Bibliography

Aarts, E. H. L. and J. Korst (1989). *Simulated annealing and Boltzmann machines : a stochastic approach to combinatorial optimization and neural computing*. Chichester: Wiley.

Anonymous (1996). On-line collection of references concerning theory and application of Kohonen's self-organising map and related methods. Available on the world-wide-web at `http://liinwww.ira.uka.de/bibliography/Neural/SOM.LVQ.html`, or via `ftp` at `cochlear.hut.fi`.

Balakrishnan, P. V., M. C. Cooper, V. S. Jacob, and P. A. Lewis (1994). A study of the classification capabilities of neural networks using unsupervised learning: a comparison with *k*-means clustering. *Psychometrika* ***59***(4), 509–525.

Baldi, P. and K. Hornik (1989). Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks* **2**(1), 53–58.

Baldi, P. and K. Hornik (1995). Learning in linear neural networks: a survey. *IEEE Transactions on Neural Networks* ***6***(4), 837–858.

Barrett, P. J. (1995). *Exploratory Database Visualisation: the application & assessment of data and dimensionality reduction*. Ph. D. thesis, Neural Computing Research Group, Aston University, Aston Street, Birmingham, UK, B4 7ET.

Bauer, H.-U. and K. R. Pawelzik (1992). Quantifying the neighbourhood preservation of self-organizing feature maps. *IEEE Transactions on Neural Networks* ***3***(4), 570–579.

Bellman, R. E. (1961). *Adaptive Control Processes*. New Jersey: Princeton University Press.

Bezdek, J. C. and N. R. Pal (1995). An index of topological preservation for feature extraction. *Pattern Recognition* ***28***(3), 381–391.

Bishop, C. M. (1991). Improving the generalisation properties of radial basis function neural networks. *Neural Computation 3*, 579–588.

Bishop, C. M. (1993). Curvature driven smoothing: a learning algorithm for feedforward networks. *IEEE Transactions on Neural Networks* ***4***(5), 882–884.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford: Clarendon Press.

Bishop, C. M., M. Svensén, and C. K. I. Williams (1996). GTM: A principled alternative to the self-organizing map. Technical Report NCRG/96/015, Neural Computing Research Group, Aston University, Aston Street, Aston Triangle, Birmingham B4 7ET, UK. Submitted to *Neural Computation*.

Biswas, G., A. K. Jain, and R. C. Dubes (1981). Evaluation of projection algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* ***PAMI-3***(6), 701–708.

Blasdel, G. and G. Salama (1986). Voltage sensitive dyes reveal a modular organization in monkey striate cortex. *Nature 321*, 579–585.

Bottou, L. and Y. Bengio (1995). Convergence properties of the *k*-means algorithms. In G. Tesauro, D. S. Touretzky, and T. K. Leen (Eds.), *Advances in Neural Information Processing Systems 7*, pp. 585–592. Cambridge, Mass: MIT Press.

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

Bryan, J. G. (1951). The generalized discriminant function: mathematical foundation and computational routine. *Harvard Educational Review 21*, 90–95.

Carroll, J. D. and P. Arabie (1980). Multidimensional scaling. *Annual Review of Psychology 31*, 607–649.

Chang, C. L. and R. C. T. Lee (1973). A heuristic relaxation method for nonlinear mapping in cluster analysis. *IEEE Transactions on Systems, Man and Cybernetics 3*, 197–200.

Chatfield, C. and A. J. Collins (1980). *Introduction to Multivariate Analysis*. London: Chapman & Hall.

Cox, T. F. and G. Ferry (1993). Discriminant analysis using non-metric multidimensional scaling. *Pattern Recognition* ***26***(1), 145–153.

Crace, J. (1995). Latest in the rating game. *The Guardian*, June 24.

Davison, M. L. (1983). *Multidimensional Scaling*. New York: Wiley.

de Leeuw, J. (1988). Covergence of the majorisation method for multidimensional scaling. *Journal of Classification 5*, 163–180.

de Soete, G., L. Hubert, and P. Arabie (1987). The comparative performance of simulated annealing on two problems of combinatorial data analysis. In *First Conf. IFCS, Technical University of Aachen*.

