[191] Jabri, M., and B. Flower. Weight perturbation: An optimal architecture and learning technique for analog VLSI feedforward and recurrent multilayer networks. *Neural Computation* 3(4):546–565, 1991.

[192] Jabri, M.A., and B. Flower. Weight perturbation: An optimal architecture and learning technique for analog VLSI feedforward and recurrent multilayer networks. *IEEE Transactions on Neural Networks* 3(1):154–157, 1992.

[193] Jacobs, R.A. Initial experiments on constructing domains of expertise and hierarchies in connectionist systems. In D. Touretzky, G. Hinton, and T. Sejnowski, editors, *Proceedings of the 1988 Connectionist Models Summer School*, pp. 144–1153. Morgan Kaufmann, San Mateo, 1988.

[194] Jacobs, R.A. Increased rates of convergence through learning rate adaptation. *Neural Networks* 1(4):295–307, 1988.

[195] Ji, C., R.R. Snapp, and D. Psaltis. Generalizing smoothness constraints from discrete samples. *Neural Computation* 2(2):188–197, 1990.

[196] Jia, Q., K. Hagiwara, N. Toda, and S.Usui. Equivalence relation between the backpropagation learning process of an FNN and that of an FNNG. *Neural Networks* 7(2):411, 1994.

[197] Johansson, E.M., F.U. Dowla, and D.M. Goodman. Backpropagation learning for multilayer feed-forward neural networks using the conjugate gradient method. *International Journal of Neural Systems* 2(4):291–301, 1991.

[198] Jordan, F., and G. Clement. Using the symmetries of a multi-layered network to reduce the weight space. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 2, pp. 391–396. IEEE, New York, 1991.

[199] Jordan, M.I. Constrained supervised learning. *Journal of Mathematical Psychology* 36(3):396–425, 1992.

[200] Jordan, M.I., and R.A. Jacobs. Modularity, unsupervised learning, and supervised learning. In S. Davis, editor, *Connectionism: Theory and Practice*, pp. 21–29. Oxford University Press, Oxford, 1992.

[201] Jordan, M.I., and D.E. Rumelhart. Forward models: Supervised learning with a distal teacher. *Cognitive Science* 16(3):307–354, 1992.

[202] Judd, J.S., editor. *Neural Network Design and the Complexity of Learning*. MIT Press, Cambridge, 1990.

[203] Judd, S. Learning in neural networks. In D. Haussler and L. Pitt, editors, *Proceedings of the 1988 Workshop on Computational Learning Theory*, pp. 2–8. Morgan Kaufmann, San Mateo, 1988.

[204] Judd, S. On the complexity of loading shallow neural networks. *Journal of Complexity* 4(3):177–192, 1988.

[205] Kamimura, R., T. Takagi, and S. Nakanishi. Improving generalization performance by information minimization. In *IEEE International Conference on Neural Networks (Orlando)*, pp. 143–147. IEEE, New York, 1994.

[206] Kandel, E.R., and J.H. Schwartz, editors. *Principles of Neural Science*, 2nd edition. Elsevier, New York, 1985.

[207] Karnin, E.D. A simple procedure for pruning back-propagation trained neural networks. *IEEE Transactions on Neural Networks* 1(2):239–242, 1990.

[208] Kasparian, V., C. Batur, H. Zhang, and J. Padovan. Davidon least squares based learning algorithm for feedforward neural networks. *Neural Networks* 7(4):661–670, 1994.

[209] Keesing, R., and D. G. Stork. Evolution and learning in neural networks: The number and distribution of learning trials affect the rate of evolution. In R.P. Lippmann, J.E. Moody, and D.S Touretzky, editors, *Advances in Neural Information Processing (Denver)* (3), pp. 804–810. Morgan Kaufmann, San Mateo, 1991.

[210] Kinsella, J.A. Comparison and evaluation of variants of the conjugate gradient method for efficient learning in feed-forward neural networks with backward error propagation. *Network: Computation in Neural Systems* 3(1):27–35, 1992.

