

This excerpt from

Neural Smithing.
Russell D. Reed and Robert J. Marks II.
© 1999 The MIT Press.

is provided in screen-viewable form for personal use only by members of MIT CogNet.

Unauthorized use or dissemination of this information is expressly forbidden.

If you have any questions about this material, please contact
cognetadmin@cognet.mit.edu.

References

- [1] Abu-Mostafa, Y.S. The Vapnik-Chervonenkis dimension: Information versus complexity in learning. *Neural Computation* 1(3):312–317, 1989.
- [2] Achenie, L.E.K. Computational experience with a quasi Newton based training of the feedforward neural network. In *World Congress on Neural Networks (San Diego)*, vol. 3, pp. 607–612. Lawrence Erlbaum, Hillsdale, NJ, 1994.
- [3] Ackley, D. H. *A Connectionist Machine for Genetic Hillclimbing*. Kluwer, Boston, 1987.
- [4] Ackley, D. H. and M. S. Littman. Learning from natural selection in an artificial environment. In *Proceedings of the International Joint Conference on Neural Networks (San Diego)*, vol. 1, p. 189, IEEE, New York, 1990.
- [5] Ahmad, S., and G. Tesauro. Scaling and generalization in neural networks: a case study. In D. Touretzky, G. Hinton, and T. Sejnowski, editors, *Proceedings of the 1988 Connectionist Summer School*, pp. 3–10. Morgan Kaufmann, San Mateo, 1988.
- [6] Akaike, H. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* 21:243–247, 1969.
- [7] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC-19(6):716–723, 1974.
- [8] Alpsan, D., M. Towsey, O. Ozdamar, A. Tsoi, and D.N. Ghista. Are modified back-propagation algorithms worth the effort? In *IEEE International Conference on Neural Networks (Orlando)*, vol. 1, pp. 567–571. IEEE, New York, 1994.
- [9] Alpsan, D., M. Towsey, O. Ozdamar, A.C. Tsoi, and D.N. Ghista. Efficacy of modified backpropagation and optimisation methods on a real-world medical problem. *Neural Networks* 8(6):945–962, 1995.
- [10] Arabshahi, P., J.J. Choi, R.J. Marks, II, and T.P. Caudell. Fuzzy control of backpropagation. In *Proceedings of the First IEEE International Conference on Fuzzy Systems, (FUZZ-IEEE '92), San Diego*, pp. 967–972. IEEE, New York, 1992.
- [11] Ash, T. Dynamic node creation in back-propagation networks. *Technical Report 8901*. Institute for Cognitive Science, UCSD, La Jolla, 1989.
- [12] Atlas, L., D. Cohn, R. Ladner, M.A. El-Sharkawi, R.J. Marks II, M.E. Aggoune, and D.C. Park. Training connectionist networks with queries and selective sampling. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems (Denver)* (2), pp. 566–573, Morgan Kaufmann, San Mateo, 1990.
- [13] Auer, P., M. Herbster, and M.K. Warmuth. Exponentially many local minima for single neurons. In D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, editors, *Advances in Neural Information Processing Systems* (8), pp. 316–322. MIT Press, Cambridge, 1996.
- [14] Aylward, S., D.C. St. Clair, W. Bond, B. Flachsbart, and A.K. Rigler. One-dimensional search strategies for conjugate gradient training of backpropagation neural networks. In *Proceedings of the Artificial Neural Networks in Engineering (ANNIE'92) Conference, St. Louis*, vol. 2, pp. 192–202. IEEE, New York, 1992.
- [15] Bailey, A.W. Automatic evolution of neural net architectures. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, pp. 589–592. IEEE, New York, 1990.
- [16] Baldi, P., and Y. Chauvin. Temporal evolution of generalization during learning in linear networks. *Neural Computation* 3(4):598–603, 1991.
- [17] Baldi, P., Y. Chauvin, and K. Hornik. Backpropagation and unsupervised learning in linear networks. In Y. Chauvin and D.E. Rumelhart, editors, *Backpropagation: Theory, Architectures, and Applications*, chapter 12, pp. 389–432. Erlbaum, Hillsdale, NJ, 1994.
- [18] Baldi, P., and K. Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks* 2(1):53–58, 1989.
- [19] Barnard, E. Optimization for training neural nets. *IEEE Transactions on Neural Networks* 3(2):232–240, 1992.
- [20] Barnard, E., and J.E.W. Holm. A comparative study of optimization techniques for backpropagation. *Neuro-computing* 6:19–30, 1994.

