

References

- Alpaydin, E. and F. Gürgen. 1996. "Comparison of Statistical and Neural Classifiers and Their Applications to Optical Character Recognition and Speech Classification." in *Neural Network Systems: Techniques and Applications* (Leondes, ed.). San Diego, CA: Academic Press.
- Banks, R. B. 1994. *Growth and Diffusion Phenomena: Mathematical Frameworks and Application*. New York: Springer-Verlag.
- Bartlett, P. L. 1997. "For Valid Generalization, the Size of the Weights is More Important than the Size of the Network." in: *Advances in Neural Information Processing Systems Volume 9* (Mozer, Jordan, and Petsche, eds.). Cambridge, MA: The MIT Press.
- Bauer, E. and R. Kohavi. 1999. "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants." *Machine Learning*. 36:105–139.
- Beck, A. 1997. "Herb Edelstein discusses the usefulness of data mining." DS Star. Vol. 1:No. 2. Available at www.tgc.com/dsstar/.
- Berger, J. 1980. *Statistical Decision Theory*. New York: Springer-Verlag.
- Berk, K. N. and D. E. Booth. 1995. "Seeing a Curve in Multiple Regression." *Technometrics* 37:4.
- Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. New York: Oxford University Press.
- Blake, C., E. Keogh, and C. J. Merz. 1998. UCI Repository of machine learning databases. Irvine, CA: University of California, Department of Information and Computer Science. Available at <http://archive.ics.uci.edu/ml/>
- Box, G. E. P. and G. M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day Inc.
- Breiman, L. et al. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.
- Breiman, L. 1996a. "Technical Note: Some Properties of Splitting Criteria." *Machine Learning*. 24:41–47.
- Breiman, L. 1996b. "Bagging Predictors." *Machine Learning*. 24:123–140.
- Breiman, L. 1998. "Arcing Classifiers (with discussion)." *Annals of Statistics*. 26:801–849.
- [BFOS], Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. New York: Chapman & Hall.
- Broomhead, D. S. and David Lowe. 1988. [Radial basis functions, multi-variable functional interpolation and adaptive networks](#) (Technical report). *RSRE*. 4148.
- Broyden, C. G. 1970 "The Convergence of a Class of Double-rank Minimization Algorithms." *Journal of the Institute of Mathematics and Its Applications* 6:76–90.
- Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu. 1995. "A limited memory algorithm for bound constrained optimization." *SIAM Journal on Scientific Computing*. 16, 1190-1208.
- Bryson, A. E. and Y. C. Ho. 1969. *Neural Networks: Computers with Intuition*. Singapore: World Scientific.
- Cai, Z. and C. L. Tsai. 1999. "Diagnostics for Nonlinearity in Generalized Linear Models." *Computational Statistics and Data Analysis* 29:445–469.
- Carroll, R. J. and D. Ruppert. 1988. *Transformation and Weighting in Regression*. New York: Chapman & Hall.

- Chatfield, C. and J. Faraway. 1998. "Time Series Forecasting with Neural Networks: A Comparative Study Using the Airline Data." *Applied Statistics* 47:231–250.
- Chester, D.L. 1990. "Why Two Hidden-Layers are Better than One" in IJCNN-90-WASH-DC. Lawrence Erlbaum. Vol. 1:265–268.
- Cuellar, M.P., M. Delgado, and M.C. Pegalajar. 2006. "An Application of Non-Linear Programming to Train Recurrent Neural Networks in Time Series Prediction Problems." *Enterprise Information Systems VII* (Springer Netherlands) 95–102.
- Cybenko, G. 1988. *Continuous Valued Neural Networks with Two Hidden Layers Are Sufficient*. Technical Report. Dept. of Computer Science. Tufts University: Medford, MA.
- David Shepard Associates. 1999. *The New Direct Marketing: How to Implement a Profit-Driven Database Marketing Strategy*. New York: McGraw-Hill.
- De Vries, B. and J. C. Principe. 1992. "The Gamma Filter - A New Class of Adaptive IIR Filters with Restricted Feedback." *IEEE Transactions*. Signal Processing. In Press.
- De Vries, B. and J. C. Principe. 1992. "The Gamma Model - A New Model for Temporal Processing." *Neural Networks* Vol.5:565–576.
- De Vries, B. and J. C. Principe. 1992. *Short Term Memory Structures for Dynamic Neural Networks*. Princeton, NJ: David Sarnoff Research Center.
