni, S., 1991. *Problems of computational and information complexity in machine vision and learning.* PhD thesis, MIT, Department of Electrical Engineering and Computer Science.

Kullback, S. & Leibler, R., 1951. On Information and Sufficiency. *Annals of Statistics*, *22*, 79–86.

Kyburg, H., 1983. The reference class. *Philosophy of Science*, *50*, 374–397.

Kyburg, H., 1991. Evidential probability. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, (pp. 1196–1202). Sydney: Morgan Kaufmann.

Landau, G. & Schnidt, J., 1993. An algorithm for approximate tandem repeats. *Lecture Notes in Computer Science*, *684*, 134–151.

Langley, P. & Iba, W., 1993. Average-case analysis of a nearest neighbor algorithm. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, (pp. 889–894). Chambery, France: Morgan Kaufmann.

Lapedes, A. & R.Farber, 1987. *Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling.* Technical Report LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM.

Larsen, R. J. & Marx, M. L., 1981. *An Introduction to Mathematical Statistics and its Applications.* Englewood Cliffs, NJ: Prentice-Hall.

le Cun, Y., Denker, J., & Solla, S., 1990. Optimal Brain Damage. In *Advances in Neural Information Processing Systems 2*, (pp. 598–605). Morgan Kaufmann.

Leacock, C., Towell, G., & Voorhees, E. M., 1993. Towards Building Contextual Representations of Word Senses Using Statistical Models. In *Proceedings of SIGLEX Workshop: Acquisition of Lexical Knowledge from Text*, (pp. 10–20). Association for Computational Linguistics.

Lewis, D. D. & Catlett, J., 1994. Heterogeneous Uncertainty Sampling for Supervised Learning. In *Proceedings of the Eleventh International Conference on Machine Learning*, (pp. 149–156). Morgan Kaufmann.

Liebert, W., Pawelzik, K., & Schuster, H., 1991. Optimal Embedding of Chaotic Attractors from Topological Considerations. *Europhysics Letters*, *14*, 521–526.

Liebert, W. & Schuster, H., 1989. Proper Choice of the Time Delay for the Analysis of Chaotic Time Series. *Phys. Lett. A*, *142*, 107–111. **Please fill in full name of Journal!**

Lindley, D. V., 1968. The Choice of Variables in Multiple Regression (with discussion). *Journal of the Royal Statistics Society, Series B, 30*, 31–66.

Linial, N., Mansour, Y., & Rivest, R., 1988. Some results on learnability and the Vapnik-Chervonenkis dimension. In *Proceedings of the First Workshop on Computational Learning Theory.* ACM Press.

Little, J. A. & Rubin, D. B., 1987. *Statistical Analysis with Missing Data.* New York: Wiley.

Littlestone, N., 1988. Learning quickly when irrelevant attributes abound: A new linear threshold algorithm. *Machine Learning, 2*, 285–318.

Littlestone, N., 1989. From online to batch learning. In *Proceedings of the Second Workshop on Computational Learning Theory*, (pp. 269–284). ACM Press.

Littlestone, N. & Warmuth, M. K., 1989. The weighted majority algorithm. In *Proceedings of Symposium on Foundations of Computer Science*, (pp. 256–261). Proceedings of IEEE.

Liu, Y., 1993. Neural Network Model Selection Using Asymptotic Jackknife Estimator and Cross-Validation Method. In Giles, C., Hanson, S., & Cowan, J. (Eds.), *Advances in neural information processing system 5*, (pp. 599–606). Morgan Kaufmann.

Liu, Y., 1994. Robust Parameter Estimation and Model Selection for Neural Network Regression. In J. D. Cowan, G. Tesauro, & J. Alspector (Eds.), *Advances in neural information processing systems 6.* San Mateo, CA: Morgan Kaufmann.

Liu, Y., 1995. Unbiased Estimate of Generalization Error and Model Selection in Neural Network. *Neural Networks, 8*(2), 215–219.

Livstone, M., Farrell, J., & Baker, W., 1992. A Computationally Efficient Algorithm for Training Recurrent Connectionist Networks. In *Proceedings 1992 American Control Conference*, volume 2, (pp. 555–561). Chicago. **Please fill in Publisher!**

Ljung, L., 1987. *System Identification: Theory for the User.* Englewood Cliff New Jersey, USA: Prentice Hall.