Domine, D., J. Devillers, M. Chastrette, and W. Karcher (1993). Non-linear mapping for structure-activity and structure-property modelling. *Journal of Chemometrics 7*, 227–242.

Dong, D. and T. J. McAvoy (1996). Nonlinear principal component analysis — based on principal curves and neural networks. *Computers and Chemical Engineering* ***20***(1), 65–78.

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

Dzwinel, W. (1994). How to make Sammon's mapping useful for multidimensional data structures analysis. *Pattern Recognition* ***27***(7), 949–959.

Ekman, G. (1954). Dimensions of color vision. *Journal of Psychology 38*, 467–474.

Erwin, E., K. Obermayer, and K. Schulten (1992). Self-organizing maps: ordering, convergence properties and energy functions. *Biological Cybernetics 67*, 47–55.

Fisher, R. A. (1936). The use of multiple measurements on taxonomic problems. *Annals of Eugenics 7*, 179–188.

Friedman, J. H. (1994). An overview of predictive learning and function approximation. In V. Cherkassky, J. H. Friedman, and H. Wechsler (Eds.), *From Statistics to Neural Networks — Theory and Pattern Recognition Applications*, Volume 136 of *NATO ASI Series F: Computer and Systems Sciences*, pp. 1–61. Berlin: Springer-Verlag.

Friedman, J. H. (1995). Introduction to computational learning and statistical prediction. Tutorial 6, ICANN95, Paris.

Friedman, J. H. and J. W. Tukey (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers C-23*, 881–889.

Fritzke, B. (1994). Growing cell structures — a self-organizing network for unsupervised and supervised learning. *Neural Networks* ***7***(9), 1441–1460.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition* (Second ed.). London: Academic Press.

Gallinari, P., S. Thiria, F. Badran, and F. Fogelman-Soulié (1991). On the relations between discriminant analysis and multilayer perceptrons. *Neural Networks 4*, 349–360.

Gamelin, T. W. and R. E. Greene (1983). *Introduction to Topology*. Philadelphia: Saunders College Publishing.

Geman, S., E. Bienenstock, and R. Doursat (1992). Neural networks and the bias/variance dilemma. *Neural Computation* ***4***(1), 1–58.

Golub, G. and W. Kahan (1965). Calculating the singular values and pseudo-inverse of a matrix. *SIAM Numerical Analysis, B* **2**(2), 205–224.

Goodhill, G. J., S. Finch, and T. J. Sejnowski (1995). Quantifying neighbourhood preservation in topographic mappings. Technical Report INC-9505, Institute for Neural Computation, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA.

Goodhill, G. J., M. W. Simmen, and D. J. Willshaw (1995). An evaluation of the use of multidimensional scaling for understanding brain connectivity. *Philosophical Transactions of the Royal Society of London Series B 348*, 265–280.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika 53*, 325–338.

Gower, J. C. and P. Legendre (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification 3*, 5–48.

Gower, J. C. and G. J. S. Ross (1969). Minimal spanning trees and single linkage cluster analysis. *Applied Statistics* ***18***(1), 54–64.

Hastie, T. and W. Stuetzle (1989). Principal curves. *Journal of the American Statistical Association 84*, 502–516.

Haykin, S. (1994). *Neural Networks: a comprehensive foundation*. New York: Macmillan.

Heiser, W. J. (1991). A generalized majorisation method for least squares multidimensional scaling of pseudodistances that may be negative. *Psychometrika* ***56***(1), 7–27.

Higher Education Funding Council for England (1994). The 1992 RAE database. Available through JANET on NISSWAIS.

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

Hofmann, T. (1995). Personal communication.

Hofmann, T. and J. Buhmann (1995). Multidimensional scaling and data clustering. In G. Tesauro, D. S. Touretzky, and T. K. Leen (Eds.), *Advances in Neural Information Processing Systems 7*, pp. 459–466. Cambridge, Mass: MIT Press.

Huber, P. J. (1985). Projection pursuit. *The Annals of Statistics* ***13***(2), 435–475.