- [21] Barron, A.R. Approximation bounds for superpositions of a sigmoidal function. In *Proceedings of the IEEE International Symposium on Information Theory*. IEEE Press, New York, 1991.
- [22] Barron, A.R. Universal approximation bounds of a sigmoidal function. *Technical Report 58*. Statistics Department, University of Illinois, Urbana-Champaign, 1991.
- [23] Barron, A.R. Universal approximation bounds for superposition of a sigmoidal function. *IEEE Transactions on Information Theory* 39(3):930–945, 1993.
- [24] Barto, A.G. Reinforcement learning and adaptive critic methods. In D.A. White and D.A. Sofge, editors, *Handbook of Intelligent Control*, chapter 12. Van Nostrand Reinhold, New York, 1992.
- [25] Barto, A.G., R.S. Sutton, and C.W. Anderson. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-13(5):834–846, 1983 (reprinted in J.A. Anderson and E. Rosenfeld, editors, *Neurocomputing*. 1988).
- [26] Battiti, R. Accelerated backpropagation learning: Two optimization methods. *Complex Systems*, 3:331–342, 1989.
- [27] Battiti, R. Optimization methods for back-propagation: Automatic parameter tuning and faster convergence. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, 593–596. IEEE, New York, 1990.
- [28] Battiti, R. First- and second-order methods for learning: Between steepest descent and Newton’s method. *Neural Computation* 4(2):141–166, 1992.
- [29] Battiti, R., and G. Tecchiolli. Learning with first, second, and no derivatives: A case study in high energy physics. *Neurocomputing* 6:181–206, 1994.
- [30] Baum, E.B. On the capabilities of multilayer perceptrons. *Journal of Complexity* 4:193–215, 1988.
- [31] Baum, E.B. A proposal for more powerful learning algorithms. *Neural Computation* 1(2):201–207, 1989.
- [32] Baum, E.B. When are k-nearest neighbor and back propagation accurate for feasible sized sets of examples? In L.B. Almeida and C.J. Wellekens, editors, *Neural Networks, Proceedings EURASIP Workshop*, vol. 412 of *Lecture Notes in Computer Science*, pp. 2–25. Springer-Verlag, New York, February 1990.
- [33] Baum, E.B., and D. Haussler. What size net gives valid generalization? In David S Touretzky, editor, *Advances in Neural Information Processing Systems* (1), pp. 81–90. Morgan Kaufmann, San Mateo, 1989.
- [34] Baum, E.B., and D. Haussler. What size net gives valid generalization? *Neural Computation* 1:151–160, 1989.
- [35] Baum, E.B., and R.L. Rivest. Training a 3-node neural network is NP-complete. In *Proceedings of the 1988 Workshop on Computational Learning Theory*, pp. 9–18. Morgan Kaufmann, San Mateo, 1988.
- [36] Baxt, W.G. Improving the accuracy of an artificial neural network using multiple differently trained networks. *Neural Computation* 4(5):772–780, 1992.
- [37] Becker, S., and Y. Le Cun. Improving the convergence of back-propagation learning with second order methods. In D. Touretzky, G. Hinton, and T. Sejnowski, editors, *Proceedings of the 1988 Connectionist Models Summer School*, pp. 29–37. Morgan Kaufmann, San Mateo, 1988.
- [38] Bello, M.G. Enhanced training algorithms, and integrated training/architecture selection for multilayer perceptrons. *IEEE Transactions on Neural Networks*, 3(6):864–875, 1992.
- [39] Bishop, C. Exact calculation of the Hessian matrix for the multilayer perceptron. *Neural Computation* 4(4):494–501, 1992.
- [40] Bishop, C.M. Curvature-driven smoothing in backpropagation neural networks. *IEEE Transactions on Neural Networks* 4:882–884, 1993.
- [41] Bishop, C.M. A fast procedure for retraining the multilayer perceptron. *International Journal of Neural Systems* 2(3):229–236, 1991.
- [42] Bishop, C.M. Improving the generalization properties of radial basis function neural networks. *Neural Computation* 3(4):579–588, 1991.

- [43] Bishop, C.M. Curvature-driven smoothing: A learning algorithm for feedforward networks. *IEEE Transactions on Neural Networks* 4(5):882–884, 1993.
- [44] Bishop, C.M. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, 1995.
- [45] Bishop, C.M. Training with noise is equivalent to Tikhonov regularization. *Neural Computation*, 7(1):108–116, 1995.
- [46] Block, H.D. The perceptron: A model for brain functioning. I. *Reviews of Modern Physics*, 34:123–135, 1962. (Reprinted in J.A. Anderson and E. Rosenfeld, editors, *Neurocomputing*. MIT Press, Cambridge, 1988).
- [47] Blue, J. Training feed-forward neural networks using conjugate gradients. In *SPIE*, vol. 1661, p. 179. Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, 1992.
- [48] Blum, E.K., and L.K. Li. Approximation theory and feedforward networks. *Neural Networks* 4(4):511–515, 1991.
- [49] Blumer, A., A. Ehrenfeucht, D. Haussler, and M. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the Association for Computing Machinery* 36(4):929–965, 1989.
- [50] Booker, L. Improving search in genetic algorithms. In L. Davis, editor, *Genetic Algorithms and Simulated Annealing*. Morgan Kaufmann, San Mateo, 1990.
- [51] Booker, L. Using classifier systems to implement distributed representations. In *Proceedings of the International Joint Conference on Neural Networks (San Diego)*, vol. 1, p. 39. IEEE, New York, 1990.
- [52] Booker, L.B., D.E. Goldberg, and J.H. Holland. Classifier systems and genetic algorithms. In J. Carbonell, editor, *Machine Learning: Paradigms and Methods*. MIT Press, Cambridge, 1990.
- [53] Bose, N.K., and A.K. Garga. Neural network design using Voronoi diagrams: Preliminaries. In *Proceedings of the International Joint Conference on Neural Networks (Baltimore)*, vol. 3, pp. 127–132. IEEE Press, New York, 1992.
- [54] Bose, N.K., and A.K. Garga. Neural network design using Voronoi diagrams. *IEEE Transactions on Neural Networks* 4(5):778–787, 1993.
- [55] Bose, N.K., and P. Liang. *Neural Network Fundamentals with Graphs, Algorithms, and Applications*. McGraw-Hill, New York, 1996.
- [56] Botros, S.H., and C.G. Atkeson. Generalization properties of radial basis functions. In R. P. Lippmann, J. E. Moody, and D. S. Touretzky, editors, *Advances in Neural Information Processing Systems* (3), pp. 707–713. Morgan Kaufmann, San Mateo, 1991.
- [57] Boulard, H., and Y. Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics* 59:291–294, 1988.
- [58] Brady, M.L., and R. Raghavan. Gradient descent fails to separate. In *Proceedings of the IEEE International Conference on Neural Networks (San Diego)*, vol. 1, pp. 649–656. IEEE, New York, 1988.
- [59] Brady, M.L., R. Raghavan, and J. Slawny. Back propagation fails to separate where perceptrons succeed. *IEEE Transactions on Circuits and Systems* 36:665–674, 1989.
- [60] Breiman, L. Current research. In D.H. Wolpert, editor, *The Mathematics of Generalization*, pp. 361–368. Addison-Wesley, Reading, MA, 1995.
- [61] Breiman, L. Reflections after refereeing papers for NIPS. In D.H. Wolpert, editor, *The Mathematics of Generalization*, pp. 11–15. Addison-Wesley, Reading, MA, 1995.
- [62] Buntine, W.L., and A.S. Weigend. Computing second derivatives in feed-forward networks: A review. *IEEE Transactions on Neural Networks* 5(3):480–488, 1994.
- [63] Carpenter, G.A., S. Grossberg, N. Markuzon, J.H. Reynolds, and D.B. Rosen. Fuzzy ARTMAP: An adaptive resonance architecture for incremental learning of analog maps. In *Proceedings of the International Joint Conference on Neural Networks (Baltimore)*, vol. 3, pp. 309–314. IEEE, New York, 1992.
- [64] Carroll, S.M., and B.W. Dickson. Construction of neural nets using the Radon transform. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, pp. 607–611. IEEE, New York, 1989.