Lowe, D., 1989. Adaptive radial basis function nonlinearities and the problem of generalization. *Proceedings of the First IEE International Conference on Artificial Neural Networks*, (pp. 171–175).

Lynch, M. & Rayner, P., 1989. Properties and implementation of the nonlinear vector space connectionist model. In *Proceedings of the IEE First International Conference on Artificial Neural Networks*. London. **Please fill in Publisher!**

Lyndon, R. & Schutzenberg, M., 1962. The equation $a^M = b^N c^P$ in a free group. *Michigan Math. J.*, *9*, 289–298.

Maass, W., 1995. Vapnik-Chervonenkis Dimension of Neural Networks. In (Arbib, 1995), (pp. 1000–1002).

Maass, W., 1996. Lower bounds for the computational power of networks of spiking neurons. *Neural Computation*, *8*(1).

MacKay, D., 1991. *Bayesian methods for adaptive models*. PhD thesis, California Institute of Technology.

MacKay, D., 1992a. Bayesian interpolation. *Neural Computation*, *4*, 415–447.

MacKay, D., 1992b. A practical Bayesian framework for backpropagation. *Neural Computation*, *4*, 448–472.

Mallows, C. L., 1973. Some Comments on C_p. *Technometrics*, *15*, 661–675.

Marchand, M., Golea, M., & Rusján, P., 1990. A convergence Theorem for Sequential Learning in Two-Layer Perceptrons. *Europhysics Letters*, *11*, 487–492.

Mato, G. & Parga, N., 1992. Generalization Properties of Multilayerd Neural Networks. *Journal of Physics A: Mathematical and General*, *25*, 5047–5054.

Matsuoka, K., 1992. Noise Injection into Inputs in Back-Propagation Learning. *IEEE Transactions on Systems, Man and Cybernetics*, *22*, 436–440.

Mazurov, V. D., Krivonogov, A. I., & Kazantsev, V. S., 1987. Solving of optimization and identification problems by the committee methods. *Pattern Recognition*, *20*(4), 371–378.

McKenna, T., Davis, J., & Zornetzer, S. F. (Eds.), 1992. *Single Neuron Computation*. Academic Press.

McKusick, K. & Langley, P., 1991. Constraints on Tree Structure in Concept Formation. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, (pp. 810–816). Sydney: Morgan Kaufmann.

McLachlan, G. & Basford, K., 1988. *Mixture models: Inference and applications to clustering.* Marcel Dekker.

Meir, R. & Fontanari, F., 1992. Calculation of learning curves for inconsistent algorithms. *Physical Review A*, *45*, 8874–8884.

Mel, B. W., 1992. NMDA-Based Pattern Discrimination in a Modeled Cortical Neuron. *Neural Computation*, *4*(4), 502–516.

Mézard, M. & Nadal, J., 1989. Learning in Feedforward Neural Network: the Tiling Algorithm. *Journal of Physics A: Mathematical and General*, *22*, 2191–2203.

Michalski, R., 1980. Learning by being told and learning from examples: an experimental comparison of the two methods of knowledge acquisition in the context of developing an expert system for soybean disease diagnosis. *International Journal of Policy Analysis and Information Systems*, *4*(2), 125–161.

Michalski, R. & Stepp, R., 1983. Learning from observation: Conceptual clustering. In R. Michalski, J. Carbonell, & T. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach.* Palo Alto: Tioga.

Miller, I. & Freund, J., 1985. *Probability and Statistics for Engineers (3rd ed.).* New Jersey: Prentice Hall.

Miller, N. R., 1986. Information, electorates, and democracy: Some extensions and interpretations of the Condorcet Jury Theorem. In B. Grofman & G. Owen (Eds.), *Information Pooling and Group Decision Making* (pp. 173–192). Jai Press Inc., Greenwich, Connecticut.

Mingers, J., 1989. An Empirical Comparison of Pruning Methods for Decision Tree Induction. *Machine Learning*, *4*, 227–243.

Minsky, M., 1975. A framework for representing knowledge. In P. H. Winston (Ed.), *The Psychology of Computer Vision* (pp. 211–277). McGraw-Hill.

Mitchell, T. M., 1980. *The need for biases in learning generalizations.* Technical Report CBM-TR-117, Rutgers University.

Miyano, S., 1993. Learning Theory Toward Genome Informatics. In *Proceedings of the Fourth Workshop on Algorithmic Learning Theory (ALT'93)*, (pp. 19–36). Springer.