- Dorffner, G. 1996. "Neural Networks for Time Series Processing." *Neural Network World* 6:447–468.
- Efron, B. 1983. "Estimating the Error Rate of a Prediction Rule: Improvement on Cross Validation." *Journal of the American Statistical Association* 78:316–331.
- Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani. 2004. "Least Angle Regression (with Discussion)." *Annals of Statistics*, 32: 407–499.
- Ellingsen, B.K. 1994. "A Comparative Analysis of Backpropagation and Counterpropagation Neural Networks." Department of Information Science. University of Bergen: Bergen, Norway.
- Elman, J.L. 1990. "Finding Structure in Time." *Cognitive Science* 14:179–221.
- Ezekiel, M. 1924. "A Method for Handling Curvilinear Correlation for any Number of Variables." *Journal of the American Statistical Association* 19:431–453.
- Fahlman, S. E. 1988 "Faster-learning variations on back-propagation: an empirical study." *Proceedings of the 1988 Connectionist Models Summer School*. (D. Touretzky, G. E. Hinton, and T. J. Sejnowski, eds.).pp. 31–51. San Mateo, CA. Morgan Kaufmann.
- Fahlman, S. E. and C. Lebiere. 1990. "The Cascade-Correlation Learning Architecture." in *Advances in Neural Information Processing Systems Volume 2*. (D Touretzky, ed.). San Mateo, CA: Morgan Kaufmann.
- Fan, J. and I. Gijbels. 1996. *Local Polynomial Modeling and its Applications*. New York: Chapman & Hall.
- Fletcher, R. 1970. "A New Approach to Variable Metric Algorithms." *Computer Journal* 13: 317–322.
- Fletcher, R. 1987. *Practical Methods of Optimization*. New York: Wiley.
- Freund, Y. and R. E. Schapire. 1996. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting." *Journal of Computer and System Science*. 55:119–139.
- Friedman, J. H. 1991. "Multivariate Adaptive Regression Splines (with discussion)." *Annals of Statistics* 19:1–141.

- Friedman, J. H. 1994. "An Overview of Predictive Learning and Function Approximation." In *From Statistics to Neural Networks. Theory and Pattern Recognition Applications*. (Cherkasy, Friedman, Wechsler, eds.) New York: Springer-Verlag.
- Friedman, J. H. 1997. "On Bias, Variance, 0/1-Loss, and the Curse-of-Dimensionality." *Data Mining and Knowledge Discovery* 1:55–77.
- Friedman, J. H. 2001. "Greedy function approximation: A gradient boosting machine." *The Annals of Statistics*. 29:1189–1232.
- Friedman, J. H. 2002. "Stochastic gradient boosting." *Computational Statistics & Data Analysis*. 38:367–378.
- Friedman, J. H. and W. Stuetzle 1981. "Projection Pursuit Regression." *Journal of the American Statistical Association*. 76:817–823.
- Furnival, G. M. and R. W. Wilson. 1974. "Regression by Leaps and Bounds." *Technometrics*, 16, 499–511.
- Georges, J. E. 2003. "Beyond Expectations: Quantifying Variability in Predictive Models." *Proceedings of the M2003 SAS Data Mining Conference*. Cary, NC: SAS Institute Inc.
- Georges, J. E. 2004. "Qualities to Quantities: Using Non-numeric Data in Parametric Prediction." *Proceedings of the M2004 SAS Data Mining Conference*. Cary, NC: SAS Institute Inc.
- Goldfarb, D. 1970. "A Family of Variable Metric Updates Derived by Variational Means." *Mathematics of Computation* 24: 23–24.
- Hand, D. J. 1997. *Construction and Assessment of Classification Rules*. New York: Wiley.
- Hand, D. J. 2005. "What you get is what you want? – Some dangers of black box data mining." *M2005 Conference Proceedings*. Cary, NC: SAS Institute Inc.
- Hand, D. J. 2006. "Classifier technology and the illusion of progress." *Statistical Science* 21:1–14.
- Hand, D. J. and W. E. Henley. 1997. "Statistical classification methods in consumer credit scoring: a review." *Journal of the Royal Statistical Society A*. 160:523–541.
- Hand, David, Heikki Mannila, and Padraic Smyth. 2001. *Principles of Data Mining*. Cambridge, Massachusetts: The MIT Press.
