References

Abiteboul, S., Hull, R., and Vianu, V. (1995) *Foundations of Databases*. Reading, MA: Addison-Wesley.

Adriaans, P., and Zantige, D. (1996) *Data Mining*. Harlow, UK: Addison-Wesley.

Agrawal, R., Aggarwal, C., and Prasad, V. (in press) A tree projection algorithm for finding frequent itemsets. *Journal of Parallel and Distributed Computing*.

Agrawal, R., Imielenski, T., and Swami, A. (1993) Mining association rules between sets of items in large databases. *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD'98)*, New York: ACM Press, pp. 207–216.

Agrawal, R., Lin, K.I., Sawhney, H.S., and Shim, K. (1995) Fast similarity search in the presence of noise, scaling, and translation in time-series databases. *Proceedings of VLDB-95,* pp. 490–501.

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., and Verkamo, A.I. (1996) Fast discovery of association rules. *Advances in Knowledge Discovery and Data Mining,* U.M., Fayyad, G., Piatetsky-Shapiro, P., Smyth, and Uthurasamy, R. (eds.). Menlo Park, CA: AAAI Press, pp. 307–328.

Agrawal, R., and Srikant, R. (1994) Fast algorithms for mining association rules in large databases. *Proceedings of the Twentieth International Conference on Very Large Data Bases (VLDB'94)*, pp. 487–499.

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information The-*

ory, B.N. Petrov and F. Csaki (eds.), Academiai Kiado, Budapest, pp. 267–281.

Alon, N., and Spencer, J.H. (1992) *The Probabilistic Method*. New York: Wiley.

Anderberg, M.R. (1973) *Cluster Analysis for Applications*. New York: Academic Press.

Applebaum, D. (1996) *Probability and Information: An Integrated Approach*, Cambridge, U.K.: Cambridge University Press.

Aronis, J.M., and Provost, F.J. (1997) Increasing the efficiency of data mining algorithms with breadth-first marker propagation. *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, Heckerman, D., Mannila, H., and Pregibon, D. (eds.). Menlo Park, CA: AAAI Press, pp. 119–122.

Asimov, D. (1985) The grand tour: a tool for viewing multidimensional data. *SIAM Journal of Scientific and Statistical Computing*, 6, pp. 128–143.

Atkeson, C.W., Schaal, S.A., and Moore, A.W. (1997) Locally weighted learning. *Artificial Intelligence Review*, 11, pp. 75–133.

Azzalini, A., and Bowman, A.W. (1990) A look at some data on the Old Faithful geyser. *Applied Statistics*, 39, pp. 357–365.

Babcock, C. (1994) Parallel processing mines retail data. *Computer World*, 6.

Ballard, D.H. (1997) *An Introduction to Natural Computation*. Cambridge, MA: MIT Press.

Banfield, J.D., and Raftery, A.E. (1993) Model-based Gaussian and non-Gaussian clustering. *Biometrics*, 49, pp. 803–821.

Barnett, V. (1982) *Comparative Statistical Inference*. Chichester, U.K.: Wiley.

Baum, L.E., and Petrie, T. (1966) Statistical inference for probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 37, pp. 1554–1563.

Bayardo, R.J., and Agrawal, R. (1999) Mining the most interesting rules. *Proc. 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99)*. New York: ACM Press, pp. 145–154.

Becker, R.A., Cleveland, W.S., and Wilks, A.R. (1987) Dynamic graphics for data analysis. *Statistical Science,* 2, pp. 355–395.

Becker, R.A., Eick, S.G., and Wilks, A.R. (1995) Visualizing network data. *IEEE Transactions on Visualization and Computer Graphics,* 1(1), pp. 16–28.

Bennett, K., Fayyad, U., and Geiger, D. (1999) Density-based indexing for approximate nearest-neighbor queries. *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY: ACM Press, pp. 233–243.

Bernardo, J.M., and Smith, A.F.M. (1994) *Bayesian Theory*. New York, NY: Wiley.

Berndt, D.J., and Clifford, J. (1996) Finding patterns in time-series, a dynamic programming approach. *Advances in Knowledge Discovery and Data Mining*, Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., and Uthurasamy, R. (eds). Menlo Park, CA: AAAI/MIT Press, pp. 229–248.

Berry, M.J.A., and Linoff, G. (1997) *Data Mining Techniques for Marketing, Sales, and Customer Support*. New York: Wiley.

Berry, M.J.A., and Linoff, G. (2000) *Mastering Data Mining*. New York: Wiley.

Berry, M.W. (1992) Large scale singular value computations. *International Journal of Supercomputer Applications* 6(1), pp. 13–49.

Berry, M.W., Drmvac, Z., and Jessup, E.R. (1999) Matrices, vector-spaces, and information retrieval, *SIAM Review*, 41(2), pp. 335–362.

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999) When is "nearest neighbor" meaningful? *Proceedings of the 7th International Conference on Data Theory, ICDT'99,* Lecture Notes in Computer Science, LNCS, Number 1540. New York: Springer-Verlag, pp. 217–235.

Bhandari, I., Colet, E., Parker, J., Pines, Z., Pratap, R., and Ramanujam, K. (1997) Advanced Scout: data mining and knowledge discovery in NBA data. *Data Mining and Knowledge Discovery,* 1(1), pp. 121–125.

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

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975) *Discrete Multivariate Analysis,* Cambridge, MA: MIT Press.

Blasius, J., and Greenacre, M. (1998) *Visualization of Categorical Data*. San Diego, CA: Academic Press.