- [65] Cater, J.P. Successfully using peak learning rates of 10 (and greater) in back-propagation networks with the heuristic learning algorithm. In *Proceedings of the IEEE First International Conference on Neural Networks (San Diego)*. IEEE, New York, 1987.
- [66] Caudill, M. Evolutionary neural networks. *AI Expert*, March 1991.
- [67] Chan, L.-W., and F. Fallside. An adaptive training algorithm for back propagation networks. *Computer Speech and Language* 2:205–218, 1987.
- [68] Chang, E. I., and R. P. Lippmann. Using genetic algorithms to improve pattern classification performance. In R.P. Lippmann, J.E. Moody, and D.S. Touretzky, editors, *Advances in Neural Information Processing (Denver, 1990)* (3), pp. 797–803. Morgan Kaufmann, San Mateo, 1991.
- [69] Charalambous, C. Conjugate gradient algorithm for efficient training of artificial neural networks. *IEEE Proceedings* 139(3):301–310, 1992.
- [70] Chauvin, Y. A back-propagation algorithm with optimal use of hidden units. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems (Denver, 1988)* (1), pp. 519–526. Morgan Kaufmann, San Mateo, 1989.
- [71] Chauvin, Y. Principal component analysis by gradient descent on a constrained linear Hebbian cell. In *Proceedings of the International Joint Conference on Neural Networks (Washington, D.C.)*, vol. 1, pp. 373–380. IEEE, New York, 1989.
- [72] Chauvin, Y. Dynamic behavior of constrained back-propagation networks. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems* (2), pp. 642–649. Morgan Kaufmann, San Mateo, 1990.
- [73] Chauvin, Y. Generalization performance of overtrained back-propagation networks. In L.B. Almeida and C.J. Wellekens, editors, *Neural Networks, Proceedings EUROSIP Workshop*, vol. 412 of *Lecture Notes in Computer Science*, pp. 46–55. Springer-Verlag, New York, Feb. 1990.
- [74] Chauvin, Y. Generalization dynamics in LMS trained linear networks. In R. Lippmann, J. Moody, and D. Touretzky, editors, *Advances in Neural Information Processing Systems* (3), pp. 890–896. Morgan Kaufmann, San Mateo, 1991.
- [75] Chen, C.H., and H. Lai. A comparison study of the gradient descent and the conjugate gradient backpropagation neural networks. In *World Congress on Neural Networks (Portland)*, vol. 3, pp. 401–406. Erlbaum, Hillsdale, NJ, 1993.
- [76] Chen, J.R., and P. Mars. Stepsize variation methods for accelerating the back propagation algorithm. In *Proceedings of the International Joint Conference on Neural Networks (Washington, D.C.)*, vol. 1, pp. 601–604. IEEE, New York, 1990.
- [77] Chen, O.T.-C., and B.J. Sheu. Optimization schemes for neural network training. In *IEEE International Conference on Neural Networks (Orlando)*, vol. 2, pp. 817–822. IEEE, New York, 1994.
- [78] Chen, S., S.A. Billings, and R. Hecht-Nielsen. On the geometry of feedforward neural network error surfaces. *Neural Computation* 5(6):910–927, 1993.
- [79] Cherkassky, V., and F. Mulier. Statistical and neural network techniques for nonparametric regression. In *Selecting Models from Data*, Lecture Notes in Statistics 89, pp. 383–392. Springer Verlag, New York, 1994.
- [80] Chester, D.L. Why two hidden layers are better than one. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, pp. 265–268. IEEE, New York, 1990.
- [81] Choi, J. J., P. Arabshahi, R. J. Marks II, and T. P. Caudell. Fuzzy parameter adaptation in neural systems. In *Proceedings of the International Joint Conference on Neural Networks (Baltimore)*, vol. 1, pp. 232–235. IEEE, New York, 1992.
- [82] Cichocki, A., and R. Unbehauen. *Neural Networks for Optimization and Signal Processing*. Wiley, 1993.
- [83] Clemen, R.T. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5(4):559–583, 1989.
- [84] Codrington, C.W., and M.F. Tenorio. Adaptive gain networks. In *IEEE International Conference on Neural Networks (Orlando)*, vol. 1, pp. 339–344. IEEE, New York, 1994.