- Harrell F. E., K. L. Lee, and D. B. Mark. 1996. "Multivariate Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors." *Statistics in Medicine* 15:361–387.
- Harrell, F. E. 2006. *Regression Modeling Strategies*. New York: Springer-Verlag New York, Inc.
- Harrison, D. and D. L. Rubinfeld. 1978. "Hedonic Prices in the Demand for Clean Air." *Journal of Environmental Economics and Management* 5:81–102.
- Hassibi, B. and D. G. Stork. 1993. "Second Order Derivatives for Network Pruning: Optimal Brain Surgeon." In *Advances in Neural Information Processing Systems, Volume 5*. (Hansen, Cowan, and Giles, eds.) San Mateo, CA: Morgan Kaufmann.
- Hastie, T. J. and R. J. Tibshirani. 1986. "Generalized Additive Models (with discussion)." *Statistical Science* 1:297–318.
- Hastie, T. J. and R. J. Tibshirani. 1990. *Generalized Additive Models*. New York: Chapman & Hall.
- Hastie, T. J. and R. J. Tibshirani, and Jerome Friedman. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag New York, Inc.
- Hebb, D. O. 1949. *The Organization of Behavior*. New York: Wiley.

- Hecht-Nielsen, R. 1987. "Counterpropagation Networks." *Applied Optics* 26:4979–4984.
- Hertz, J., A. Krogh, and R. G. Palmer. 1991. *Introduction to the Theory of Neural Computation*. Redwood City, CA: Addison-Wesley Publishing Co.
- Hettich, S. and S. D. Bay. 1999. The UCI KDD Archive [<http://kdd.ics.uci.edu>]. Irvine, CA: University of California, Department of Information and Computer Science.
- Hinton, G. E., S. Osindero, and Y. W. Teh. 2006. "A Fast Learning Algorithm for Deep Belief Networks." *Neural Computation* 18. MIT. 1527–1554.
- Hinton, G. E., N. Shrivastava, and K. Swersky. 2013. "Overview of Mini-Batch Gradient Descent" Video from Coursera - University of Toronto - Course: Neural Networks for Machine Learning: Published on Nov 5, 2013. Available at <https://www.coursera.org/course/neura>.
- Hoaglin, D. C., F. Mosteller, and J. W. Tukey. 1983. *Understanding Robust and Exploratory Data Analysis*. New York: Wiley.
- Hoerl, A. E. and R. W. Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12:55–67.
- Hogg, R. V. 1979. "Statistical Robustness: One View of Its Use in Applications Today." *The American Statistician* 33:108–115.
- Holden, K. 1995. "Vector Autoregression Modeling and Forecasting." *Journal of Forecasting* 14:159–166.
- Holland, J. H. 1976. "Studies of Spontaneous Emergence of Self-Replicating Systems Using Cellular Automata and Formal Grammars." In *Automata, Languages, Development*. (A. Lindenmayer and G. Rozenberg, eds.) Amsterdam: North-Holland.
- Holland, J. H. 1992. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, 2nd Edition. Cambridge, MA: The MIT Press.
- Huber, P. J. 1964. "Robust Estimation of a Location Parameter." *Annals of Mathematical Statistics* 35:73–101.
- Jacobs, R. A., M. I. Jordan, S. J. Nowlan, and G. E. Hinton. 1991. "Adaptive Mixture of Local Experts." *Neural Computation* 3:79–87.
- Jaeger, H. 2001. "[The 'echo state' approach to analyzing and training recurrent neural networks](#)". *GMD Report 148*. GMD - German National Research Institute for Computer Science.
- Jaeger, H. 2002. "[Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the echo state network approach](#)". *GMD Report 159*. Fraunhofer Institute AIS.
- Jaeger, H. 2007. "Echo state network." *Scholarpedia* 2(9):2330.
- Johnson, D. E. 1998. *Applied Multivariate Methods for Data Analysis*. Pacific Grove, CA: Druxbury Press.
- Johnson, G. 1991. *In the Palaces of Memory*. Vintage Books: Random House, Inc. NY, NY.
- Jordan, M. I. and R. A. Jacobs. 1994. "Hierarchical Mixture of Experts and the EM algorithm." *Neural Computation* 6:181–214.
- Kass, G. V. 1980. "An exploratory technique for investigating large quantities of categorical data." *Applied Statistics*. 29:119–127.