Blum, T., Keislaer, D., Wheaton, J., and Wold, E. (1997) Audio databases with content-based retrieval. *Intelligent Multimedia Information Retrieval*, Maybury, M. T. (ed.). Menlo Park, CA: AAAI Press, pp. 113–135.

Box, G.E.P., and Jenkins, G.M. (1976) *Time Series Analysis: Forecasting and Control*. Oakland, CA: Holden Day.

Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994) *Time Series Analysis: Forecasting and Control,* 3rd ed. Englewood Cliffs, NJ: Prentice Hall.

Bradley, P.S., Fayyad, U.M., and Mangasarian, O.L. (1999) Mathematical programming for data mining: formulation and challenges. *INFORMS Journal on Computing*, 11, pp. 217–238.

Bradley, P.S., Fayyad, U.M., Reina, C. (1998) Scaling clustering algorithms to large databases. In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining,* R. Agrawal, P. Stolorz, and G. Piatetsky-Shapiro (eds.), Menlo Park, CA: AAAI Press, pp. 9–15.

Breese, J.S., Heckerman, D., and Kadie, C. (1998) Empirical analysis of predictive algorithms for collaborative filtering. *Proceedings 14th Conference on Uncertainty in Artificial Intelligence,* San Francisco, CA: Morgan Kaufmann, pp. 43–52..

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984) *Classification and Regression Trees*. Belmont, CA: Wadsworth Statistical Press.

Brijs, T., Goethals, B., Swinnen, G., Vanhoof, K., and Wets, G. (2000) A data mining framework for optimal product selection in retail supermarket data: the generalized PROFSET model. *Proceedings of the ACM Seventh International Conference on Knowledge Discovery and Data Mining,* New York: ACM Press, pp. 300–304.

Brin, S., and Page, L. (1998) The anatomy of a large-scale hypertextual Web search engine. *Proceedings of the Seventh International World-Wide Web Conference,* Brisbane, Australia, pp. 107–117.

Brin, S., Motwani, R., and Silverstein, C. (1997) Beyond market baskets: generalizing association rules to correlations. *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD'97),* New York: ACM Press, pp. 265–276.

Brooks, S.P., and Morgan, B.J.T. (1995) Optimisation using simulated annealing. *The Statistician,* 44, pp. 241–257.

Buckley, C., and Salton, G. (1995) Optimization of relevance feedback weights. *Proceedings of the 18th Annual ACM 1995 SIGIR Conference*, pp. 351-356.

Buja, A., Cook, D., and Swayne, D.F. (1996) Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics,* 5(1), 78–99.

Buntine, W. (1992) Learning classification trees. *Statistics and Computing,* 2, pp. 63–73.

Buntine, W., Fischer, B., and Pressburger, T. (1999) Towards automated synthesis of data mining programs. In *Proceedings of the Fifth ACM Conference on Knowledge Discovery and Data Mining,* S. Chaudhuri and D. Madigan (eds.), New York, NY: ACM Press, pp. 372–376.

Burges, C.J.C. (1998) A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery,* 2, pp. 121–167.

Burnham, K.P., and Anderson, D.R. (1998) *Model Selection and Inference: a Practical Information Theoretic Approach*. New York: Springer-Verlag.

Böhning, D. (1998) *Computer Assisted Analysis of Mixtures,* Boca Raton, FL: Chapman and Hall.

Cadez, I.V., McLaren, C.E., Smyth, P., and McLachlan, G.J. (1999) Hierarchical models for screening of iron-deficient anemia. *Proceedings of the 1999 International Conference on Machine Learning*, I. Bratko and S. Dzeroski (eds.), San Francisco, CA: Morgan Kaufmann, pp. 77–86.

Cadez, I.V., Heckerman, D., Meek, C., Smyth, P., and White, S. (2000) Visualization of navigation patterns on a Web site using model-based clustering. *Proceedings of the ACM Seventh International Conference on Knowledge Discovery and Data Mining*, New York, NY: ACM Press, pp. 280–284.

Card, S.K., MacKinlay, J.D., and Shneiderman, B. (eds.) (1999) *Readings in Information Visualization: Using Vision to Think.* San Francisco, CA: Morgan Kaufmann.

Carmines, E.G., and Zeller, R.A. (1979) *Reliability and Validity Assessment*. Beverly Hills, CA: Sage Publications.

Carr, D.B., Littlefield, R.J., Nicholson, W.L., and Littlefield, J.S. (1987) Scatterplot matrix techniques for large N. *Journal of the American Statistical Association,* 82(398), pp. 424–436.

Casti, J.L. (1990) *Searching for Certainty: What Scientists Can Know about the Future*. New York: Willam Morrow.

Celeux, G., and Govaert, G. (1995) Gaussian parsimonious clustering models. *Pattern Recognition,* 28, pp. 781–793.

Chambers, J.M., Cleveland, W.S., Kleiner, B., and Tukey, P.A. (1983) *Graphical Methods for Data Analysis*. Pacific Grove: Wadsworth and Brooks/Cole.

Chatfield, C. (1996) *The Analysis of Time Series: An Introduction*. London: Chapman and Hall.

Chatterjee, S., Handcock, M.S., and Simonoff, J.S. (1995) *A Casebook for a First Course in Statistics and Data Analysis*. New York: Wiley.

Chaudhuri, S. (1998) An overview of query optimization in relational systems. *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems,* New York: ACM Press, pp. 34–43.