[211] Kirkpatrick, S., C.D. Gelatt Jr., and M.P. Vecchi. Optimization by simulated annealing. *Science* 220:671–680, 1983 (reprinted in J.A. Anderson and E. Rosenfeld, editors, *Neurocomputing*, MIT Press, Cambridge, 1988).

[212] Kohonen, T. *Self-Organization and Associative Memory*, vol. 8 of *Springer Series in Information Sciences*, 2nd edition. Springer-Verlag, Berlin, 1988.

[213] Koiran, P., and E.D. Sontag. Neural networks with quadratic VC dimension. In D.S. Touretsky, M.C. Mozer, and M.E. Hasselmo, editors, *Advances in Neural Information Processing Systems* (8), pp. 197–203. MIT Press, Cambridge, 1996.

[214] Koistinen, P., and L. Holmström. Kernel regression and backpropagation training with noise. In *Proceedings of the International Joint Conference on Neural Networks (Singapore)*, pp. 367–372. IEEE, New York, 1991.

[215] Koistinen, P., and L. Holmström. Kernel regression and backpropagation training with noise. In J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems* (4), pp. 1035–1039. Morgan Kaufmann, San Mateo, 1992.

[216] Kolen, J.F., and J.B. Pollack. Backpropagation is sensitive to initial conditions. *Complex Systems* 4(3):269–280, 1990.

[217] Kolen, J.F., and J.B. Pollack. Back propagation is sensitive to initial conditions. In *Advances in Neural Information Processing Systems* (3), pp. 860–867. Morgan Kaufmann, San Mateo, 1991.

[218] Kollias, S., and D. Anastassiou. An adaptive least squares algorithm for the efficient training of multilayered networks. *IEEE Transactions on Circuits and Systems* 36:1092–1101, 1989.

[219] Kolmogorov, A.N. On the representation of continuous functions of several variables by superpositions of continuous functions of one variable and addition. *Doklady Akademii Nauk SSSR* 114(5):953–956, 1957 (in Russian).

[220] Kolmogorov, A.N. On the representation of continuous functions of several variables by superpositions of continuous functions of one variable and addition. *American Mathematical Society Translations* 28:55–59, 1963.

[221] Koza, J. *Genetic Programming*. MIT Press, Cambridge, 1992.

[222] Koza, J.R. A genetic approach to the truck backer upper problem and the inter-twined spiral problem. In *Proceedings of the International Joint Conference on Neural Networks (Baltimore)*, vol. 4, pp. 310–318. IEEE, New York, 1992.

[223] Koza, J.R., and M.A. Keane. Cart centering and broom balancing by genetically breeding populations of control strategy programs. In *Proceedings of the International Joint Conference on Neural Networks (San Diego)*, vol. 1, p. 198. IEEE, New York, 1990.

[224] Koza, J.R., and J.P. Rice. Genetic generation of both the weights and architecture for a neural network. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 2, p. 397. IEEE, New York, 1991.

[225] Kramer, A.H., and A. Sangiovanni-Vincentelli. Efficient parallel learning algorithms for neural networks. In *Advances in Neural Information Processing Systems* (1), pp. 40–48. Morgan Kaufmann, San Mateo, 1989.

[226] Kreinovich, V.Y. Arbitrary nonlinearity is sufficient to represent all functions by neural networks: A theorem. *Neural Networks* 4(3):381–383, 1991.

[227] Krogh, A., and J.A. Hertz. A simple weight decay can improve generalization. In J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems* (4), pp. 950–957. Morgan Kaufmann, San Mateo, 1992.

[228] Kruschke, J.K. Creating local and distributed bottlenecks in hidden layers of back-propagation networks. In D. Touretzky, G. Hinton, and T. Sejnowski, editors, *Proceedings of the 1988 Connectionist Models Summer School*, pp. 120–126, Morgan Kaufmann, San Mateo, 1989.

[229] Kruschke, J.K. Improving generalization in back-propagation networks with distributed bottlenecks. In *Proceedings of the International Joint Conference on Neural Networks (Washington, D.C.)*, vol. 1, pp. 443–447. IEEE, New York, 1989.