- Kohavi, R., D. Sommerfield, and J. Dougherty. 1996. "Data Mining using MLC." *International Journal on Artificial Intelligence Tools*. 6:234–245.

- Landwehr, J. D., D. Pregibon, and A. C. Shoemaker. 1984. "Graphical Methods for Assessing Logistic Regression Models (with discussion)." *Journal of the American Statistical Association* 79:61–83.
- Larsen, W. A. and S. J. McCleary. 1972. "The Use of Partial Residual Plots in Regression Analysis." *Technometrics* 14:781–790.
- LeCun, Y., P. Y. Simard, and B. Pearlmutter. 1993. "Autonomic learning rate maximization by on-line estimation of Hessian's eigenvectors" In *Advances in Neural Information Processing Systems* (S. J. Hanson, J. D. Cowan, and C. L. Giles, eds.) Vol. 5:156–163. San Mateo, CA: Morgan Kaufmann.
- LeCun Y., L. Bottou, G. B. Orr, and K. R. Müller. 1998. "Efficient BackProp." In *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science* (G. B. Orr and K. R. Müller, eds). Vol. 1524. Berlin: Springer,
- Leisch, F., A. Trapletti, and K. Hornik. 1999. "Stationarity and Stability of Autoregressive Neural Network Processes." In *Advances in Neural Information Processing Systems 11*. (Kearns, Solla, Cohn, eds.) Cambridge, MA: The MIT Press.
- Loh, W. and Y. Shih. 1997. "Split Selection Methods for Classification Trees." *Statistica Sinica*. 7:815–840.
- Loh, W. and N. Vanichsetakul. 1988. "Tree-Structured Classification Via Generalized Discriminant Analysis (with discussion)." *Journal of the American Statistical Association*. 83:715–728.
- Lukosevicius, M. and H. Jaeger. 2010. "Reservoir Approaches to Recurrent Neural Network Training." *Computer Science Revue* 3(3):127–149.
- Lukosevicius, M. 2012. "A Practical Guide to Applying Echo State Networks." In *Neural Networks: Tricks of the Trade, 2nd ed.* (G. Montavon, G. B. Orr, and K. R. Müller, eds.) Springer LNCS 7700:659–686.
- Maass W., T. Natschlaeger, and H. Markram. 2002. "Real-time computing without stable states: A new framework for neural computation based on perturbations." *Neural Computation*. 14(11):2531–2560.
- Maldonado, M., J. Dean, W. Czika, and S. Haller. 2014. "Leveraging Ensemble Models in SAS® Enterprise Miner™." *Proceedings of the SAS Global Forum 2014 Conference*. Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings14/SAS133-2014.pdf>.
- Masters, T. 1993. *Practical Neural Network Recipes in C++*. Boston: Academic Press Inc. Harcourt Brace & Co. Publishers.
- Melssen, W., R. Wehrens, and L. Buydens. 2006. "Supervised Kohonen networks for classification problems." *Chemometrics and Intelligent Laboratory Systems* 83:99–113.
- McCullagh, P. and J. A. Nelder. 1989. *Generalized Linear Models, Second Edition*. New York: Chapman & Hall.
- McLachlan, G. J. 1992. *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley.
- Minsky, Marvin. 1986. *The Society of Mind*. New York: Simon and Schuster.
- Mitra, S. 2000. "Neuro-Fuzzy Rule Generation: Survey in Soft Computing Framework." *IEEE Transactions on Neural Networks*. Vol. II, No 3 (May).
- Moody, J. 1994. "Prediction Risk and Architecture Selection for Neural Networks." In *From Statistics to Neural Networks. Theory and Pattern Recognition Applications*. (Cherkasy, Friedman, Wechsler, eds.) New York: Springer-Verlag.

- Morgan, J. N. and J. A. Sonquist. 1963. "Problems in the Analysis of Survey Data, and a Proposal." *Journal of the American Statistical Association*. 58:415–434.
- Mosteller, F. and J. W. Tukey. 1977. *Data Analysis and Regression*. Reading, MA: Addison-Wesley.
- Murthy, S. K., S. Kasif, and S. Salzberg. 1994. "A System of Induction of Oblique Decision Trees." *Journal of Artificial Intelligence Research*. 2:1–32.
- Murthy, S. K. and S. Salzberg. 1995. "Lookahead and Pathology in Decision Tree Induction." *Proceedings of IJCAI-95, Montreal*. 1025–1031.