- [85] Cohn, D., and G. Tesauro. Can neural-networks do better than the Vapnik–Chervonenkis bounds? In R.P. Lippmann, J.E. Moody, and D.S. Touretzky, editors, *Advances in Neural Information Processing Systems (Denver, 1990)* (3), pp. 911–917. Morgan Kaufmann, San Mateo, 1991.
- [86] Corwin, E.M., A.M. Logar, and W.J.B. Oldham. An iterative method for training multilayer networks with threshold functions. *IEEE Transactions on Neural Networks* 5(3):507–508, 1994.
- [87] Cover, T. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, 14:326–334, 1965.
- [88] Crowder, III, R.S. Predicting the Mackey-Glass timeseries with cascade-correlation learning. In D.S. Touretzky, J.L. Elman, T.J. Sejnowski, G.E. Hinton, editors, *Connectionist Models: Proceedings of the 1990 Summer School*, pp. 117–123. Morgan Kaufmann, San Mateo, 1991.
- [89] Le Cun, Y. Une procédure d’apprentissage pour réseau à seuil asymétrique. In *COGNITIVA 85: A la Frontière de l’Intelligence Artificielle des Science de la Connaissance des Neurosciences (Paris)*, pp. 599–604. CESTA, Paris, 1985.
- [90] Le Cun, Y. Learning process in an asymmetric threshold network. In E. Bienenstock, F. Fogelman Soulié, and G. Weisbuch, editors, *Disordered Systems and Biological Organization*, pp. 233–240. Springer-Verlag, New York, 1986.
- [91] Le Cun, Y., J.S. Denker, and S.A. Solla. Optimal brain damage. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems (Denver)* (2), pp. 598–605. Morgan Kaufmann, San Mateo, 1990.
- [92] Le Cun, Y., I. Kanter, and S.A. Solla. Eigenvalues of covariance matrices: Application to neural-network learning. *Physical Review Letters* 66(18):2396–2399, 1991.
- [93] Le Cun, Y., I. Kanter, and S.A. Solla. Second order properties of error surfaces: Learning time and generalization. In R.P. Lippmann, J.E. Moody, and D.S. Touretzky, editors, *Advances in Neural Information Processing Systems* (3), pp. 918–924. Morgan Kaufmann, San Mateo, 1991.
- [94] Le Cun, Y., P.Y. Simard, and B. Pearlmutter. Automatic learning rate maximization by on-line estimation of the Hessian eigenvectors. In S.J. Hanson, J.D. Cowan, C.L. Giles, editors, *Advances in Neural Information Processing Systems* (5), pp. 156–163. Morgan Kaufmann, San Mateo, 1993.
- [95] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Technical Report 856*, Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, 1988.
- [96] Cybenko, G. Approximation by superpositions of sigmoids. *Mathematics of Control, Signals, and Systems* 2:303–314, 1989.
- [97] Darken, C., and J. Moody. Note on learning rate schedules for stochastic optimization. In R.P. Lippmann, J.E. Moody, and D.S. Touretzky, editors, *Advances in Neural Information Processing Systems* (3), pp. 832–838. Morgan Kaufmann, San Mateo, 1991.
- [98] Darken, C., and J. Moody. Towards faster stochastic gradient search. In J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems* (4), pp. 1009–1016. Morgan Kaufmann, San Mateo, 1992.
- [99] DasGupta, B., H.T. Siegelmann, and E. Sontag. On the intractability of loading neural networks. In V. Roychowdhury, K.-Y. Siu, and A. Orlitsky, editors, *Theoretical Advances in Neural Computation and Learning*, pp. 357–389. Kluwer, Boston, 1994.
- [100] Davis, H.F., and A.D. Snider. *Introduction to Vector Analysis*, 4th edition. Allyn and Bacon, Boston, 1979.
- [101] Davis, L. Mapping classifier systems into neural networks. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems* (1), pp. 49–56. Morgan Kaufmann, San Mateo, 1989.
- [102] de Garis, H. Genetic programming: Modular neural evolution for Darwin machines. In *Proceedings of the International Joint Conference on Neural Networks (San Diego)*, vol. 1, p. 194. IEEE, New York, 1990.
- [103] de Groot, C., and D. Wurtz. “Plain backpropagation” and advanced optimization algorithms: A comparative study. *Neurocomputing* 6:153–161, 1994.
- [104] de Villiers, J., and E. Barnard. Backpropagation neural nets with one and two hidden layers. *IEEE Transactions on Neural Networks* 4(1):136–141, 1993.

- [105] Denoeux, T., and R. Lengellé. Initializing back propagation networks with prototypes. *Neural Networks* 6(3):351–363, 1993.
- [106] Diamantaras, K.I., and S.Y. Kung. *Principal Component Neural Networks, Theory and Applications*. John Wiley & Sons, New York, 1996.
- [107] Dietterich, T., and G. Bakiri. Error-correcting output codes: A general method for improving multiclass inductive learning programs. In D.H. Wolpert, editor, *The Mathematics of Generalization*, pp. 395–407. Addison-Wesley, Reading, MA, 1995.
- [108] Dodd, N. Optimisation of network structure using genetic techniques. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, pp. 965–970. IEEE, New York, 1990.
- [109] Dominic, S., R. Das, D. Whitley, and C. Anderson. Genetic reinforcement learning for neural networks. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 2, pp. 71–76. IEEE, New York, 1991.
- [110] Drago, G.P., and S. Ridella. Statistically controlled activation weight initialization (SCAWI). *IEEE Transactions on Neural Networks* 3(4):627–631, 1992.
- [111] Drucker, H., C. Cortes, L.D. Jackel, Y. LeCun, and V. Vapnik. Boosting and other ensemble methods. *Neural Computation* 6(6):1289–1301, 1994.
- [112] Drucker, H., and Y. Le Cun. Double backpropagation and increasing generalization performance. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 2, pp. 145–150. IEEE, New York, 1991.
- [113] Duda, R.O., and P.E. Hart. *Pattern Classification and Scene Analysis*. Wiley, New York, 1973.
- [114] Eaton, H.A.C., and T.L. Olivier. Learning coefficient dependence on training set size. *Neural Networks* 5(2):283–288, 1992.
- [115] Efron, B. *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia, 1982.
- [116] Efron, B., and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.
- [117] Ehrenfeucht, A., D. Haussler, M. Kearns, and L. Valiant. A general lower bound on the number of examples needed for learning. In *Proceedings of the 1988 Workshop on Computational Learning Theory*. Morgan Kaufmann, San Mateo, 1988.
- [118] Elman, J.L., and D. Zipser. Learning the hidden structure of speech. *Journal of the Acoustical Society of America* 83(4):1615–1626, 1988.
- [119] Fahlman, S. E. The recurrent cascade-correlation architecture. In R.P. Lippmann, J.E. Moody, and D.S. Touretzky, editors, *Advances in Neural Information Processing Systems (Denver, 1990)* (3), pp. 190–196. Morgan Kaufmann, San Mateo, 1991.
- [120] Fahlman, S. E., and C. Lebiere. The cascade-correlation learning architecture. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems (Denver, 1989)* (2), pp. 524–532. Morgan Kaufmann, San Mateo, 1990.
- [121] Fahlman, S.E. Faster-learning variations of back-propagation: An empirical study. In D. Touretzky, G. Hinton, and T. Sejnowski, editors, *Proceedings of the 1988 Connectionist Models Summer School*, pp. 38–51. Morgan Kaufmann, San Mateo, 1988.
- [122] Finnoff, W., F. Hergert, and H.G. Zimmermann. Extended regularization methods for nonconvergent model selection. In S.J. Hanson, J.D. Cowan, and C.L. Giles, editors, *Advances in Neural Information Processing Systems* (5), pp. 630–637. Morgan Kaufmann, San Mateo, 1993.
- [123] Finnoff, W., F. Hergert, and H.G. Zimmermann. Improving model selection by nonconvergent methods. *Neural Networks* 6(6):771–783, 1993.
- [124] Fitch, J.P., S.K. Lehman, and F.U. Dowla. Ship wake detection procedure using conjugate gradient trained artificial neural networks. *IEEE Transactions on Geoscience and Remote Sensing* 29(5):718, 1991.
- [125] Flower, B., and M. Jabri. Summed weight neuron perturbation: An O(N) improvement over weight perturbation. In S.J. Hanson, J.D. Cowan, and C.L. Giles, editors, *Advances in Neural Information Processing Systems* (5), pp. 212–219. Morgan Kaufmann, San Mateo, 1993.