[230] Kruschke, J.K., and J.R. Movellan. Benefits of the gain: Speeded learning and minimal hidden layers in back-propagation networks. *IEEE Transactions on Systems, Man, and Cybernetics* 21(1):273–280, 1991.

[231] Kurkova, V. Kolmogorov's theorem and multilayer neural networks. *Neural Networks* 5(3):501–506, 1992.

[232] Kurkova, V., and P.C. Kainen. Functionally equivalent feedforward neural networks. *Neural Computation* 6(3):544–558, 1994.

[233] Lang, K.J., and M.J. Witbrock. Learning to tell two spirals apart. In D. Touretzky, G. Hinton, and T. Se-jnowski, editors, *Proceedings of the 1988 Connectionist Models Summer School*, pp. 52–59. Morgan Kaufmann, San Mateo, 1989.

[234] Lapedes, A., and R. Farber. How neural nets work. In D. Anderson, editor, *Neural Information Processing Systems (Denver 1987)*, pp. 442–456. American Institute of Physics, New York, 1988.

[235] Lari-Najafi, H., M. Nasiruddin, and T. Samad. Effect of initial weights on back-propagation and its variations. In *IEEE International Conference on Systems, Man, and Cybernetics*, vol. 1, pp. 218–219. IEEE, New York, 1989.

[236] Le Cun, Y. Generalization and network design strategies. Technical Report CRG-TR-89-4, University of Toronto, Department of Computer Science, 1989.

[237] Lee, J., and Z. Bien. Improvement of function approximation capability of backpropagation neural net-works. In *Proceedings of the International Joint Conference on Neural Networks (Singapore)*, vol. 2, pp. 1367–1372. IEEE, New York, 1991.

[238] Lee, J.S.-J., J.-N. Hwang, D.T. Davis, and A.C. Nelson. Integration of neural networks and decision tree classifiers for automated cytology screening. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 1, pp. 257–262. IEEE, New York, 1991.

[239] Lee, Y., and R.P. Lippmann. Practical characteristics of neural network and conventional pattern classifiers on artificial and speech problems. In *Advances in Neural Information Processing Systems* (2), pp. 168–177. Morgan Kaufmann, San Mateo, 1990.

[240] Lee, Y., S.-H. Oh, and M.W. Kim. The effect of initial weights on premature saturation in back-propagation learning. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 1, pp. 765–770. IEEE, New York, 1991.

[241] Lee, Y., S.-H. Oh, and M.W. Kim. An analysis of premature saturation in back propagation learning. *Neural Networks* 6(5):719–728, 1993.

[242] Levin, A.U., T.K. Leen, and J.E. Moody. Fast pruning using principal components. In J.D. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems* (6), pp. 35–42. Morgan Kaufmann, San Mateo, 1994.

[243] Levin, E., N. Tishby, and S.A. Solla. A statistical approach to learning and generalization in layered neural networks. *Proceedings of the IEEE* 78(10):1568–1574, Oct. 1990.

[244] Lin, J.-N., and R. Unbehauen. On the realization of a Kolmogorov network. *Neural Computation* 5(1):18–20, 1993.

[245] Lincoln, W.P., and J. Skrzypek. Synergy of clustering multiple backpropagation networks. In D.S. Touret-zky, editor, *Advances in Neural Information Processing Systems (Denver, 1989)* (2), pp. 650–657. Morgan Kauf-mann, San Mateo, 1990.

[246] Linden, A., and J. Kindermann. Inversion of multilayer nets. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 2, pp. 425–430. IEEE, New York, 1989.

[247] Lippmann, R.P. An introduction to computing with neural nets. *ASSP Magazine*, pp. 4–22. April 1987.

[248] Littmann, E., and H. Ritter. Cascade network architectures. In *Proceedings of the International Joint Conference on Neural Networks (Baltimore)*, pp. 398–404. IEEE, New York, 1992.

[249] Lui, H.C. Analysis of decision contour of neural network with sigmoidal nonlinearity. In *Proceedings of the International Joint Conference on Neural Networks (Washington, D.C.)*, pp. 655–658. IEEE, New York, 1990.