Jain, A. K. and J. Mao (1992). Artificial neural network for nonlinear projection of multivariate data. In *IJCNN International Joint Conference on Neural Networks*, Volume 3, pp. 335–340. New York: IEEE.

Johnes, J., J. Taylor, and B. Francis (1993). The research performance of UK universities: a statistical analysis of the results of the 1989 Research Selectivity Exercise. *Journal of the Royal Statistical Society, A* ***156***(2), 271–286.

Joint Performance Indicators Working Group (1994). Explanatory and statistical material to accompany recomendations for research indicators. Technical report, Higher Eduction Funding Council for England, Northavon House, Coldharbour Lane, Bristol, BS16 1QD.

Jolliffe, I. T. (1986). *Principal Component Analysis*. New York: Springer-Verlag.

Kaas, J. H., R. J. Nelson, M. Sur, C.-S. Lin, and M. M. Merzenich (1979). Multiple representations of the body within the primary somatosensory cortex of primates. *Science 204*, 521–523.

Kakusho, O. and R. Mizoguchi (1983). A new algorithm for non-linear mapping with applications to dimension and cluster analyses. *Pattern Recognition* ***16***(1), 109–117.

Karhunen, J. (1994). Stablilty of Oja's PCA subspace rule. *Neural Computation* ***6***(4), 739–747.

Karhunen, J. and J. Joutsensalo (1994). Representation and separation of signals using nonlinear PCA type learning. *Neural Networks* ***7***(1), 113–127.

Klein, R. W. and R. C. Dubes (1989). Experiments in projection and clustering by simulated annealing. *Pattern Recognition* ***22***(2), 213–220.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics 43*, 59–69.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE* ***78***(9), 1464–1480.

Kohonen, T. (1995). *Self-Organizing Maps*. Berlin: Springer-Verlag.

Koontz, W. L. G. and K. Fukunaga (1972). A nonlinear feature extraction algorithm using distance transformation. *IEEE Transactions On Computers* ***C-21***(1), 56–63.

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal* ***37***(2), 233–243.

Kruskal, J. B. (1964a). Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis. *Psychometrika* ***29***(1), 1–27.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrika* ***29***(2), 115–129.

Kruskal, J. B. (1971). Comments on 'A nonlinear mapping for data structure analysis'. *IEEE Transactions on Computers C-20*, 1614.

Krzanowski, W. J. and F. H. C. Marriott (1994). *Multivariate Analysis Part I: Distributions, Ordination and Inference*. London: Edward Arnold.

LeBlanc, M. and R. Tibshirani (1994). Adaptive principal surfaces. *Journal of the American Statistical Association* ***89***(425), 53–64.

Lee, R. C. T., J. R. Slagle, and H. Blum (1977). A triangulation method for the sequential mapping of points from *N*-space to two-space. *IEEE Transactions on Computers C-26*, 288–292.

Levine, S. H. and J. G. Kreifeldt (1994). Uniquely representing point patterns with minimal information. *IEEE Transactions on Systems, Man and Cybernetics* ***SMC-24***(6), 895–900.

Li, X., J. Gasteiger, and J. Zupan (1993). On the topology distortion in self-organising feature maps. *Biological Cybernetics 70*, 189–198.

Linde, Y., A. Buzo, and R. M. Gray (1980). An algorithm for vector quantizer design. *IEEE Transactions on Communications COM-28*, 84–95.

Lippmann, R. P. (1994). Neural networks, Bayesian *a posteriori* probabilities, and pattern classification. In V. Cherkassky, J. H. Friedman, and H. Wechsler (Eds.), *From Statistics to Neural Networks — Theory and Pattern Recognition Applications*, Volume 136 of *NATO ASI Series F: Computer and Systems Sciences*, pp. 83–104. Berlin: Springer-Verlag.

Lowe, D. (1993). Novel 'topographic' nonlinear feature extraction using radial basis functions for concentration coding in the 'artificial nose'. In *3rd IEE International Conference on Artificial Neural Networks*. London: IEE.

Lowe, D. (1995). Radial basis function networks. In M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*, pp. 779–782. Cambridge, Mass: MIT Press.