Moody, J. & Darken, C., 1989. Fast learning in networks of locally-tuned processing units. *Neural Computation*, *1*(2), 281–294.

Moody, J. & Darken, C., 1990. Fast adaptive k-means clustering: Some emperical results. In *Proceedings of the International Joint Conference on Neural Networks*. Piscataway, NJ: IEEE Service Center.

Moody, J. E., 1992. The effective number of parameters, an analysis of generalization and regularization in nonlinear learning system. In Moody, J. E., Hanson, S. J., & Lippmann, R. P. (Eds.), *Advances in neural information processing system 4*, (pp. 847–854). Morgan Kaufmann.

Moody, J. E., Hanson, S. J., & Lippmann, R. P. (Eds.)., NIPS*91, 1992. *Advances in Neural Information Processing Systems 4*. Morgan Kaufmann.

Mooney, R. J., Shavlik, J. W., Towell, G. G., & Gove, A., 1989. An Experimental Comparison of Symbolic and Connectionist Learning Algorithms. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*. Morgan Kaufmann.

Moore, A. & Lee, M., 1994. Efficient algorithms for minimizing cross validation error. In *Proceedings of the Eleventh International Conference on Machine Learning*, (pp. 190–198). New Brunswick, NJ: Morgan Kaufmann.

Mostow, J., 1989. Design by derivational analogy: issues in the automated replay of design plans. *Artificial Intelligence*, *40*, 119–184.

Murata, N., Yoshizawa, S., & Amari, S., 1993. Learning Curves, Model Selection and Complexity of Neural Networks. In Giles, C., Hanson, S., & Cowan, J. (Eds.), *Advances in neural information processing system 5*. Morgan Kaufmann.

Murphy, P. & Aha, D., 1992. *UCI Repository of machine learning databases* [Machine-readable data repository]. Irvine, CA: University of California, Department of Information and Computer Science.

Myers, E. & Miller, W., 1989. Approximate matching of regular expression. *Bulletin of Mathematical Biology*, *51*(1), 5–37.

Neal, R., 1993. Bayesian learning via stochastic dynamics. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in neural information processing systems 5* (pp. 475–482). San Mateo, CA: Morgan Kaufmann.

Neal, R. M., 1992. Connectionist learning of belief networks. *Artificial Intelligence*, *56*, 71–113.

Nowlan, S. J., 1991. *Soft Competitive Adaptation: Neural Network Learning Algorithms based on Fitting Statistical Mixtures.* CMU-CS-91-126, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Opitz, D. W. & Shavlik, J. W., 1994. Using Genetic Search to Refine Knowledge-Based Neural Networks. In *Proceedings of the Eleventh International Conference on Machine Learning*, (pp. 208–215). Morgan Kaufmann.

Opper, M. & Haussler, D., 1991. Generalization Performance of Bayes Optimal Classification Algorithm for Learning a Perceptron. *Physical Review Letters*, *66*, 2677–2680.

Opper, M., Kinzel, W., Kleinz, J., & Nehl, R., 1990. On the Ability of the Optimal Perceptron to Generalize. *Journal of Physics A: Mathematical and General*, *23*, L581–L586.

Parker, D., 1987. Optimal Algorithms for Adaptive Neural networks: Second Order Backpropagation, Second Order Direct Propagation, And Second Order Hebbian Learning. In *Proceeding of the First IEEE International Conference on Neural Networks*, (pp. 593–600). San Diego, CA: IEEE Computer Society Press. Volume 2.

Patterson, J., 1990. An Optimum Multilayer Perceptron Neural receiver for Signal Detection. *IEEE Transactions on Neural Networks*, *4*(1).

Pawelzik, K. & Schuster, H., 1991. Unstable Periodic Orbits and Prediction. *Phys. Rev. A*, *43*, 1808–1812. **Please fill in full name of Journal!**

Pazzani, M. J. & Sarret, W., 1990. Average Case Analysis of Conjunctive Learning Algorithm. In *Proceedings of the Seventh International Conference on Machine Learning*, (pp. 339–47). Austin, TX: Morgan Kaufmann.

Pazzani, M. J. & Sarrett, W., 1992. A framework for the average case analysis of conjunctive learning algorithms. *Machine Learning*, *9*, 349–372.

Pearl, J., 1988. *Probabilistic Reasoning in Intelligent Systems.* San Mateo, CA: Morgan Kaufmann.