[250] Maass, W., G. Schnitger, and E.D. Sontag. On the computational power of sigmoid versus boolean threshold circuits. In *Proceedings of the 32nd Annual Symp. on Foundations of Computer Science*, pp. 767–776. IEEE Computer Society Press, Los Alamitos, CA, 1991.

[251] MacKay, D.J.C. Bayesian interpolation. *Neural Computation* 4(3):415–447, 1992.

[252] MacKay, D.J.C. The evidence framework applied to classification networks. *Neural Computation* 4(5):720–736, 1992.

[253] MacKay, D.J.C. A practical Bayesian framework for backpropagation networks. *Neural Computation* 4(3):448–472, 1992.

[254] Makhoul, J. Pattern recognition properties of neural networks. In B.H. Juang, S.Y. Kung, and C.A. Kamm, editors, *Neural Networks for Signal Processing: Proceedings of the 1991 IEEE Workshop*, pp. 173–187. IEEE Press, New York, 1991.

[255] Makhoul, J., A. El-Jaroudi, and R. Schwartz. Formation of disconnected decision regions with a single hidden layer. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, pp. 455–460. IEEE, New York, 1989.

[256] Marchand, M., M. Golea, and P. Ruján. A convergence theorem for sequential learning in two-layer perceptrons. *Europhysics Letters* 11(6):487–492, 1990.

[257] Marks, R.J., II. *Introduction to Shannon Sampling and Interpolation Theory*. Springer-Verlag, New York, 1991.

[258] Masters, T., editor. *Advanced Algorithms for Neural Networks, A C++ Sourcebook*. Wiley, New York, 1995.

[259] Matsuoka, K. An approach to generalization problem in back-propagation learning. In *International Neural Network Conference (Paris)*, 2:765–768, 1990.

[260] Matsuoka, K. Noise injection into inputs in back-propagation learning. *IEEE Transactions on Systems, Man, and Cybernetics* 22(3):436–440, 1992.

[261] McClelland, J.L., and D.E. Rumelhart. Training hidden units: The generalized delta rule. In *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*, pp. 121–160. MIT Press, Cambridge, 1988.

[262] McInerney, J.M., K.G. Haines, S. Biafore, and R. Hecht-Nielsen. Can back propagation error surfaces have non-global minima? Technical report, Department of Electrical and Computer Engineering, University of California at San Diego, August 1988.

[263] McInerney, J.M., K.G. Haines, S. Biafore, and R. Hecht-Nielsen. kack propagation error surfaces can have local minima. In *Proceedings of the International Joint Conference on Neural Networks (Washington, D.C.)*, vol. 3, pp. 627, 1989 (abstract).

[264] Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemistry and Physics* 21:1087–1091, 1953.

[265] Mézard, M., and J.-P. Nadal. Learning in feedforward layered networks: The tiling algorithm. *Journal of Physics A* 22:2191–2203, 1989.

[266] Minai, A.A., and R.D. Williams. Acceleration of back-propagation through learning rate and momentum adaptation. In *Proceedings of the International Joint Conference on Neural Networks (Washington, D.C.)*, vol. 1, pp. 676–679. IEEE, New York, 1990.

[267] Minnix, J.I. Fault tolerance of the backpropagation neural network trained on noisy inputs. In *Proceedings of the International Joint Conference on Neural Networks (Baltimore)*, vol. 1, pp. 847–852. IEEE, New York, 1992.

[268] Minsky, M., and S. Papert. *Perceptrons, Expanded Edition*. MIT Press, Cambridge, 1988.

[269] Mitchison, G.J., and R.M. Durbin. Bounds on the learning capacity of some multi-layer networks. *Biological Cybernetics* 60:345–356, 1989.

[270] Moller, M.F. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks* 6(4):525–533, 1993.

[271] Moody, J., and C. Darken. Learning with localized receptive fields. In *Proceedings of the 1988 Connectionist Models Summer School*, pp. 133–143. Morgan Kaufmann, San Mateo, 1988.