- NC Health and Human Services. North Carolina State Center for Health Statistics. "Statistics and Reports." Available at <http://www.schs.state.nc.us/data/archivedvitalstats.cfm#vol1>.
- Ng, A. 2013. "Stochastic Gradient Descent" Video from Coursera - Stanford University - Course: Machine Learning: Published on Nov 1, 2013. Available at <https://www.coursera.org/course/ml>.
- Olah, C. "Understanding LSTM Networks." Available at <http://colah.github.io/posts/2015-08-Understanding-LSTMs>. Posted Aug 27, 2015. Accessed March 7, 2016.
- Pao, Y. 1989. *Adaptive Pattern Recognition and Neural Networks*. Reading, MA: Addison-Wesley.
- Parker, D. B. 1985. *Learning Logic: Technical Report TR-47*. Center for Computational Research in Economics and Management Science. MIT. Cambridge, MA.
- Patil, G. P. and C. Taillie. 1982. "Diversity as a Concept and its Measurement (with discussion)." *Journal of the American Statistical Association*. 77:548–567.
- Pearlmutter, B. 1995. "Gradient calculations for dynamic recurrent neural networks: A survey." *Neural Networks*. IEEE Transactions. 6(5):1212–1218.
- Piatetsky-Shapiro, G. 1998. "What Wal-Mart might do with Barbie association rules." *Knowledge Discovery Nuggets*. 98:1. Available at <http://www.kdnuggets.com/>.
- Poh, H., J. Yao, and T. Jasic. 1998. "Neural networks for the Analysis and Forecasting of Advertising and Promotion Impact." *International Journal of Intelligent Systems in Accounting, Finance, and Management* 7:253–268.
- Potts, W. J. E. 1999. "Generalized Additive Neural Networks." In *KDD-99 Proceedings*. (Chaudhuri, Madigan, eds.). ACM.
- Prechelt, L. 1996. "A Quantitative Study of Experimental Evaluations of Neural Network Learning Algorithms: Current Research Practice." *Neural Networks* 9:457–462.
- Prechelt, L. 1997. "Investigation of the CasCor Family of Learning Algorithms." *Neural Networks* 10:885–896.
- Prechelt, L. 1998. "Automatic Early Stopping Using Cross Validation: Quantifying the Criteria." *Neural Networks* 11:761–767.
- Principe, J.C., N. R. Euliano, and W. C. Lefebvre. 2000. *Neural and Adaptive Systems*. New York: Wiley.
- Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.
- Raftery, A. E. 1995. "Bayesian Model Selection in Social Research (with discussion)." In *Sociological Methodology 1995* (Marsden ed.) New York: Blackwell.
- Rao, J. S. and W. J. E. Potts. 1997. "Visualizing Bagged Decision Trees." *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining* (Heckerman, Mannila, Pregibon, and Uthurusamy, eds.). Menlo Park, CA: AAAI Press.
- Rendle, S. 2010. "Factorization Machines." The Institute for Scientific and Industrial Research. Osaka University, Japan.

- Rendle, S. 2012. "Factorization Machines with libFM." *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 3, article 57.
- Ridgeway, Greg. 1999. "The State of Boosting." *Computing Science and Statistics*. 31:171–181.
- Riedmiller, M. and H. Braun. 1993. "A direct adaptive method for faster back propagation learning: The RPROP algorithm." *Proceedings of the International Conference on Neural Networks*. San Francisco, CA. pp. 586–591.
- Riedmiller, M. 1994. "Supervised Learning in Multi-layer Perceptrons – From Backpropagation to Adaptive Learning Algorithms." *Int. Journal of Computer Standards and Interfaces* 16.
- Ripley, B. D. 1996. *Pattern Recognition and Neural Networks*. New York: Cambridge University Press.
- Rubinkam, M. 2006. "Internet Merchants Fighting Costs of Credit Card Fraud." AP Worldstream. The Associated Press.
- Rud, Olivia Parr. 2001. *Data Mining Cookbook: Modeling Data, Risk, and Customer Relationship Management*. New York: Wiley.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. "Learning Representations by Back-Propagating Errors." *Nature* 323:533–536.
- Runkle, D. E. 1987. "Vector Autoregressions and Reality (with discussion)." *Journal of Business and Economic Statistics* 5:437–454.