Chaudhuri, S., and Dayal, U. (1997) An overview of data warehousing and OLAP technology. *Proceedings of the 1997 ACM/SIGMOD Conference,* New York: ACM Press, pp. 65–75.

Cheeseman, P., and Stutz. J. (1996) Bayesian classification (AutoClass): theory and results. In *Advances in Knowledge Discovery and Data Mining*, U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (eds.), Cambridge, MA: AAAI/MIT Press, pp. 153–180.

Cheng, X., and Wallace, J.M. (1993) Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: spatial patterns. *Journal of the Atmospheric Sciences*, 50, pp. 2674–2696.

Cherkassky, V.S., and Muller, F. (1998). *Learning from Data: Concepts, Theory, and Methods*. New York: Wiley.

Chernoff, H. (1973) The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association,* 68, pp. 361–368.

Chickering, D.M., and Heckerman, D. (1997) Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables. *Machine Learning*, 29(2/3), pp. 181–244.

Chickering, D.M., Heckerman, D., and Meek, C. (1997) A Bayesian approach to learning Bayesian networks with local structure. *Proceedings of Thirteenth Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann, pp. 80–89.

Chipman, H., George, E.I., and McCulloch, R.E. (1998) Bayesian CART model search (with discussion). *Journal of the American Statistical Association*, 93, pp. 935–960.

Clark, P., and Niblett, T. (1989) The CN2 induction algorithm. *Machine Learning*, 3(4), pp. 261–283.

Clearwater, S., and Stern, E. (1991) A rule-learning program in high-energy physics event classification. *Computational Physics Communications*, 67, pp. 159–182.

Cleveland, W.S., and McGill, M.E. (eds.) (1988) *Dynamic Graphics for Statistics*. Belmont, CA: Wadsworth and Brooks/Cole.

Cleveland, W.S., and Devlin, S.J. (1988) Locally-weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83, pp. 597–610.

Cochran, W.G. (1977) *Sampling Techniques*. New York: Wiley.

Cohen, W. (1995) Fast effective rule induction. *Proceedings of the Twelfth International Conference on Machine Learning*, San Mateo, CA: Morgan Kaufmann, pp. 115–123.

Cook, R.D., and Weisberg, S. (1994) *An Introduction to Regression Graphics*. New York: Wiley.

Cook, R.D., and Weisberg, S. (1999) *Applied Regression Including Computing and Graphics*. New York: Wiley.

Cook, W.J., Cunningham, W.H., Pulleyblank, W.R., and Schrijver, A. (1998) *Combinatorial Optimization*. New York: Wiley.

Corman, T.H., Leiserson, C. E., and Rivest, R.L. (1990) *Introduction to Algorithms*. Cambridge, MA: MIT Press.

Cortes, C., and Pregibon, D. (1998) Giga-mining. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, R. Agrawal and P. Stolorz (eds.), Menlo Park, CA: AAAI Press, pp. 174–178.

Cox D.R., and Wermuth, N. (1996) *Multivariate Dependencies: Models, Analysis, and Interpretation*. London: Chapman and Hall.

Cox, D.R., and Hinkley, D.V. (1974) *Theoretical Statistics*. London: Chapman and Hall.

Cox, T.F., and Cox, M.A.A. (1994) *Multidimensional Scaling*. London: Chapman and Hall.

Crawford, S.L. (1989) Extensions to the CART algorithm. *International Journal of Man-Machine Studies*, 31, pp. 197–217.

Cressie, N.A.C. (1981) *Statistics for Spatial Data*, New York: Wiley.

Crowder, M. J., and Hand, D. J. (1990) *Analysis of Repeated Measures*. London: Chapman and Hall.

Daly, F., Hand, D.J., Jones, M.C., Lunn, A.D., and McConway, K. (1995) *Elements of Statistics*, Wokingham, U.K.: Addison-Wesley.

Dasarathy, B.V. (ed.) (1991) *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. Los Alamitos, CA: IEEE Computer Society Press.

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

Dawes, R.M., and Smith, T.L. (1985) Attitude and opinion measurement. In *The Handbook of Social Psychology*, Volume I (3rd edition), G. Lindzey and E. Aronson (eds.), New York: Random House, pp. 509–566.

Dawid, A.P. (1984) Statistical theory: The prequential approach (with discussion). *Journal of the Royal Statistical Society A*, 147, pp. 178–292.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., and Harshman, R. (1990) Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41, pp. 391–407.

DeFinetti, B. (1974, 1975) *Theory of Probability*, Vols. 1 and 2. Chichester, U.K.: Wiley.

Della Pietra, S., Della Pietra, V., and Lafferty, J. (1997) Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4), pp. 380–393.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B,* 39, pp. 1–38.

Devijver, P.A., and Kittler, J. (1982) *Pattern Recognition: A Statistical Approach*. Englewood Cliffs, NJ: Prentice-Hall.

Devroye, L. (1984) *Nonparametric Density Estimation: the L1 View*. New York: Wiley.

Devroye, L., Gyorfi, L., and Lugosi, G. (1996) *A Probabilistic Theory of Pattern Recognition*. New York: Springer-Verlag.

Devroye, L.P., and Wagner, T.J. (1982) Nearest neighbor methods in discrimination. In *Handbook of Statistics,* vol. 2, P.R. Krishnaiah and L.N. Kanal, (eds.) Amsterdam: North-Holland, pp. 193–197.

Diaconis, P., and Shahshahani, M. (1984) On non-linear functions of linear combinations. *SIAM Journal of Scientific Computing,* 5, pp. 175–191.