[272] Moody, J., and C. Darken. Fast learning in networks of locally-tuned processing units. *Neural Computation* 1:281–294, 1989.

[273] Moody, J.E. The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems. In J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems* (4), pp. 847–854. Morgan Kaufmann, San Mateo, 1992.

[274] Moore, B. Theory of networks for learning. In *SPIE* Vol. 1294, *Applications of Artificial Neural Networks*, pp. 22–30. Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, 1990.

[275] Mozer, M.C., and P. Smolensky. Skeletonization: A technique for trimming the fat from a network via relevance assessment. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems (Denver, 1988)* (1), pp. 107–115. Morgan Kaufmann, San Mateo, 1989.

[276] Mukhopadhyay, S., A. Roy, L.S. Kim, and S. Govil. A polynomial time algorithm for generating neural networks for pattern classification: Its stability properties and some test results. *Neural Computation* 5(2):317–330, 1993.

[277] Müller, B., and J. Reinhardt. *Neural Networks, An Introduction*. Springer-Verlag, Berlin, 1990.

[278] Muroga, S. *Threshold logic and its Applications*. Wiley, New York, 1971.

[279] Murphy, O.J. Nearest neighbor pattern classification perceptrons. *Proceedings of the IEEE* 78(10):1595–1598, 1990.

[280] Murphy, O.J. An information theoretic design and training algorithm for neural networks. *IEEE Transaction on Circuits and Systems* 38(12):1542–1547, 1991.

[281] Musavi, M.T., K.H. Chan, D.M. Hummels, and K. Kalantri. On the generalization ability of neural network classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(6):659–663, 1994.

[282] Natarajan, B.K. *Machine Learning, A Theoretical Approach*. Morgan Kaufmann, San Mateo, 1991.

[283] Nguyen, D.H., and B. Widrow. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In *Proceedings of the International Joint Conference on Neural Networks (San Diego)*, vol. 3, pp. 211–226. IEEE, New York, 1990.

[284] Nilsson, N.J. *The Mathematical Foundations of Learning Machines*. Morgan-Kaufmann, San Mateo, 1990 (reprint).

[285] Nowlan, S.J., and G.E. Hinton. Adaptive soft weight tying using Gaussian mixtures. In J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems* (4), pp. 993–1000. Morgan Kaufmann, San Mateo, 1992.

[286] Nowlan, S.J., and G.E. Hinton. Simplifying neural networks by soft weight-sharing. *Neural Computation* 4(4):473–493, 1992.

[287] Oh, S., R.J. Marks, II, and M.A. El-Sharkawi. Query based learning in a multilayered perceptron in the presence of data jitter. In M.A. El-Sharkawi and R.J. Marks, II, editors, *Applications of Neural Networks to Power Systems (Seattle)* pp. 72–75. IEEE Press, New York, 1991.

[288] Oja, E. Neural networks, principal components, and subspaces. *International Journal of Neural Systems* 1(1):61–68, 1089.

[289] Oja, E. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology* 15:267–273, 1982.

[290] Oja, E., and J. Karhunen. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *Journal of Mathematical Analysis and Applications* 106:69–84, 1985.

[291] Okada, H., et al. Initializing multilayer neural networks with fuzzy logic. In *Proceedings of the International Joint Conference on Neural Networks (Baltimore)*, vol. 1, pp. 239–244. IEEE, New York, 1992.

[292] Oppenheim, A.V., A.S. Willsky, and I.T. Young. *Signals and Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1983.

[293] Pados, D.A., and P. Papantoni-Kazakos. A note on the estimation of the generalization error and the prevention of overfitting. In *IEEE International Conference on Neural Networks (Orlando)*, vol. 1, pp. 321–325. IEEE, New York, 1994.

[294] Palubinskas, G. Data-driven weight initialization of back-propagation for pattern recognition. In *Proceedings of the International Conferance on Artificial Neural Networks (ICANN'94)*, pp. 851–854. Springer-Verlag, London, 1994.

[295] Parker, D.B. Learning logic. *Technical Report* TR-47. Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Cambridge, 1985.