Pearl, J., 1989. Probabilistic semantics for nonmonotonic reasoning: A survey. In *Proceedings of the First International Conference on Knowledge Repesentation and Reasoning*, (pp. 505–516). Toronto: Morgan Kaufmann.

Pearlmutter, B., 1989. Learning State Space Trajectories in Recurrent Neural Networks. *Neural Computation*, *1*, 239–269.

Peebles, P. Z., 1987. *Probability, Random Variables and Random Signal Principles*. Singapore: McGraw Hill International.

Peitgen, H.-O. & Richter, P. H., 1986. *The beauty of fractals: images of complex dynamical systems*. Berlin: Springer-Verlag.

Poggio, T. & Girosi, F., 1989. *A theory of networks for approximation and learning*. A.I. Lab Memo 1140, MIT, Cambridge, MA.

Polat, U. & Sagi, D., 1994. Spatial Interactions in Human Vision: From Near to Far via Experience-dependent Cascades of Connections. *Proceedings of the National Academy of Science USA*, *91*, 1206–1209.

Pomerleau, D. A., 1993. *Neural Network Perception for Mobile Robot Guidance*. Boston: Kluwer.

Porter, B. W., Bareiss, R., & Holte, R. C., 1990. Concept learning and heuristic classification in weak-theory domains. *Artificial Intelligence*, *45*(1-2), 229–263.

Pouskoris, G. & Feldkamp, L., 1991. Decoupled Extended Kalman Filter Training of Feedforward Layered Networks. In *International Joint Conference on Neural Networks, vol. I*, (pp. 771–777). Seattle: IEEE Computer Society Press.

Principe, J., Rathie, A., & Kuo, J., 1992. Prediction of Chaotic Time Series with Neural Networks and the Issue of Dynamic Modeling. *Bifurcation and Chaos*, *2*(4), 989.

Quinlan, J. R., 1986. Induction of Decision Trees. *Machine Learning*, *1*(1).

Quinlan, J. R., 1989. Unknown attribute values in induction. In *Proceedings of the Sixth International Conference on Machine Learning*, (pp. 164–168). Morgan Kaufmann.

Quinlan, J. R., 1993. *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann.

Rao, N. S. V., 1995. Fusion rule estimation in multiple sensor systems with unknown noise densities. *Journal of the Franklin Institute*, *331B*(5), 509–530.

Rao, N. S. V. & Oblow, E. M., 1994. Majority and location-based fusers for PAC concept learners. *IEEE Transactiosn on Systems, Man and Cybernetics*, *24*(5), 713–727.

Rao, N. S. V., Oblow, E. M., Glover, C. W., & Liepins, G. E., 1994. N-learners Problem: Fusion of concepts. *IEEE Transactions on Systems, Man and Cybernetics*, *24*(2), 319–327.

Rayner, P. & Lynch, M., 1989. A new connectionist model based on a nonlinear adaptive filter. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. Morgan Kaufmann.

Reiter, R., 1987. Nonmonotonic reasoning. *Annual Review of Computer Science*, *2*, 147–186.

Renka, R., 1988. Multivariate Interpolation of large sets of scattered data. *ACM transactions on mathematical software*, *14*(2), 139–148.

Riedmiller, M., 1994. Advanced supervised learning in multi-layer perceptrons: from backpropagation to adaptive learning algorithms. *International Journal of Computer Standards and Interfaces*, *5*. (special issue on neural networks).

Rissanen, J., 1990. *Complexity of models*. Redwood City,: Addison-Wesley.

Rivest, R. & Sloan, R., 1988. Learning complicated concepts reliably and usefully. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, (pp. 635–640). Morgan Kaufmann.

Romanik, K., 1992. Approximate testing and learnability. In *Proceedings of the Fifth Workshop on Computational Learning Theory*, (pp. 327–332). ACM Press.

Rosenblatt, F., 1958. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, *65*.

Roth, D., 1993. On the hardness of approximate reasoning. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, (pp. 613–618). Morgan Kaufmann.

Roux, J.-C., Simoyi, R. H., & Swinney, H. L., 1983. Observation of a Strange Attractor. *Physica D*, *8*, 257–266.

Rumelhart, D., Widrow, B., & Lehr, M., 1994a. Models of the brain and evolution: the basic ideas in neural networks. *Communications of the ACM*, *37*, 87–92.