Diebolt, J., and Robert, C.P. (1994) Bayesian estimation of finite mixture distributions. *Journal of the Royal Statistical Society B,* 56, pp. 363–375.

Dietterich, T.G. (1998) Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7) pp. 1895-1924.

Digby, P., and Kempton, R. (1987) *Multivariate Analysis of Ecological Communities*. London: Chapman and Hall.

Diggle, P.J., Liang, K-Y., and Zeger, S.L. (1994) *Analysis of Longitudinal Data*. Oxford, U.K.: Clarendon Press.

Domingos, P. (1996) Unifying instance-based and rule-based induction. *Machine Learning*, 24, pp. 141–168.

Domingos, P. (1999) A general method for making classifiers cost-sensitive. *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York: ACM Press, pp. 155–164.

Domingos, P., and Hulten, G. (2000) Mining high-speed data streams. *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* New York: ACM Press, pp.71–80.

Domingos, P., and Pazzani, M. (1997) On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning,* 29, pp. 103–130.

Draper, N.R., and Smith, H. (1981) *Applied Regression Analysis,* New York: Wiley.

Dryden, I.L., and Mardia, K.V. (1998) *Statistical Shape Analysis*. Chichester, UK: Wiley.

Du Mouchel, W., Volinsky, C., Johnson, T., Cortes, C., and Pregibon, D. (1999) Squashing flat files flatter. *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* New York: ACM Press, pp. 6–15.

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

Duda, R.O., Hart, P.E., and Stork, D.J. (2001) *Pattern Recognition* New York: Wiley.

Dumais, S.T., Platt, J., Heckerman, D., and Sahami, M. (1998) Inductive learning algorithms and representations for text categorization. *Proceedings of the ACM Seventh International Conference on Information and Knowledge Management,* New York: ACM Press, pp. 148–155.

Dunn, G. (1989) *Design and Analysis of Reliability Studies*. London: Arnold.

Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge, U.K.: Cambridge University Press.

Edwards, D. (1995) *Introduction to Graphical Modeling*. New York: Springer Verlag.

Edwards, A.W.F. (1972) *Likelihood*. Baltimore, MD: Johns Hopkins University Press, expanded edition.

Efron, B., and Tibshirani, R.J. (1993) *An Introduction to the Bootstrap,* New York: Chapman and Hall.

Ein-Dor, P., and Feldmesser, J. (1987) Attributes of the performance of central processing units: a relative performance prediction model. *Communications of the ACM,* 30, pp. 308–317.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Science,* 95(25), pp. 14863–68.

Elliott, R.J., Aggoun, L., and Moore, J.B. (1995) *Hidden Markov Models*. New York: Springer-Verlag.

Everitt, B.S. (1981) A Monte Carlo investigation of the likelihood ratio test for the number of components in a mixture of normal distributions. *Multivariate Behavioural Research,* 16, pp. 171–180.

Everitt, B.S., and Hand, D.J. (1981) *Finite Mixture Distributions*. London: Chapman and Hall.

Everitt, B.S., and Dunn, G. (1991) *Applied Multivariate Data Analysis*. New York: Halstead Press.

Everitt, B.S., Gourlay, A.J., and Kendell, R.E. (1971) An attempt at validation of traditional psychiatric syndromes by cluster analysis. *British Journal of Psychiatry,* 138, pp. 336–339.

Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., and Equitz, W. (1994) Efficient and effective querying by image content. *Journal of Intelligent Information Systems,* 3, pp. 231–262.

Faloutsos, C., Ranganathan, M., and Manolopoulos, Y. (1994) Fast subsequence matching in time-series databases. *Proceedings of the 1994 Annual ACM SIGMOD Conference*, New York, NY: ACM Press, pp. 419–429.

Fan, J., and Gijbels, I. (1996) *Local Polynomial Modeling and its Applications.* London: Chapman and Hall.

Fawcett, T., and Provost, F. (1997) Adaptive fraud detection. *Data Mining and Knowledge Discovery,* 1(3), pp. 291–316.

Fayyad, U.M., Djorgovski S.G., and Weir N. (1996) Automating the analysis and cataloging of sky surveys. In *Advances in Knowledge Discovery and Data Mining* U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (eds.), Menlo Park, CA: AAAI Press, pp. 471–493.

Fayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P. (1996) From data mining to knowledge discovery: an overview. In *Advances in Knowledge Discovery and Data Mining*, U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (eds.). Menlo Park, CA: AAAI Press. pp. 1–34.

Feller, W. (1968) *An Introduction to Probability Theory and its Applications*, Vol. 1 (3rd ed.) New York: Wiley.

Feng, Z.D., and McCulloch, C.E. (1996) Using bootstrap likelihood ratios in finite mixture models. *Journal of the Royal Statistical Society B*, 58(3), pp. 609–617.

Fenton, N.E. (1991) *Software Metrics*. London: Chapman and Hall.

Fine, T.L. (1999) *Feedforward Neural Network Methodology.* NewYork: Springer.

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

Fletcher, R. (1987) *Practical Methods of Optimization*. New York: Wiley.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. (1995) Query by image and video content. *IEEE Computer*, 28(9), pp. 23–31.

Florek K., Lukasziwicz J., Perkal J., Steinhaus H., and Zubrzycki S. (1951) Sur la liaison et la division des points d'un ensemble fini. *Colloquium Mathematicum*, 2, pp. 282–285.

Frakes, W.B., and Baeza-Yates, R. (eds.) (1992) *Information Retrieval: Data Structures and Algorithms*, Englewood Cliffs, N.J.: Prentice Hall.