[296] Parker, D.B. Optimal algorithms for adaptive networks: Second-order back propagation, second-order direct propagation, and second-order Hebbian learning. In *Proceedings of the IEEE First International Conference on Neural Networks (San Diego)*, vol. 2, pp. 593–600. IEEE, New York, 1987.

[297] Pearlmutter, B.A. Fast exact multiplication by the Hessian. *Neural Computation* 6(1):147–160, 1994.

[298] Pearlmutter, B.A., and R. Rosenfeld. Chaitin-Kolmogorov complexity and generalization in neural networks. In R. Lippmann, J. Moody, and D. Touretzky, editors, *Advances in Neural Information Processing Systems (3)*, pp. 925–931. Morgan Kaufmann, San Mateo, 1991.

[299] Plaut, D.C., S.J. Nowlan, and G.E. Hinton. Experiments on learning by back propagation. *Technical Report* CMU-CS-86-126, Carnegie-Mellon University, Pittsburgh, 1986.

[300] Poggio, T., and F. Girosi. Networks for approximation and learning. *Proceedings of the IEEE* 78(9):1481–1497, Sept. 1990.

[301] Poston, T., C.-N. Lee, Y. Choie, and Y. Kwon. Local minima and back propagtion. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 2, pp. 173–176. IEEE, New York, 1991.

[302] Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988.

[303] Raudys, S., and M. Skurikhina. The role of the number of training samples on weight initialization of artificial neural net classifier. In *RNNS/IEEE Symposium on Neuroinformatics and Neurocomputing (Rostov-on-Don, Russia)*, pp. 343–353. IEEE Press, New York, 1992.

[304] Rechenberg, I. Artificial evolution and artificial intelligence. In R. Forsyth, editor, *Machine Learning, Principles and Techniques*, pp. 83–103. Chapman and Hall Computing, New York, 1989.

[305] Reed, R., R.J. Marks, II, and S. Oh. An equivalence between sigmoidal gain scaling and training with noisy (jittered) input data. In *RNNS/IEEE Symposium on Neuroinformatics and Neurocomputing (Rostov-on-Don, Russia)*, pp. 120–127. IEEE Press, New York, 1992.

[306] Reed, R., R.J. Marks, II, and S. Oh. Similarities of error regularization, sigmoid gain scaling, target smoothing, and training with jitter. *IEEE Transactions on Neural Networks* 6(3):529–538, May 1995.

[307] Reed, R., S. Oh, and R.J. Marks, II. Regularization using jittered training data. In *Proceedings of the International Joint Conference on Neural Networks (Baltimore)*, vol. 3, pp. 147–152. IEEE, New York, 1992.

[308] Reed, R.D. Pruning algorithms—a survey. *IEEE Transactions on Neural Networks*, 4(5):740–744, 1993.

[309] Reed, R.D., J.E. Sanders, and R.J. Marks, II. Neural network aided prosthetic alignment. In *1995 IEEE International Conference on Systems, Man, and Cybernetics, Vancouver, British Columbia, Oct. 1995*, vol. 1, pp. 505–508, IEEE, New York, 1995.

[310] Refregier, Ph. and J.-M. Vignolle. An improved version of the pseudo-inverse solution for classification and neural networks. *Europhysics Letters* 10(4):387–392, 1989.

[311] Reklaitis, G.V., A. Ravindran, and K.M. Ragsdell. *Engineering Optimization, Methods and Applications*. Wiley, 1983.

[312] Rezgui, A., and Nazif Tepedelenlioglu. The effect of the slope of the activation function on the back propagation algorithm. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, pp. 707–710. IEEE, New York, 1990.

[313] Ricotti, L.P., S. Ragazzini, and G. Martinelli. Learning of word stress in a sub-optimal second order back-propagation neural network. In *Proceedings of the IEEE International Conference on Neural Networks (San Diego)*, vol. 1, pp. 355–361. IEEE, New York, 1988.

[314] Riedmiller, M. Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms. *Computer Standards & Interfaces* 16, 1994.