Rumelhart, D., Widrow, B., & Lehr, M., 1994b. Neural networks: applications in industry business and science. *Communications of the ACM*, *37*, 93–105.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J., 1986. Learning Internal Representation by Error Propagation. In D. Rumelhart, J. McClelland, & the PDP Research Group (Eds.), *Parallel Distributed Processing, Volume 1: Foundation* chapter 8, (pp. 318–362). The MIT Press.

Sauer, T., Yorke, J., & Casdagli, M., 1991. Embedology. *J. Stat. Phys.*, *65*, 579–616. **Please fill in full name of Journal!**

Schafer, J., 1994. *Analysis of Incomplete Multivariate Data by Simulation.* London: Chapman & Hall.

Scharf, L., 1991. *Statistical Signal Processing*, (pp. 43–51). New York: Addison-Wesley.

Schetzen, M., 1980. *The Volterra and Wiener Theories of Nonlinear Systems.* Wiley.

Schlang, M., Haft, M., & Abraham-Fuchs, K., 1994. Neural Networks for Reconstruction of Focal Events from Bioelectric/Biomagnetic Potentials. In *Proceedings of the IEEE World Congress on Computational Intelligence*, volume VI. Piscataway, NJ: IEEE Service Center.

Schlang, M., Tresp, V., Abraham-Fuchs, K., Haerer, W., & Weismueller, P., 1992. Neural Networks for Segmentation and Clustering of Biomagnetic Signals. In *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing.* Piscataway, NJ: IEEE Service Center.

Schlimmer, J., 1987. *Concept acquisition through representational adjustment.* Ph.D. Dissertation, Department of Information & Computer Science, University of California, Irvine, CA.

Schlimmer, J., 1993. Efficiently inducing determinations: A complete and efficient search algorithm that uses optimal pruning. In *Proceedings of the Tenth International Conference on Machine Learning*, (pp. 284–290). Amherst, MA: Morgan Kaufmann.

Schneider, S., Hoenig, E., Reichenberger, H., Abraham-Fuchs, K., Daalmans, G., Moshage, W., Oppelt, A., Roehrlein, G. Stefan, H., Vieth, J., Weikl, A., & Wirth, A., 1990. A Multichannel Biomagnetic System for High Resolution Functional Studies of Brain and Heart. *Radiology*, *176*, 825–830.

Schuurmans, D., 1996. *Effective Classification Learning.* PhD thesis, University of Toronto, Department of Computer Science.

Schuurmans, D. & Greiner, R., 1994. Learning default concepts. In *Proceedings of the Tenth Canadian Conference on Artificial Intelligence*, (pp. 99–106). Morgan Kaufmann.

Schuurmans, D. & Greiner, R., 1995. Sequential PAC Learning. In *Proceedings of the Eighth Workshop on Computational Learning Theory*, (pp. 268–277). ACM Press.

Schwartz, G., 1978. Estimating the Dimension of a Model. *Annual Statitics*, *6*, 461–464.

Schwarze, H. & Hertz, J., 1992. Generalization in Large Committee Machine. *Europhysics Letters*, *20*, 375–380.

Sethi, I. K., 1990. Entropy Nets: From Decision Trees to Neural Networks. *Proceedings of IEEE*, *78*.

Seung, H., Sompolinsky, H., & Tishby, N., 1992. Statistical Mechanics of Learning from Examples. *Physical Review A*, *45*, 6056–6091.

Shackelford, G. & Volper, D., 1988. Learning k-DNF with noise in the attributes. In *Proceedings of the First Workshop on Computational Learning Theory*, (pp. 97–103). ACM Press.

Shepard, D., 1968. A two-dimensional interpolation function for irregulary spaced data. In *Proceedings of the 23rd national conference*, (pp. 517–523). New York: ACM.

Shepherd, G. & Brayton, R., 1987. Logic operations are properties of computer-simulated interactions between excitable dendritic spines. *Journal of Neuroscience*, *21*, 151–166.

Shiryayev, A. N., 1978. *Optimal Stopping Rules*. New York: Springer-Verlag.

Shizawa, M., 1993. *Multi-Valued Standard Regularization Theory (2): Regularization Networks and Learning Algorithms for Approximating Multi-Valued Functions*. Technical Report TR-H-045, ATR, Kyoto, Japan.

Singer, M. & Berg, P., 1991. *Genes and Genomes*. Mill Valey: University Science Books.