Fraley, C., and Raftery, A.E. (1998) How many clusters? Which clustering method? answers via model-based cluster analysis. *Computer Journal*, 41, pp. 578–588.

Freund, Y., and Schapire, R.E. (1996) Experiments with a new boosting algorithm. In *Proceedings of the Thirteenth International Conference on Machine Learning*, San Francisco, CA: Morgan Kaufmann, pp. 148–156.

Friedman, J. (1997) On bias, variance, 0/1 loss, and the curse of dimensionality. *Data Mining and Knowledge Discovery*, pp. 55–77.

Friedman, J.H. (1991) Multivariate adaptive regression splines (with discussion). *Annals of Statistics,* 19, pp. 1–141.

Friedman, J.H. and Stuetzle, W. (1981) Project pursuit regression. *Journal of the American Statistical Association,* 76, pp. 817–823.

Friedman, J.H., and Fisher, N.I. (1999) Bump hunting in high dimensional data (with discussion). *Statistics and Computing*, 9, pp. 123–162.

Friedman, J.H.F., Hastie, T., and Tibshirani, R. (2000) Additive logistic regression: a statistical view of boosting, *Annals of Statistics,* 28, 377–386.

Friedman, N., and Goldszmidt, M. (1996) Learning Bayesian networks with local structure. *Proceedings of Twelfth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann, pp. 252–262.

Fukuda, T., Morimoto, Y., Morishita, S., and Tokuyama, T. (1996) Mining optimized association rules for numeric attributes. *Proceedings of the 15th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database and Knowledgebase Systems (PODS'96)*, New York: ACM Press, pp. 182–191.

Fukunaga, K. (1990) *Introduction to Statistical Pattern Recognition*, San Diego, CA: Academic Press.

Fukunaga, K., and Flick, T.E. (1984) An optimal global nearest neighbor metric. *IEEE Transactions on Pattern Recognition and Machine Intelligence,* 6, pp. 314–318.

Furnival, G.M., and Wilson, R.W. (1974) Regression by leaps and bounds. *Technometrics,* 16, pp. 499–511.

Gaffney, S., and Smyth, P. (1999) Trajectory clustering with mixtures of regression models. In *Proceedings of the ACM 1999 Conference on Knowledge Discovery and Data Mining,* New York, NY: ACM Press, pp. 63–72.

Ganti, V., Gehrke, J., and Ramakrishnan, R. (1999) Mining very large databases. *IEEE Computer,* 32, pp. 38–45.

Garcia-Molina, H., Ullman, J.D., and Widom, J. (1999) *Database System Implementation*. Englewood Cliffs, NJ: Prentice Hall.

Ge, X., and Smyth, P. (2000) Deformable Markov model templates for time series pattern-matching. *Proceedings of the ACM Seventh International Conference on Knowledge Discovery and Data Mining,* New York: ACM Press, pp. 81–90.

Gehrke, J., Ganti, V., Ramakrishnan, R., and Loh, W-Y. (1999) BOAT—optimistic decision tree construction. *Proceedings of the 1999 ACM SIGMOD conference*. New York: ACM Press, pp. 169–180.

Gehrke, J.E., Ramakrishnan, R., and Ganti, V. (1998) RainForest—a framework for fast decision tree construction of large datasets. *Proceedings of the 24th International Conference on Very Large Databases (VLDB'98),* pp. 416–427.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995) *Bayesian Data Analysis,* London: Chapman and Hall.

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

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Gill, P.E., Murray, W., and Wright, M.H. (1981) *Practical Optimization*. New York: Academic Press.

Glymour, C., Madigan, D., Pregibon, D., and Smyth, P. (1997) Statistical themes and lessons for data mining. *Data Mining and Knowledge Discovery*. 1(1), pp. 11–28.

Goer, J.C. (1967) A comparison of some methods of cluster analysis. *Biometrics,* 23, pp. 623–628.

Golden, R.M. (1996) *Mathematical Methods for Neural Network Analysis and Design*. Cambridge, MA: MIT Press.

Goldstein, H. (1995) *Multilevel Statistical Models* (2nd ed.). London: Arnold.

Gordon, A. (1981) *Classification: Methods for the Exploratory Analysis of Multivariate Data*. London: Chapman and Hall.

Gower, J.C. (1974) Maximal predictive classification. *Biometrics,* 30, pp. 643–654.

Gower, J.C., and Hand, D.J. (1996) *Biplots*. London: Chapman and Hall.

Gray, J., Bosworth, A., Layman, A., and Pirahesh, H. (1996) Data cube: a relational aggregation operator generalizing group-by, cross-tab, and sub-totals. *12th International Conference on Data Engineering (ICDE'96),* New Orleans, Louisiana, pp. 152–159.

Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., Pellow, F., and Pirahesh, H. (1997) Data Cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals. *Data Mining and Knowledge Discovery,* 1, pp. 29–53.

Grenander, U. (1996) *Elements of Pattern Theory*. Baltimore, MD: Johns Hopkins University Press.

Grimmett, G.R., and Stirzaker, D.R. (1992) *Probability and Random Processes*. (2nd ed.) Oxford: Clarenden Press.

Gusfield, D. (1997) *Algorithms on Strings, Trees and Sequences*. New York, NY: Cambridge University Press.

Hall, D.J., and Ball, G.B. (1965) ISODATA: A novel method of cluster analysis and pattern classification. Technical Report, Stanford Research Institute, Menlo Park, California.