- [126] Franzini, M.A. Speech recognition with back propagation. In *Proceedings of the Ninth Annual Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 3, pp. 1702–1703. IEEE Press, IEEE, New York, 1987.
- [127] Frasconi, P., M. Gori, and A. Tesi. Backpropagation for linearly-separable patterns: a detailed analysis. In *IEEE International Conference on Neural Networks (San Francisco)*, vol. 3, pp. 1818–1822. IEEE, New York, 1993.
- [128] Frean, M. The upstart algorithm: A method for constructing and training feedforward neural networks. *Neural Computation* 2(2):198–209, 1990.
- [129] Friedman, J. H., and W. Stuetzle. Projection pursuit regression. *Journal of the American Statistical Association* 76(376):817–823, 1981.
- [130] Fukunaga, K. *Introduction to Statistical Pattern Recognition*. Academic Press, New York, 1972.
- [131] Funahashi, K. On the approximate realization of continuous mappings by neural networks. *Neural Networks* 2(3):183–192, 1989.
- [132] Gallant, A.R., and H. White. There exists a neural network that does not make avoidable mistakes. In *Proceedings of the IEEE International Conference on Neural Networks (San Diego)*, vol. 1, pp. 657–664. IEEE, New York, 1988 (reprinted in [395]).
- [133] Gallant, S.I. Perceptron-based learning algorithms. *IEEE Transactions on Neural Networks* 1(2):179–191, June 1990.
- [134] Gallinari, P., S. Thiria, F. Badran, and F. Fogelman-Soulie. On the relations between discriminant analysis and multilayer perceptrons. *Neural Networks* 4(3):349–360, 1991.
- [135] Geman, S., and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6:721–741, 1984. (reprinted in J.A. Anderson and E. Rosenfeld, editors, *Neurocomputing*, MIT Press, Cambridge, 1988).
- [136] Georgiou, G.M., and C. Koutsougeras. Embedding domain information in backpropagation. In *Proceedings of SPIE Conference on Adaptive and Learning Systems, Orlando, Fla.* Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, 1992.
- [137] Gibson, G.J., and C.F.N. Cowan. On the decision regions of multilayer perceptrons. *Proceedings of the IEEE* 78(10):1590–1594, Oct. 1990.
- [138] Gill, P.E., W. Murray, and M.H. Wright. *Practical Optimization*. Academic Press, New York, 1981.
- [139] Girosi, F., and T. Poggio. Representation properties of networks: Kolmogorov's theorem is irrelevant. *Neural Computation* 1(4):465–469, 1989.
- [140] Girosi, F., and T. Poggio. Networks for learning. In P. Antognetti and V. Milutinović, editors, *Neural Networks: Concepts, Applications, and Implementations*, chapter 6. Prentice Hall, Englewood Cliffs, NJ, 1991.
- [141] Goldberg, D. E. *Genetic Algorithms in Search, Optimization, & Machine Learning*. Addison-Wesley, Reading, MA, 1989.
- [142] Goldman, M. Communication complexity and lower bounds for threshold circuits. In V. Roychowdhury, K.-Y. Siu, and A. Orlitsky, editors, *Theoretical Advances in Neural Computation and Learning*, pp. 85–125. Kluwer, Boston, 1994.
- [143] Goldmann, M., J. Håstad, and A. Razborov. Majority gates vs. general weighted threshold gates. In *Proceedings of the 7th Annual Structure in Complexity Theory Conference*, pp. 2–13. IEEE Computer Society Press, Los Alamitos, CA, 1991.
- [144] Gori, M., and M. Maggini. Optimal convergence of on-line backpropagation. *IEEE Transactions on Neural Networks* 7(1):251–254, 1996.
- [145] Gori, M., and A. Tesi. On the problem of local minima in backpropagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(1):76–86, 1992.
- [146] Green, R.W., and C.M. DeAngelis. An empirical comparison of backpropagation training algorithms. In *World Congress on Neural Networks (Portland)*, vol. 4, pp. 401–406. Erlbaum, Hillsdale, NJ, 1993.

- [147] Guyon, I., V. Vapnik, B. Boser, L. Bottou, and S.A. Solla. Structural risk minimization for character recognition. In J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems* (4), pp. 471–478. Morgan Kaufmann, San Mateo, 1992.
- [148] Hagan, M., and M. Menhaj. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks* 5(6):989–993, 1994.
- [149] Hamamoto, M., J. Kamruzzaman, and Y. Kumagai. Generalization ability of artificial neural network using Fahlman and Lebiere’s learning algorithm. In *Proceedings of the International Joint Conference on Neural Networks (Baltimore)*, vol. 1, pp. 613–618. IEEE Press, New York, 1992.
- [150] Hamey, L.G.C. Comments on “Can backpropagation error surface not have local minima.” *IEEE Transactions on Neural Networks* 5(5):844–845, 1994.
- [151] Hampson, S.E., and D.J. Volper. Linear function neurons: Structure and training. *Biological Cybernetics* 53:203–217, 1986.
- [152] Hansen, L.K. Stochastic linear learning: Exact test and training set averages. *Neural Networks* 6(3):393–396, 1993.
- [153] Hanson, S.J. Meiosis networks. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems (Denver, 1989)* (2), pp. 533–541. Morgan Kaufmann, San Mateo, 1990.
- [154] Hanson, S.J., and L.Y. Pratt. Comparing biases for minimal network construction with back-propagation. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems (Denver, 1988)* (1), pp. 177–185. Morgan Kaufmann, San Mateo, 1989.
- [155] Harp, S. A., and T. Samad. Genetic optimization of self-organizing feature maps. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 1, pp. 341–346, 1991.
- [156] Harp, S. A., T. Samad, and A. Guha. Designing application-specific neural networks using the genetic algorithm. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems* (2), pp. 447–454. Morgan Kaufmann, San Mateo, 1989.
- [157] Hassibi, B., and D.G. Stork. Optimal brain surgeon. In S.J. Hanson, J.D. Cowan, and C.L. Giles, editors, *Advances in Neural Information Processing Systems (Denver, 1992)* (5), pp. 164–171. Morgan Kaufmann, San Mateo, 1993.
- [158] Hassibi, B., D.G. Stork, and G. Wolff. Optimal brain surgeon: Extensions and performance comparisons. In J.D. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems* (6), pp. 263–270. Morgan Kaufmann, San Mateo, 1994.
- [159] Hassibi, B., D.G. Stork, and G.J. Wolff. Optimal brain surgeon and general network pruning. In *IEEE International Conference on Neural Networks (San Francisco)*, pp. 293–299. IEEE, New York, 1993.
- [160] Haykin, S. *Neural Networks, A Comprehensive Foundation*. Macmillan, New York, 1993.
- [161] Hecht-Nielsen, R. Kolmogorov’s mapping neural network existence theorem. In *Proceedings of the IEEE First International Conference on Neural Networks (San Diego)*, vol. 3, pp. 11–13. IEEE, New York, 1987.
- [162] Hecht-Nielsen, R. Theory of the backpropagation neural network. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, pp. 593–605. IEEE, New York, 1989.
- [163] Hecht-Nielsen, R. On the algebraic structure of feedforward network weight spaces. In R. Eckmiller, editor, *Advanced Neural Computers*, pp. 129–135. Elsevier, New York, 1990.
- [164] Hecht-Nielsen, R. The munificence of high dimensionality. In I. Aleksander and J. Taylor, editors, *Proceedings of the 1992 International Conference on Artificial Neural Networks (ICANN-92)*, vol. 2, pp. 1017–1030. Elsevier, 1992.
- [165] Hergert, F., W. Finnoff, and H.G. Zimmermann. A comparison of weight elimination methods for reducing complexity in neural networks. In *Proceedings of the International Joint Conference on Neural Networks (Baltimore)*, vol. 3, pp. 980–987. IEEE, New York, 1992.
- [166] Hertz, J., A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, Reading, MA, 1991.