[315] Riedmiller, M., and H. Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *IEEE International Conference on Neural Networks (San Francisco)*, vol. 1, pp. 586–591. IEEE, New York, 1993.

[316] Rigler, A.K., J.M. Irvine, and T.P. Vogl. Rescaling of variables in back propagation learning. *Neural Networks* 4(2):225–229, 1991.

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

[318] Robbins, H., and S. Monro. A stochastic optimization method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

[319] Rogers, D. Predicting weather using a genetic memory: a combination of Kanerva's sparse distributed memory with Holland's genetic algorithms. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems* (2), pp. 455–464. Morgan Kaufmann, San Mateo, 1989.

[320] Rohwer, R. Time trials on second-order and variable-learning-rate algorithms. In *Advances in Neural Information Processing Systems* (3), pp. 977–983. Morgan Kaufmann, San Mateo, 1991.

[321] Rojas, R. Optimal weight initialization for neural networks. In *Proceedings of the International Conference on Artificial Neural Networks (ICANN'94)*, pp. 577–580. Springer-Verlag, London, 1994.

[322] Romaniuk, S.G., and L.O. Hall. Dynamic neural networks with the use of divide and conquer. In *Proceedings of the International Joint Conference on Neural Networks (Baltimore)*, vol. 1, pp. 658–663. IEEE, New York, 1992.

[323] Röscheisen, M., R. Hofmann, and V. Tresp. Neural control for rolling mills: Incorporating domain theories to overcome data deficiency. In J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems* (4), pp. 659–666. Morgan Kaufmann, San Mateo, 1992.

[324] Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review* 65:386–408, 1958 (reprinted in J.A. Anderson and E. Rosenfeld, editors, *Neurocomputing*. MIT Press, Cambridge, 1988).

[325] Rosenblatt, F., editor. *Principles of Neurodynamics*. Spartan Books, Washington, D.C., 1962.

[326] Roy, S., and J.J. Shynk. Analysis of the momentum LMS algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processsing* 38:2088–2098, 1990.

[327] Ruderman, D.L., and W. Bialek. Seeing beyond the Nyquist limit. *Neural Computation* 4(5):682–690, 1992.

[328] Rumelhart, D.E., R. Durbin, R. Golden, and Y. Chauvin. Backpropagation: The basic theory. In Y. Chauvin and D.E. Rumelhart, editors, *Backpropagation: Theory, Architectures, and Applications*, chapter 1. Erlbaum, Hillsdale, NJ, 1994.

[329] Rumelhart, D.E., G.E. Hinton, and R.J. Williams. Learning internal representations by back-propagating errors. *Nature* 323:533–536, 1986 (reprinted in J.A. Anderson and E. Rosenfeld, editors, *Neurocomputing*. MIT Press, Cambridge, 1988).

[330] Rumelhart, D.E., G.E. Hinton, and R.J. Williams. Learning internal representations by error propagation. In D.E. Rumelhart and J.L. McClelland, editors, *Parallel Distributed Processing*, Vol. I, chapter 8. MIT Press, Cambridge, 1986 (reprinted in J.A. Anderson and E. Rosenfeld, editors, *Neurocomputing*. MIT Press, Cambridge, 1988).

[331] Saarinen, S., R. Bramley, and G. Cybenko. Ill-conditioning in neural network training problems. *SIAM Journal of Scientific Computing* 14(3):693–714, 1993.

[332] Samad, T. Backpropagation improvements based on heuristic arguments. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, pp. 565–568. IEEE, New York, 1990.

[333] Sanger, T.D. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks* 2(6):459–473, 1989.

[334] Sanger, T.D. An optimality principle for unsupervised learning. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems* (1), pp. 11–19. Morgan Kaufmann, San Mateo, 1989.

[335] Sankar, A., and R.J. Mammone. Optimal pruning of neural tree networks for improved generalization. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 2, pp. 219–224. IEEE, New York, 1991.

[336] Schreibman, D.V., and E.M. Norris. Speeding up back propagation by gradient correlation. In *Proceedings of the International Joint Conference on Neural Networks (Washington, D.C.)*, vol. 1, pp. 723–726. IEEE, New York, 1990.