- Sarle, W. S. 1994a. "Neural Networks and Statistical Models." *Proceedings of the Nineteenth Annual SAS® Users Group International Conference*. Cary, NC: SAS Institute Inc. 1538–1550.
- Sarle, W. S. 1994b. "Neural Network Implementation in SAS® Software." *Proceedings of the Nineteenth Annual SAS® Users Group International Conference*. Cary, NC: SAS Institute Inc. 1550–1573.
- Sarle, W. S. 1995. "Stopped Training and Other Remedies for Overfitting." *Proceedings of the 27th Symposium on the Interface*.
- Sarle, W. S., ed. 1997. Neural Network FAQ. URL: <ftp://ftp.sas.com/pub/neural/FAQ.html>.
- Sarle, W. S. 1999. "How to Measure the Importance of Inputs?" URL: <ftp://ftp.sas.com/pub/neural/importance.html>.
- Sarle, W. S. 2000. (revised June 23) "How to measure importance of inputs?" Cary, N.C. [<http://mu.dmt.ibaraki.ac.jp/yanai/neu/faq/importance.html>]
- SAS Institute Inc. 2016. *SAS® Enterprise Miner 14.2: High-Performance Procedures*. Cary, NC: SAS Institute Inc.
- Scheffe, H. 1959. *The Analysis of Variance*. New York: Wiley.
- SCHS: North Carolina State Center for Health Statistics. 2012. "Selected Statistics for 2000 and 1996-00." Available at <http://www.schs.state.nc.us/data/vital/volume1/2000/nc.html>.
- SCHS: North Carolina State Center for Health Statistics. 2003. "Selected Statistics for 2001 and 1997-2001." Available at <http://www.schs.state.nc.us/data/vital/volume1/2001/nc.html>.
- Schwarz, G. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6:461–464.
- Seber, G. A. F. and C. J. Wild. 1989. *Nonlinear Regression*. New York: Wiley.
- Shanno, D. F. 1970. "Conditioning of Quasi-Newton Methods for Function Estimation." *Mathematics of Computation* 24:647–656.

- Shannon, C. E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal*. 27:379–423.
- Suykens, J., B. D. Moor, and J. Vandewalle. 2008. "Toward optical signal processing using photonic reservoir computing." *Optics Express* 16(15):11182–11192.
- Tibshirani, R. and G. E. Hinton. 1995 "Coaching variables for regression and classification." *Statistics and Computing* 8:25–33.
- Tsukimoto, H. 2000. "Extracting Rules from Trained Neural Networks." *IEEE Transactions on Neural Networks*. Vol. II, No. 2 (March).
- Vapnik, V. N. 1995. *The Nature of Statistical Learning*. New York, NY: Springer.
- Vapnik, V., S. Golowich, and A. Smola. 1997. "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing." In *Advances in Neural Information Processing System 9*. 281-287. Also in *Proceedings of the 1996 conference on Neural Information Processing Systems*. MIT Press.
- Vracko, M. 2005. "Kohonen Artificial Neural Network and Counterpropagation Neural Network in Molecular Toxicity Studies." *Current Computer-Aided Drug Design* 1:73–78.
- Wasserman, P. D. 1993. *Advanced Methods in Neural Computing*. New York: Van Nostrand Reinhold.
- Weigend, A. S. and N. A. Gershenfeld. 1994. *Time Series Prediction: Forecasting the Future and Understanding the Past*. Reading, MA: Addison-Wesley.
- Weiss, S. M. and C. A. Kulikowski. 1991. *Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*. San Mateo, CA: Morgan Kaufmann.
- Werbos, P. 1974. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Ph.D. Thesis. Harvard University.
- Wold, H. 1966. "Estimation of Principal Components and Related Models by Iterative Least Squares". In Krishnaiah, P.R. *Multivariate Analysis*. New York: Academic Press. 391–420.
- Wujek, B., P. Hall, and F. Güneş. 2016. "Best Practices for Machine Learning Applications." *Proceedings of the SAS Global Forum 2016 Conference*. Cary, NC: SAS Institute Inc. Available <http://support.sas.com/resources/papers/proceedings16/SAS2360-2016.pdf>
- Zahavi, J. and N. Levin. 1997. "Applying Neural Computing to Target Marketing." *Journal of Direct Marketing* 11:5–22.
- Zhou, Z.-H. 2012. *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL: Chapman & Hall/CRC.