- [167] Hertz, J.A., and A. Krogh. Statistical dynamics of learning. In T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, editors, *Artificial Neural Networks (ICANN-91)*, vol. 1, pp. 125–131. Elsevier, New York, 1991.
- [168] Higashino, J., B.L. de Greef, and E.H.J. Persoon. Numerical analysis and adaptation method for learning rate of back propagation. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, pp. 627–630. IEEE, New York, 1990.
- [169] Hinton, G.E. Connectionist learning procedures. *Artificial Intelligence* 40(1):143–150, 1989.
- [170] Hirose, Y., K. Yamashita, and S. Hijiya. Back-propagation algorithm which varies the number of hidden units. *Neural Networks* 4(1):61–66, 1991.
- [171] Hoehfeld, M., and S.E. Fahlman. Learning with limited numerical precision using the cascade-correlation algorithm. *IEEE Transactions on Neural Networks* 3(4):602–611, 1992.
- [172] Holden, S.B. How practical are VC dimension bounds. In *IEEE International Conference on Neural Networks (Orlando)*, pp. 327–332. IEEE, New York, 1994.
- [173] Holland, J. H. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, 1975.
- [174] Holmström, L., and P. Koistinen. Using additive noise in back-propagation training. *IEEE Transactions on Neural Networks* 3(1):24–38, Jan. 1992.
- [175] Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks* 4(2):251–257, 1991.
- [176] Hornik, K. Some new results on neural network approximation. *Neural Networks* 6(8):1069–1072, 1993.
- [177] Hornik, K., M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks* 2(5):359–366, 1989 (reprinted in [395]).
- [178] Hosmer, D.W., and S. Lemeshow. *Applied Logistic Regression*. Wiley, New York, 1989.
- [179] Hsiung, J.T., W. Suwatanakul, and D.M. Himmelblau. Should back propagation be replaced by more effective optimization algorithms? In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 1, pp. 353–356. IEEE, New York, 1991.
- [180] Huang, W.Y., and R.P. Lippmann. Neural net and traditional classifiers. In D. Anderson, editor, *Neural Information Processing Systems (Denver 1987)*, pp. 387–396. American Institute of Physics, New York, 1988.
- [181] Hush, D.R., B. Horne, and J.M. Salas. Error surfaces for multi-layer perceptrons. *IEEE Transactions on Systems, Man, and Cybernetics* 22(5):1152–1161, 1992.
- [182] Hush, D.R., and J.M. Salas. Improving the learning rate of back-propagation with the gradient reuse algorithm. In *Proceedings of the IEEE International Conference on Neural Networks (San Diego)*, vol. 1, pp. 441–447. IEEE, New York, 1988.
- [183] Hush, D.R., J.M. Salas, and B. Horne. Error surfaces for multi-layer perceptrons. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 1, pp. 759–764. IEEE, New York, 1991.
- [184] Hwang, J.-N., J.J. Choi, S. Oh, and R.J. Marks II. Query learning based on boundary search and gradient computation of trained multilayer perceptrons. In *Proceedings of the International Joint Conference on Neural Networks (San Diego)*, vol. 3, pp. 57–62. IEEE, New York, June 1990.
- [185] Hwang, J.-N., S.-R. Lay, M. Maechler, D. Martin, and J. Schimert. Regression modeling in back-propagation and projection pursuit learning. *IEEE Transactions on Neural Networks* 5(3):342–353, 1994.
- [186] Irie, B., and S. Miyake. Capabilities of three-layered perceptrons. In *Proceedings of the IEEE International Conference on Neural Networks (San Diego)*, vol. 1, pp. 641–648. IEEE, New York, 1988.
- [187] Ishikawa, M. A structural learning algorithm with forgetting of link weights. *Technical Report TR-90-7*. Electrotechnical Laboratory, Tsukuba-City, Japan, 1990.
- [188] Ito, Y. Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory. *Neural Networks* 4(3):385–394, 1991.
- [189] Izui, Y., and A. Pentland. Analysis of networks with redundancy. *Neural Computation* 2(2):226–238, 1990.
- [190] Izui, Y., and A. Pentland. Speeding up back propagation. In *Proceedings of the International Joint Conference on Neural Networks (Washington, D.C.)*, vol. 1, pp. 639–642. IEEE, New York, 1990.

- [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. Sejnowski, 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 networks. 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. Touretzky, editor, *Advances in Neural Information Processing Systems (Denver, 1989)* (2), pp. 650–657. Morgan Kaufmann, 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. back 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 non-linear 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, 1989.
- [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 Conference 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 propagation. 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 Processing* 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.