[337] Schwartz, D.B., V.K. Samalan, S.A. Solla, and J.S. Denker. Exhaustive learning. *Neural Computation* 2(3):374–385, 1990.

[338] Segee, B.E., and M.J. Carter. Fault tolerance of pruned multilayer networks. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 2, pp. 447–452. IEEE, New York, 1991.

[339] Séquin, C.H., and R.D. Clay. Fault tolerance in feed-forward artificial neural networks. In Paolo Antognetti and Veljko Milutinović, editors, *Neural Networks: Concepts, Applications, and Implementations*, vol. 4, pp. 111–141. Prentice–Hall, Englewood Cliffs, NJ, 1991.

[340] Sethi, I.K. Entropy nets: From decision trees to neural networks. *Proceedings of the IEEE* 78(10):1605–1613, Oct. 1990.

[341] Shanno, D.F. Recent advances in numerical techniques for large scale optimization. In W.T. Miller, III, R.S. Sutton, and P.J. Werbos, editors, *Neural Networks for Control*, pp. 171–178. MIT Press, Cambridge, 1990.

[342] Shavlik, J.W. A framework for combining symbolic and neural learning. In V. Honavar and L. Uhr, editors, *Artificial Intelligence and Neural Networks: Steps Toward Principled Integration*, pp. 561–580. Academic Press, New York, 1994.

[343] Shynk, J.J., and S. Roy. The LMS algorithm with momentum updating. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 2651–2654. IEEE, New York, 1988.

[344] Sietsma, J., and R.J.F. Dow. Neural net pruning—why and how. In *Proceedings of the IEEE International Conference on Neural Networks (San Diego)*, vol. 1, pp. 325–333. IEEE, New York, 1988.

[345] Sietsma, J., and R.J.F. Dow. Creating artificial neural networks that generalize. *Neural Networks* 4(1):67–79, 1991.

[346] Silva, F.M., and L.B. Almeida. Acceleration techniques for the backpropagation algorithm. In L.B. Almeida and C.J. Wellekens, editors, *Neural Networks, Proceedings EURASIP Workshop*, vol. 412 of *Lecture Notes in Computer Science*, pp. 110–119. Springer-Verlag, New York, 1990.

[347] Siu, K.-Y., V. Roychowdhury, and T. Kailath. *Discrete Neural Computation, A Theoretical Foundation*. Prentice-Hall, Englewood Cliffs, NJ, 1995.

[348] Siu, K.-Y., V.P. Roychowdhury, and T. Kailath. Depth-size tradeoffs for neural computation. *IEEE Transactions on Computers* 40(12):1402–1412, 1991.

[349] Smyth, S.G. Designing multi layer perceptrons from nearest neighbor systems. *IEEE Transactions on Neural Networks* 3(2):329–333, 1992.

[350] Sontag, E.D. Feedback stabilization using two-hidden-layer nets. *Technical Report* SYCON-90-11, Rutgers Center for Systems and Control, Princeton, NJ, 1990.

[351] Sontag, E.D. Feedback stabilization using two-hidden-layer nets. *IEEE Transactions on Neural Networks* 3(6):981–990, 1992.

[352] Sontag, E.D., and H.J. Sussmann. Backpropagation can give rise to spurious local minima even for networks without hidden layers. *Complex Systems* 3:91–106, 1989.

[353] Sontag, E.D., and H.J. Sussmann. Backpropagation separates when perceptrons do. In *Proceedings of the International Joint Conference on Neural Networks (Washington, D.C.)*, vol. 1, pp. 639–642. IEEE Press, New York, 1989.

[354] Sperduti, A., and A. Starita. Speed up learning and network optimization with extended back propagation. *Neural Networks* 6(3):365–383, 1993.

[355] Sprecher, D.A. On the structure of continuous functions of several variables. *Transactions of the American Mathematical Society* 115(3):340–355, 1965.

[356] Sprecher, D.A. A universal mapping for Kolmogorov's superposition theorem. *Neural Networks* 6(8):1089–1094, 1993.