Skalak, D., 1994. Prototype and feature selection by sampling and random mutation hill-climbing algorithms. In *Proceedings of the Eleventh International Conference on Machine Learning*, (pp. 293–301). New Brunswick, NJ: Morgan Kaufmann.

Sloan, R., 1988. Types of noise in data for concept learning. In *Proceedings of the First Workshop on Computational Learning Theory*, (pp. 91–96). Morgan Kaufmann.

Sompolinsky, H., Tishby, N., & Seung, H., 1990. Learning from Examples in Large Neural Networks. *Physical Review Letters*, *65*, 1683–1686.

Southcott, M. & Bogner, R., 1993. Classification of incomplete data using neural networks. In *Proceedings of the Fourth Australian Conference on Neural Networks (ACNN '93)*, (pp. 220–223). Sydney, Australia: Sydney University Electrical Engineering.

Specht, D., 1967. Generation of polynomial discriminant functions for pattern recognition. *IEEE Transactions on Electronic Computers*, *16*(3), 308–319.

Specht, D. F., 1991. A General Regression Neural Network. *IEEE Transactions on Neural Networks*, *2*(6), 568–576.

Srihari, S. N., 1982. Reliability analysis of majority vote systems. *Information Sciences*, *26*, 243–256.

Stefan, H., Schneider, S., Abraham-Fuchs, K., Bauer, J., Feistel, H., Pawlik, G., Neubauer, U., Roehrlein, G., & Huk, H. J., 1990. Magnetic source localisation in focal epilepsy: Multichannel magnetoencephalography correlated with magnetic resonance brain imaging. *Brain*, *113*.

Stone, M., 1977. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistics Society, Series B, 39*(1), 44–47.

Su, H. & McAvoy, T., 1991. Identification of Chemical processes Using Recurrent Networks. In *Proceedings of the 1991 American Control Conference*. Boston, MA. **Please fill in Publisher!**

Takens, F., 1980. Detecting Strange Attractors in Turbulence. In D. Rand & L. Young (Eds.), *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics* (pp. 366–381). New York: Springer-Verlag.

Tanida, N. & Yokomori, T., 1992. Polynomial-time identification of strictly regular languages in the limit. *IEICE Transactions on Information and Systems*, *E75-D*, 125–132. (IEICE = The Institute of Electronics, Information and Communication Engineers).

Thagard, P. & Holyoak, K. J., 1989. Why indexing is the wrong way to think about analog retrieval. In *Proceedings of a Workshop on Case-based Reasoning*, (pp. 36–40). Pensacola Beach, FL: Morgan Kaufmann.

Tishby, N., Levin, E., & Solla, S. A., 1989. Consistent inference of probabilities in layered networks: predictions and generalization. *Proceedings IEEE International Conference on Neural Networks*, *II*, 403–409.

Towell, G., 1996. Using Unlabeled Data for Supervised Learning. In *Proceedings of the Fourth Annual Workshop on Computational Learning Theory and Natural Learning Systems*. MIT Press. (This volume).

Towell, G. G. & Shavlik, J. W., 1993. Extracting Refined Rules from Knowledge-Based Neural Networks. *Machine Learning*, *13*.

Towell, G. G. & Shavlik, J. W., 1994. Knowledge-Based Artificial Neural Networks. *Artificial Intelligence*, *70*, 119–165.

Townsend-Weber, T. & Kibler, D., 1994. Instance-based prediction of continuous values. In *Working Notes of the AAAI94 Workshop on Case-Based Reasoning*, (pp. 30–35). Seattle, WA: AAAI Press.

Tresp, V., Ahmad, S., & Neuneier, R., 1994. Training neural networks with deficient data. In Cowan, J. D., Tesauro, G., & Alspector, J. (Eds.), *Advances in neural information processing systems 6*. San Francisco, CA: Morgan Kaufmann.

Tresp, V., Hollatz, J., & Ahmad, S., 1993. Network Structuring and Training using Rule-based Knowledge. In Giles, C. L., Hanson, S. J., & Cowan, J. D. (Eds.), *Advances in neural information processing systems 5*. San Mateo, CA: Morgan Kaufmann.

Tresp, V., Leuthaeusser, I., Schlang, M., Neuneier, R., Abraham-Fuchs, K., & Haerer, W., 1992. The Neural Impulse Response Filter. In *Artifical Neural Networks II*. Elsevier.