Lowe, D. and M. E. Tipping (1995). A novel neural network technique for exploratory data analysis. In *Proceedings of ICANN '95 (Scientific Conference)*, Volume 1, pp. 339–344. Paris: EC2 & Cie.

Lowe, D. and M. E. Tipping (1996). Feed-forward neural networks and topographic mappings for exploratory data analysis. *Neural Computing and Applications 4*, 83–95.

Lowe, D. and A. R. Webb (1990). Exploiting prior knowledge in network optimization: an illustration from medical prognosis. *Network: Computation in Neural Systems* ***1***(3), 299–323.

MacLeod, D. (1995). Head hunters. *The Guardian*, September 26.

MacQueen, J. (1967). Some methods of classification and analysis of multivariate observations. In L. M. LeCam and J. Neyman (Eds.), *Proceedings of the 5th Berkeley Symposium on Mathematics, Statistics and Probability*, pp. 281–297. Berkeley, CA: University of California Press.

Mao, J. and A. K. Jain (1995). Artificial neural networks for feature extraction and multivariate data projection. *IEEE Transactions on Neural Networks* ***6***(2), 296–317.

Mardia, K. V. (1978). Some properties of classical multidimensional scaling. *Communications in Statistics — Theory and Methods* ***A7***(13), 1233–1241.

Mardia, K. V., J. T. Kent, and J. M. Bibby (1979). *Multivariate Analysis*. Probability and Mathematical Statistics. London: Academic Press.

Martinetz, T. and K. Schulten (1991). A "neural gas" network learns topologies. In T. Kohonen, K. Mäkisara, K. Simula, and J. Kangas (Eds.), *Artificial Neural Networks*, pp. 397–402. Amsterdam: Springer.

Mathar, R. and A. Zilinskas (1993). On global optimization in two-dimensional scaling. *Acta Applicandae Mathematicae 33*, 109–118.

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

Niemann, H. and J. Weiss (1979). A fast-converging algorithm for nonlinear mapping of high-dimensional data to a plane. *IEEE Transactions on Computers C-28*, 142–147.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology 15*, 267–273.

Oja, E. (1989). Neural networks, principal components, and subspaces. *International Journal of Neural Systems 1*, 61–68.

Oja, E. (1992). Principal components, minor components, and linear neural networks. *Neural Networks 5*, 927–935.

Park, J. and I. W. Sandberg (1991). Universal approximation using radial basis function networks. *Neural Computation* ***3***(2), 246–257.

Plumbley, M. D. (1995). Lyapunov functions for convergence of principal component algorithms. *Neural Networks* ***8***(1), 11–23.

Plybon, B. F. (1992). *An Introduction to Applied Numerical Analysis*. Boston: PWS-KENT.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical Recipes in C* (Second ed.). Cambridge: Cambridge University Press.

Pykett, C. E. (1978). Improving the efficiency of Sammon's nonlinear mapping by using clustering archetypes. *Electronics Letters* ***14***(25), 799–800.

Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, Series B 10*, 159–203.

Richardson, M. W. (1938). Multidimensional psychophysics. *Psychological Bulletin 35*, 659–660.

Ritter, H. (1991). Asymptotic level density for a class of vector quantization processes. *IEEE Transactions on Neural Networks 2*, 173–175.

Ritter, H. and K. Schulten (1986). On the stationary state of Kohonen's self-organizing sensory mapping. *Biological Cybernetics 54*, 99–106.

Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology 53*, 94–104.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, and the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Volume 1, Chapter 8, pp. 318–362. Cambridge, Mass: MIT Press.

Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers **C-18***(5), 401–409.

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

Saund, E. (1989). Dimensionality-reduction using connectionist networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-11***(3), 304–315.

Schiffman, S. S., M. L. Reynolds, and F. W. Young (1981). *Introduction to Multidimensional Scaling. Theory, Methods and Applications*. New York: Academic Press.

Schwartz, E. L., A. Shaw, and E. Wolfson (1989). A numerical solution to the generalised mapmaker's problem: flattening nonconvex polyhedral surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-11***(9), 1005–1008.