- [357] Stevenson, M., R. Winter, and B. Widrow. Sensitivity of feedforward neural networks to weight errors. *IEEE Transactions on Neural Networks* 1(1):71–80, 1990.
- [358] Stinchcombe, M., and H. White. Universal approximation using feedforward networks with non-sigmoid hidden layer activation functions. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, pp. 613–617. IEEE, New York, 1989 (reprinted in [395]).
- [359] Stone, G.O. An analysis of the delta rule and the learning of statistical associations. In D.E. Rumelhart and J.L. McClelland, editors, *Parallel Distributed Processing*, vol. 1, chapter 11. MIT Press, Cambridge, 1986.
- [360] Stone, J.V., and R. Lister. On the relative time complexities of standard and conjugate gradient back propagation. In *IEEE International Conference on Neural Networks (Orlando)*, vol. 1, pp. 84–87. IEEE, New York, 1994.
- [361] Suddarth, S.C. *The Symbolic-Neural Method For Creating Models and Control Behaviors From Examples*. Ph.D. thesis, University of Washington, Seattle, 1988.
- [362] Sussmann, H.J. Uniqueness of the weights for minimal feedforward nets with a given input-output map. *Neural Networks* 5(4):589–593, 1992.
- [363] Sutton, R.S. Two problems with backpropagation and other steepest-descent learning procedures for networks. In *Proceedings of the 8th Annual Conference of the Cognitive Science Society*, pp. 823–831. Erlbaum, Hillsdale, NJ, 1986.
- [364] Sutton, R.S. Learning to predict by the methods of temporal differences. *Machine Learning* 3:9–44, 1988.
- [365] Svarer, C., L.K. Hansen, and J. Larsen. On design and evaluation of tapped-delay neural network architectures. In *IEEE International Conference on Neural Networks (San Francisco)*, pp. 977–984. IEEE, New York, 1993.
- [366] Tawel, R. Does the neuron “learn” like the synapse. In *Advances in Neural Information Processing Systems* (1), pp. 169–176. Morgan Kaufmann, San Mateo, 1989.
- [367] Tesauro, G., and B. Janssens. Scaling relationships in back-propagation learning. *Complex Systems* 2:39–44, 1988.
- [368] Thimm, G., and E. Fiesler. Neural network initialization. In *International Workshop on Artificial Neural Networks, Malaga-Torremolinos, Spain, June 1995*, pp. 535–542. Springer-Verlag, New York, 1995.
- [369] Thimm, G., P. Moerland, and E. Fiesler. The interchangeability of learning rate and gain in backpropagation neural networks. *Neural Computation* 8(2):451–460, 1996.
- [370] Tikhonov, A.N., and V.Y. Arsenin. *Solutions of Ill-Posed Problems*. V.H. Winston & Sons, Washington D.C., 1977.
- [371] Tishby, N., E. Levin, and S.A. Solla. Consistent inference of probabilities in layered networks: Predictions and generalization. In *Proceedings of the International Joint Conference on Neural Networks (Washington, D.C.)*, vol. 2, p. 403. IEEE, New York, 1989.
- [372] Tollenaere, T. SuperSAB: Fast adaptive back propagation with good scaling properties. *Neural Networks* 3(5):561–573, 1990.
- [373] Toms, D.J. Training binary node feedforward neural networks by backpropagation of error. *IEE Electronic Letters* 26(21):1745–1746, 1990.
- [374] Towell, G. G., and J.W. Shavlik. Interpretation of artificial neural networks: Mapping knowledge-based neural networks into rules. In J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems* (4), pp. 977–984. Morgan Kaufmann, San Mateo, 1992.
- [375] Towell, G.G., and J.W. Shavlik. Using knowledge-based neural networks to refine roughly-correct information. In S.J. Hanson, T. Petsche, M. Kearns, and R.L. Rivest, editors, *Computational Learning Theory and Natural Learning Systems, Vol. 2, Intersections between Theory and Experiment*, pp. 65–80. MIT Press, Cambridge, 1994.
- [376] Valiant, L.G. A theory of the learnable. *Communications of the ACM* 27(11):1134–1142, 1984.
- [377] van der Smagt, P.P. Minimization methods for training feedforward neural networks. *Neural Networks* 7(1):1–11, 1994.

- [378] Vapnik, V. Principles of risk minimization for learning theory. In J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems* (4), pp. 831–838. Morgan Kaufmann, San Mateo, 1992.
- [379] Vapnik, V.N., and A. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications* 16:264–280, 1971.
- [380] Vogl, T.P., J.K. Mangis, A.K. Rigler, W.T. Zink, and D.L. Alkon. Accelerating the convergence of the back-propagation method. *Biological Cybernetics* 59:257–263, 1988.
- [381] Wang, J.H., T.F. Krile, and J.F. Walkup. Reduction of interconnection weights in higher order associative memory networks. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 2, pp. 177–182. IEEE, New York, 1991.
- [382] Wasserman, P.D. Experiments in translating chinese characters using backpropagation. In *Proceedings of IEEE Computer Society Conference*, pp. 399–402, February 1988.
- [383] Watrous, R.L. Learning algorithms for connectionist networks: Applied gradient methods of nonlinear optimization. In *Proceedings of the IEEE First International Conference on Neural Networks (San Diego)*, vol. 2, pp. 619–627. IEEE, New York, 1987.
- [384] Webb, A.R. Functional approximation by feed-forward networks: A least-squares approach to generalization. *R.S.R.E. Memorandum* 4453, R.S.R.E., Malvern, U.K., 1991.
- [385] Webb, A.R., and D. Lowe. The optimized internal representation of multilayer classifier networks performs nonlinear discriminant analysis. *Neural Networks* 3(4):367–375, 1990.
- [386] Weigend, A.S., D.E. Rumelhart, and B.A. Huberman. Back-propagation, weight-elimination and time series prediction. In D. Touretzky, J. Elman, T. Sejnowski, and G. Hinton, editors, *Proceedings of the 1990 Connectionist Models Summer School*, pp. 105–116. Morgan Kaufmann, San Mateo, 1991.
- [387] Weigend, A.S., D.E. Rumelhart, and B.A. Huberman. Generalization by weight-elimination applied to currency exchange rate prediction. In *Proceedings of the International Joint Conference on Neural Networks (Seattle)*, vol. 1, pp. 837–841. IEEE, New York, 1991.
- [388] Weigend, A.S., D.E. Rumelhart, and B.A. Huberman. Generalization by weight-elimination with application to forecasting. In R. Lippmann, J. Moody, and D. Touretzky, editors, *Advances in Neural Information Processing Systems* (3), pp. 875–882. Morgan Kaufmann, San Mateo, 1991.
- [389] Weiss, S.M., and C.A. Kulikowski. *Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*. Morgan Kaufmann, San Mateo, 1991.
- [390] Werbos, P.J. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Ph.D. dissertation, Harvard University, Cambridge, 1974 (published as *The Roots of Backpropagation*, Wiley, 1994).
- [391] Werbos, P.J. Backpropagation: Past and future. In *Proceedings of the IEEE International Conference on Neural Networks (San Diego)*, vol. 1, pp. 343–353. IEEE, New York, 1988.
- [392] Werbos, P.J. Backpropagation and neurocontrol: A review and prospectus. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 1, pp. 209–216. IEEE, New York, 1989.
- [393] Wessels, L.F.A., and E. Barnard. Avoiding false local minima by proper initialization of connections. *IEEE Transactions on Neural Networks* 3(6):899–905, 1992.
- [394] White, H. Some asymptotic results for learning in single hidden layer feedforward network models. *Journal of the American Statistical Association* 84:1008–1013, 1989.
- [395] White, H. *Artificial Neural Networks, Approximation and Learning Theory*. Blackwell, Cambridge, MA, 1992.
- [396] Whitley, D. The GENITOR algorithm and selection pressure: Why rank-based allocation of reproductive trials is best. In *Proceedings Third International Conference on Genetic Algorithms*. (Washington D.C.), June 1990.
- [397] Whitley, D., and C. Bogart. The evolution of connectivity: Pruning neural networks using genetic algorithms. In *Proceedings of the International Joint Conference on Neural Networks (San Diego)*, vol. 1, p. 134. IEEE, New York, 1990.