Halstead, M.H. (1977) *Elements of Software Science*. New York: Elsevier.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hamming, R.W. (1991) *The Art of Probability for Scientists and Engineers,* Redwood City, CA: Addison-Wesley.

Han, J., and Fu, Y. (1995) Discovery of multiple-level association rules from large databases, *Proceedings of the Twenty First International Conference on Very Large Data Bases (VLDB'95),* San Mateo, CA: Morgan Kaufmann, pp. 420–431.

Han, J., and Kamber, M. (2000) *Data Mining: Concepts and Techniques*, San Francisco, CA: Morgan Kaufmann.

Hand, D.J. (1981) *Discrimination and Classification*. Chichester, U.K.: Wiley.

Hand, D.J. (1982) *Kernel Discriminant Analysis*. Chichester, U.K.: Research Studies Press.

Hand, D.J. (1986) Recent advances in error rate estimation. *Pattern Recognition Letters,* 4, pp. 335–346.

Hand, D.J. (1996) Statistics and the theory of measurement (with discussion). *Journal of the Royal Statistical Society, Series A,* 159, pp. 445–492.

Hand, D.J. (1997) *Construction and Assessment of Classification Rules*. London: Wiley.

Hand, D.J., Blunt, G., Kelly, M.G., and Adams, N.M. (2000) Data mining for fun and profit. *Statistical Science*, 15, pp. 111-131.

Hand, D.J., and Crowder, M.J. (1996) *Practical Longitudinal Data Analysis*. London: Chapman and Hall.

Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J., and Ostrowski, E. (eds.) (1994) *A Handbook of Small Data Sets*. London: Chapman and Hall.

Hand, D.J., McConway, K.J., and Stanghellini, E. (1997) Graphical models of applicants for credit. *IMA Journal of Mathematics Applied in Business and Industry*, 8, pp. 143–155.

Hand, D.J., and Yu, K. (1999) Idiot's Bayes—not so stupid after all? Working paper. Department of Mathematics, Imperial College, London.

Harman, D.K. (1993) The First Text Retrieval Conference (TREC-1), NIST SP 500-207, National Institute of Standards and Technology, Gaithersburg, Md.: (annual series, 1993–1999).

Harman, D.K., (1995) *Hypertext—Information Retrieval—Multimedia: Synergieeffekte Elektronischer Informationssysteme, Proceedings of HIM'95*, R. Kuhlen and M. Rittberger (eds.), Konstanz, Germany: Universitaetsforlag Konstanz, pp. 9–28.

Harrison, D. (1993) Backing up. *Neural Computing*, pp. 98–104.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models, and the Kalman Filter*. Cambridge, UK: Cambridge University Press.

Hastie, T., and Tibshirani, R.J. (1990) *Generalized Additive Models*. London: Chapman and Hall.

Hastie, T., and Tibshirani, R.J. (1996) Discriminant adaptive nearest neighbor classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18, pp. 607–616.

Heckerman, D., Chickering, D.M., Meek, C., Rounthwaite, R., and Kadie, C. (2000) Dependency networks for inference, collaborative filtering, and data visualization. *Journal of Machine Learning Research*, 1, pp. 49–75.

Hendry, D.F. (1995) *Dynamic Econometrics*. New York: Oxford University Press.

Hilden, J. (1984) Statistical diagnosis based on conditional independence does not require it. *Computers in Biology and Medicine,* 14, pp. 429–435.

Hjort, J.S.U. (1993) *Computer Intensive Statistical Methods: Validation, Model Selection, and Bootstrap*. Boca Raton, FL: CRC Press.

Ho, T.K., Hull J.J., and Srihari, S.N. (1994) Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 16, pp. 66–75.

Hoffmann, T. (1999) Probabilistic latent sematic indexing. *Proceedings of the ACM SIGIR Conference 1999,* New York: ACM Press, pp. 50–57.

Holsheimer, M., Kersten, M., Mannila, H., and Toivonen, H. (1995) A perspective on databases and data mining. *Proceedings of the First International Conference on knowledge discovery and data mining,* Fayyad, U.M., and Uthurusamy, R. (eds.), Menlo Park, CA: AAAI Press, pp. 150–155.

Holte, R.C., (1993) Very simple classification rules perform well on most commonly used data sets. *Machine Learning*, 11, pp. 63–91.

Huba, G.J., Wingard, J.A., and Bentler, P.M. (1981) A comparison of two latent variable causal models for adolescent drug use. *Journal of Personality and Social Psychology,* 40, pp. 180–193.

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

Huber, P.J. (1980) *Robust Statistics*. New York: Wiley.

Hunter, J.S. (1980) The national system of scientific measurement. *Science*, 210, 21 November 1980, pp. 869-874.

Hyvarinen, A. (1999) Survey on independent component analysis. *Neural Computing Surveys*, 2, pp. 94–128.

Imielinski, T., and Mannila, H. (1996) A database perspective on knowledge discovery. *Communications of the ACM,* 39(11), pp. 58–64.

Imielinski, T., and Virmani, A. (1999) MSQL: A query language for database mining. *Data Mining and Knowledge Discovery* 3(4), pp. 373–408.

Imielinski, T., Virmani, A., and Abdulghani, A. (1999) DMajor application programming interface for database mining. *Data Mining and Knowledge Discovery,* 3(4), pp. 347–372.

Jacoby, W.G. (1997) *Statistical Graphics for Univariate and Bivariate Data*. London: Sage Publications.

Jain, A., and Dubes, R. (1988) *Algorithms for Clustering Data*., Englewood Cliffs, Prentice-Hall.