Shepard, R. N. (1958). Stimulus and response generalisation: Deduction of the generalisation gradient from a trace model. *Psychological Review 65*, 242–256.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function, I. *Psychometrika* **27**(2), 125–140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function, II. *Psychometrika* **27**(3), 219–246.

Shepard, R. N. and J. D. Carroll (1966). Parametric representation of nonlinear data structures. In P. Krishnaiah (Ed.), *Multivariate Analysis*, pp. 561–592. New York: Academic Press.

Siedlecki, W., K. Siedlecka, and J. Sklansky (1988). An overview of mapping techniques for exploratory pattern analysis. *Pattern Recognition **21***(5), 411–429.

SOM Programming Team (1995). SOM_PAK: the Self-Organizing Map Program Package, Version 3.1. Available by anonymous `ftp` from Helsinki University of Technology at `cochlea.hut.fi`.

Strang, G. (1988). *Linear Algebra and its Applications*. Orlando: Harcourt Brace Jovanovich.

Suga, N. and W. E. O'Neill (1979). Neural axis representing target range in the auditory cortex of the mustache bat. *Science 206*, 351–353.

Tarazaga, P. and M. W. Trosset (1993). An optimization problem on subsets of the symmetric positive-definite matrices. *Journal of Optimization Theory and Applications **79***(3), 513–524.

Tattersall, G. D. and P. R. Limb (1994). Visualisation techniques for data mining. *BT Technology Journal, British Telecom Labs, Ipswich IP5 7RE, UK **12***(4), 23–31.

Taylor, J. (1994). Measuring research performance in business and management studies in the United Kingdom: The 1992 Research Assessment Exercise. *British Journal of Management 5*, 275–288.

Taylor, J. (1995). A statistical analysis of the 1992 Research Assessment Exercise. *Journal of the Royal Statistical Society A **158***(2), 241–261.

Tikhonov, A. N. and V. Y. Arsenin (1977). *Solutions of Ill-Posed Problems*. Washington, D.C.: V.H. Winston.

Tipping, M. E. (1996). Topographic mappings, classical multidimensional scaling and the principal subspace network. Submitted for publication in *IEEE Transactions on Neural Networks*.

Torgerson, W. S. (1952). Multidimensional scaling: I. theory and method. *Psychometrika 17*, 401–419.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Webb, A. R. (1992). Non-metric multidimensional scaling using feed-forward networks. CSE1 Research Note 192, Defence Research Agency, Malvern, UK.

Webb, A. R. (1995). Multidimensional scaling by iterative majorisation using radial basis functions. *Pattern Recognition **28***(5), 753–759.

Webb, A. R. and D. Lowe (1990). The optimised internal representation of multilayer classifier networks performs nonlinear discriminant analysis. *Neural Networks **3***(4), 367–375.

Weiss, S. M. and C. A. Kulikowski (1991). *Computer Systems That Learn*. San Mateo, CA: Morgan Kaufmann.

Williams, R. (1985). Feature discovery through error-correcting learning. Technical Report 8501, Institute of Cognitive Science, University of California, San Diego.

Willshaw, D. J. and C. von der Malsburg (1976). How patterned neural connections can be set up by self-organisation. In *Proceedings of the Royal Society of London B*, Volume 194, pp. 431–445.

Wyatt, J. L. and I. M. Elfadel (1995). Time-domain solutions of Oja's equations. *Neural Computation **7***(5), 915–922.

Xu, L. (1993). Least mean square error reconstruction principle for self-organising neural-nets. *Neural Networks **6***(5), 627–648.

Yan, W.-Y., U. Helmke, and J. B. Moore (1994). Analysis of Oja's flow for neural networks. *IEEE Transactions on Neural Networks **5***(5), 674–683.

Young, F. W. and D. F. Harris (1990). Multidimensional scaling: procedure ALSCAL. In M. Norusis (Ed.), *SPSS base system: user's guide*, pp. 397–461. Chicago: SPSS.

Young, M. P. (1992). Objective analysis of the topological organization of the primate cortical visual system. *Nature 358*, 152–155.