- [398] Whitley, D., and T. Starkweather. Optimizing small neural networks using a distributed genetic algorithm. In *Proceedings of the International Joint Conference on Neural Networks (San Diego)*, vol. 1, p. 206. IEEE, New York, 1990.
- [399] Widrow, B. Generalization and information storage in networks of adaline “Neurons”. In M. C. Yovitz, G. T. Jacobi, and G. D. Goldstein, editors, *Self Organizing Systems 1962*, pp. 435–461. Spartan Books, Washington, D.C., 1962.
- [400] Widrow, B. Pattern recognition and adaptive control. In *1964 IRE WESCON Convention Record*, pp. 269–277. IRE, 1964.
- [401] Widrow, B., G.F. Groner, M.J.C. Hu, F.W. Smith, D.F. Specht, and L.R. Talbert. Practical applications for adaptive data-processing systems. In *1963 IRE WESCON Convention Record*, pp. 1–14, 1963.
- [402] Widrow, B., and M.E. Hoff. Adaptive switching circuits. In *1960 IRE WESCON Convention Record*, pp. 96–104. IRE, New York, 1960 (reprinted in J.A. Anderson and E. Rosenfeld, editors, *Neurocomputing*. MIT Press, Cambridge, 1988).
- [403] Widrow, B., and M.A. Lehr. 30 years of adaptive neural networks: Perceptron, madaline, and backpropagation. *Proceedings of the IEEE* 78(9):1415–1442, Sept. 1990.
- [404] Widrow, B., and F.W. Smith. Pattern-recognizing control systems. In *Computer and Information Sciences (COINS) Symposium Proceedings*, pp. 288–316. Spartan Books, Washington D.C., 1963.
- [405] Widrow, B., and S.D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1985.
- [406] Widrow, B., and R. Winter. Neural nets for adaptive filtering and adaptive pattern recognition. *Computer*, pp. 25–39, March 1988.
- [407] Wieland, A., and R. Leighton. Geometric analysis of neural network capabilities. In *Proceedings of the IEEE First International Conference on Neural Networks (San Diego)*, vol. 3, pp. 385–392. IEEE, New York, 1987.
- [408] Wittner, B.S., and J.S. Denker. Strategies for teaching layered networks classification tasks. In D.Z. Anderson, editor, *Proceedings of the Conference on Neural Information Processing Systems (Denver 1987)*, pp. 850–859. American Institute of Physics, New York, 1988.
- [409] Wolpert, D.H. Stacked generalization. *Neural Networks* 5(2):241–259, 1992.
- [410] Wynne-Jones, M. Node splitting: A constructive algorithm for feed-forward neural networks. In J. E. Moody, S. J. Hanson, and R. P. Lippmann, editors, *Advances in Neural Information Processing Systems* (4), pp. 1072–1079. Morgan Kaufmann, San Mateo, 1992.
- [411] Yamada, K., H. Kami, J. Tsukumo, and T. Temma. Handwritten numeral recognition by multi-layered neural network with improved learning algorithm. In *Proceedings of the International Joint Conference on Neural Networks (Washington D.C.)*, vol. 2, pp. 259–266. IEEE, New York, 1989.
- [412] Yu, X., N.K. Loh, and W.C. Miller. Training hard-limiting neurons using backpropagation algorithm by updating steepness factors. In *IEEE International Conference on Neural Networks (Orlando)*, vol. 1, pp. 526–530. IEEE, New York, 1994.
- [413] Yu, X.-H. Can backpropagation error surface not have local minima? *IEEE Transactions on Neural Networks* 3(6):1019–1021, 1992.
- [414] Yu, X.-H., and G.-A. Chen. On the local minima free condition of backpropagation learning. *IEEE Transactions on Neural Networks* 6(5):1300–1303, 1995.
- [415] Yu, X.-H., and G.-A. Chen. Corrections to “On the local minima free condition of backpropagation learning.” *IEEE Transactions on Neural Networks* 7(1):256–257, 1996.
- [416] Yu, Yeong-Ho, and R.F. Simmons. Extra output biased learning. In *Proceedings of the International Joint Conference on Neural Networks (San Diego)*, vol. 2, pp. 161–166. IEEE, New York, 1990.

This excerpt from

Neural Smithing.
Russell D. Reed and Robert J. Marks II.
© 1999 The MIT Press.

is provided in screen-viewable form for personal use only by members of MIT CogNet.

Unauthorized use or dissemination of this information is expressly forbidden.

If you have any questions about this material, please contact
cognetadmin@cognet.mit.edu.