Jensen, F.V. (1996) *An Introduction to Bayesian Networks*. New York: Springer-Verlag.

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

Jordan, M.I. (1999) *Learning in Graphical Models,* Cambridge, MA: MIT Press.

Jordan, M.I., and Jacobs, R.A. (1994) Hierarchical mixtures of experts and the EM algorithm. *Neural Computation,* 6, pp. 181-214.

Journal of the American Society for Information Science (1996) Special Issue on Evaluation, 47:1–105.

Karypis, G., and Kumar, V. (1998) A parallel algorithm for multilevel graph partitioning and sparse matrix ordering. *Journal of Parallel and Distributed Computing*, 48(1), pp. 71–95.

Kass, R., and Raftery, A. (1995) Bayes factors. *Journal of the American Statistical Association,* 90, pp. 773–795.

Kato, T., Kurita, T., and Shimogaki, H. (1991) Intelligent visual interaction with image database systems—towards the multimedia personal interface. *Information Processing* (Japan), 14, pp. 134–143.

Kaufman, L., and Rousseeuw, P.J. (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley.

Keim, D.A., and Kriegel, H.-P. (1994) VisDB: database exploration using multidimensional visualization. *IEEE Computer Graphics and Applications,* September 1994, pp. 40–49.

Kendall, M.G. (1980) *Multivariate Analysis* (2nd ed.). London: Griffin.

Keogh, E., and Smyth, P. (1997) A probabilistic approach to fast pattern matching in time series databases. *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining,* Menlo Park, CA: AAAI Press, pp. 24–30.

Kim, C-J., and Nelson, C.R. (1999) *State-Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*. Cambridge, MA: MIT Press.

Kirkpatrick, S., Gelatt, C.D. Jr., and Vecchi, M.P. (1983) Optimization by simulated annealing. *Science,* 220, pp. 671–680.

Kish, L. (1965) *Survey Sampling*. New York: Wiley.

Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., and Verkamo, A.I. (1994) Finding interesting rules from large sets of discovered association rules. *Proceedings of the Third International Conference on Information and Knowledge Management (CIKM'94),* New York: ACM Press, pp. 401–407.

Knight, K. (2000) *Mathematical Statistics*, Boca Raton, FL: Chapman and Hall.

Knuth, D. (1997). *The Art of Computer Programming: Fundamental Algorithms,* 3rd ed. Reading, MA: Addison Wesley.

Kohavi, R.(1996) Scaling up the accuracy of naive-Bayes classifiers: A decision-tree hybrid. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Portland, OR: AAAI Press, pp. 202–207.

Koontz, W.L.G., Narendra, P.M., and Fukunaga, K. (1975) A branch and bound clustering algorithm. *IEEE Transactions on Computers,* 24, pp. 908–915.

Krantz, D.H., Luce, R.D., Suppes, P., and Tversky, A. (1971) *Foundations of Measurement, Volume 1: Additive and Polynomial Representations*. New York: Academic Press.

Krzanowski, W.J., and Marriott, F.H.C. (1995) *Multivariate Analysis vol. 2: Classification, Covariance Structures, and Repeated Measurements*. London: Arnold.

Lambert, J.M., and Williams, W.T. (1966) Multivariate methods in plant ecology IV: comparison of information analysis and association analysis. *Journal of Ecology,* 54, pp. 635–664.

Lance, G.N., and Williams, W.T. (1967) A general theory of classificatory sorting strategies: 1. Hierarchical systems. *Computer Journal,* 9, pp. 373–380.

Landauer, T.K., and Dumais, S.T., (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review,* 104(2), pp. 211–240.

Lange, K. (1995) A gradient algorithm locally equivalent to the EM algorithm. *Journal of the Royal Statistical Society B,* 57, pp. 425–437.

Lange, K. (1999) *Numerical Analysis for Statisticians*. New York: Springer-Verlag.

Lapointe, F.J., and Legendre, P. (1994) A classification of pure malt Scotch whiskies. *Applied Statistics,* 43, pp. 237–257.

Lauritzen, S.L. (1996) *Graphical Models*. Oxford: Clarendon Press.

Lavine, M. (1991) Problems in extrapolation illustrated with space shuttle O-ring data. *Journal of the American Statistical Association*, 86, pp. 919–922.

Lavrac, N., and Dzeroski, S. (1994) *Inductive Logic Programming: Techniques and Applications*. Ellis Horwood.

Lawrence, R.D., Almasi, G.S., Kotlyar, V., Viveros, M.S., and Duri, S.S. (2001) Personalization of supermarket product recommendations, *Data Mining and Knowledge Discovery,* to appear.

Leamer, E.E. (1978) *Specification Searches: Ad Hoc Inference with Experimental Data*. New York: Wiley.

Lee, P.M. (1989) *Bayesian Statistics: An Introduction*. London: Edward Arnold.

Lehmann, E.L. (1986) *Testing Statistical Hypotheses*. New York: Wiley.

Lehmann, E.L., and Casella, G. (1998) *Theory of Point Estimation,* New York: Springer-Verlag.

Leighton, G., and McKinlay, P.L. (1930) *Milk Consumption and the Growth of School Children*. London: HMSO.

Leinweber, D. (personal communication) Stupid data miner tricks: Overfitting the S&P 500.

Lewis, H.R., and Papadimitriou, C.H. (1998) *Elements of the Theory of Computation*, second edition. Upper Saddle River, NJ: Prentice-Hall.

Li, M., and Vitanyi, P. (1993) *An Introduction to Kolmogorov Complexity and Its Applications*. New York: Springer.