Turney, P. D., 1995. Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree induction algorithm. *Journal of Artificial Intelligence Research*, *2*, 369–409.

Utgoff, P. E., 1989. Perceptron trees: A case study in hybrid concept representations. *Connection Science*, *1*.

Valiant, L. G., 1984. A theory of the learnable. *Communications of the ACM*, *27*(11), 1134–1142.

Vapnik, V., 1982. *Estimation of Dependences Based on Empirical Data.* New York: Springer-Verlag.

Vapnik, V. & Chervonenkis, A., 1971a. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory of Probability and Its Applications*, *16*, 264–280.

Vapnik, V. N., 1992. Principles of risk minimization for learning theory. *Advances in Neural Information Processing Systems 4*, (pp. 831–838).

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

Veloso, M. M. & Carbonell, J. G., 1993. Derivational analogy in PRODIGY: automating case acquisition, storage and utilization. *Machine Learning*, *10*(3), 249–278.

Vovk, V. G., 1990. Aggregating Strategies. In *Proceedings of Workshop on Computational Learning Theory*, (pp. 371–383). Morgan Kaufmann.

Wagner, R. A. & Seiferas, J., 1978. Correcting counter-automaton-recognizable language. *SIAM Journal of Computing*, *7*, 357–375.

Wald, A., 1947. *Sequential Analysis.* New York: John Wiley & Sons.

Wang, Ming Chen & Uhlenbeck, G., 1945. On the theory of Brownian motion II. *Reviews of Modern Physics*, *17*, 323–342.

Weigend, A., Rumelhart, D., & Huberman, B., 1990. Back-Propagation, Weight Elimination and Time Series Prediction. In *Proceedings of the Connectionist Models Summer School* (pp. 105–116).

Weiss, S. & Kapouleas, I., 1989. An Empirical Comparison of Pattern Recognition, Neural Nets, and Machine Learning Classification Methods. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence.* Morgan Kaufmann.

Werbos, P., 1990. Backpropagation Through Time: What it does and how to do it. *Proceedings of IEEE*, *78*(10), 1550–1560.

Williams, R., 1992. Training Recurrent Networks Using Extended Kalman Filter. In *Proceedings of the International Joint Conference on Neural Networks*, volume 4, (pp. 241–246). Baltimore: IEEE.

Williams, R. & Zipser, D., 1989. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Computation*, *1*, 270–280.

Winston, P. H., 1975. Learning structural descriptions from examples. In P. H. Winston (Ed.), *The Psychology of Computer Vision* (pp. 157–209). McGraw-Hill.

Wolf, A., Swift, J., Swinney, H., & Vastano, J., 1985. Determining Lyapunov Exponents from a Time Series. *Physica D*, *16*, 285–317.

Wolpert, D., 1994. *The relationship between PAC, the statistical physics framework, the Bayesian framework, and the VC framework.* Technical report, The Sante Fe Institute: SFI-TR-03-123.

Wolpert, D. & Miall, R., 1990. Detecting Chaos with Neural Networks. *Philosophical Transactions of the Royal Society of London, B*, *242*, 82–86. **Please fill in Publisher!**

Xu, L. & Yuille, A., 1993a. Robust Principal Component Analysis by Self-organizing rules based on statistical physics approach. *IEEE Transactions on Neural Networks*, *In Press.* **Is this still in press?**

Xu, L. & Yuille, A., 1993b. Self-Organizing Rules for Robust Principal Component Analysis. In Giles, C., Hanson, S., & Cowan, J. (Eds.), *Advances in neural information processing system 5.* Morgan Kaufmann.

Yamanishi, K., 1990. A learning criterion for stochastic rules. In *Proceedings of the Third Workshop on Computational Learning Theory*, (pp. 67–81). Morgan Kaufmann.

Zador, A. M., Claiborne, B. J., & Brown, T. H., 1992. Nonlinear Pattern Separation in Single Hippocampal Neurons with Active Dendritic Membrane. In (Moody et al., 1992), (pp. 51–58).

Zador, A. M. & Pearlmutter, B. A., 1996. VC dimension of an Integrate and Fire Neuron Model. *Neural Computation*, *8*(3). In press.

Zellner, A., 1984. Posterior Odds Ratios for Regression Hypotheses: General Consideration and Some Specific Results. In A. Zellner (Ed.), *Basic Issues in Econometrics* (pp. 275–305). University of Chicago Press.