Lindsey, I. (1994) *Credit Cards: The Authoritative Guide to Payment and Credit Cards*. Leighton Buzzard: Rushmere Wynne.

Lindsey, J.K. (1996) *Parametric Statistical Inference*. Oxford, U.K.: Clarendon Press.

Lindsey, J.K. (1999) *Models for Repeated Measurements,* 2nd ed. Oxford, U.K.: Oxford University Press.

Lindsey, J.K. (1999) Relationships among sample size, model selection and likelihood regions, and scientifically important differences. *Journal of the Royal Statistical Society, Series D*, 48, pp. 401–411.

Linhart, H., and Zucchini, W. (1986) *Model Selection*. New York: Wiley.

Little, R.J.A., and Rubin, D.B. (1987) *Statistical Analysis with Missing Data*. New York: Wiley.

Looney, C.G. (1997) *Pattern Recognition Using Neural Networks*. Oxford, U.K.: Oxford University Press.

Lovell, M.C. (1983) Data mining. *Review of Economics and Statistics* 65(1), pp. 1–12.

Luce, R.D., Krantz, D.H., Suppes, P., and Tversky, A. (1990) *Foundations of Measurement, Volume 3: Representation, Axiomatization, and Invariance*. San Diego, CA: Academic Press.

Luenberger, D.G. (1984) *Introduction to Linear and Nonlinear Programming*. Menlo Park, CA: Addison-Wesley.

MacDonald, I.L., and Zucchini, W. (1997) *Hidden Markov and Other Models for Discrete-valued Time Series*. London: Chapman and Hall.

MacKay, D.J.C. (1992) A practical Bayesian framework for back-propagation networks. *Neural Computation*, 4, pp. 448–472.

MacMillan, N.A., and Creelman, C.D. (1991) *Signal Detection Theory: A User's Guide*, New York, NY: Cambridge University Press.

MacNaughton-Smith, P., Williams, W.T., Dale, M.B., and Mockett, L.G. (1964) Dissimilarity analysis. *Nature,* 202, pp. 1034–1035.

MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* L.M. Le Cam, and J. Neyman (eds.) Berkeley: University of California Press, pp. 281–297.

Madigan, D., Raghavan, N., DuMouchel, W., Nason, M., Posse, C., and Ridgeway, G. (in press) Likelihood-based data squashing: A modeling approach to instance construction. *Data Mining and Knowledge Discovery*.

Mangasarian, O. (1997) Mathematical programming in data mining. *Data Mining and Knowledge Discovery*, 1(2), pp.183–201.

Mannila, H. (1997) Inductive databases and condensed representations: Concepts for data mining. *International Logic Programming Symposium 1997,* Cambridge, MA: MIT Press, pp. 21–30.

Mannila, H., Toivonen, H., and Verkamo, A.I. (1994) Efficient algorithms for discovering association rules. *Knowledge Discovery in Databases: Papers from the AAAI-94 Workshop (KDD'94),* Menlo Park, CA: AAAI Press, pp. 181–192.

Mannila, H., Toivonen, H., and Verkamo, A.I. (1997) Discovery of frequent episodes in sequences. *Data Mining and Knowledge Discovery*. 1(3), pp. 259–290.

Maritz, J.S. (1981) *Distribution-Free Statistical Methods*. London: Chapman and Hall.

Marriott, F.H.C. (1971) Practical problems in a method of cluster analysis. *Biometrics,* 27, pp. 501-514.

Maybury, M.T. (ed.) (1997) *Intelligent Multimedia Information Retrieval*. Menlo Park, CA: AAAI Press.

McCullagh, P., and Nelder, J.A. (1989) *Generalized Linear Models,* 2nd ed. London: Chapman and Hall.

McKendrick, A.G. (1926) Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society,* 44, pp. 98–130.

McLachlan, G.J. (1987) Error rate estimation in discriminant analysis: recent advances. In *Advances in Multivariate Statistical Analysis,* A.K. Gupta, ed. The Netherlands: Reidel, pp. 233–252.

McLachlan, G.J. (1987) On bootstrapping the likelihood ratio test for the number of components in a normal mixture. *Applied Statistics,* 36, pp. 318–324.

McLachlan, G.J. (1992) *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley.

McLachlan, G.J., and Basford, K.E. (1988) *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.

McLachlan, G.J., and Krishnan, T. (1998) *The EM Algorithm and Extensions.* New York: Wiley.

McLachlan, G.J., and Peel, D. (1997) On a resampling approach to choosing the number of components in normal mixture models. In *Computing Science and Statistics (Vol 28),* L. Billard, and N.I. Fisher (eds.). Fairfax Station, VA: Interface Foundation of North America, pp. 260–266.

McLachlan, G.J., and Peel, D. (1998) MIXFIT: An algorithm for the automatic fitting and testing of normal mixture models. *Proceedings of the 14th International Conference on Pattern Recognition,* vol. 1, Los Alamitos, CA: IEEE Computer Society, pp. 553–557.

McLachlan, G.J., and Peel, D. (2000) *Finite Mixture Models*. New York: Wiley.

McLaren, C.E. (1996) Mixture models in haematology: A series of case studies. *Statistical Methods in Medical Research,* 5, pp. 129–153.

Meilijson, I. (1989) A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society B,* 51, pp. 127–138.

Mendell, N.R., Finch, S.J., and Thode, H.C. (1993) Where is the likelihood ratio test powerful for detecting two component normal mixtures? *Biometrics,* 49, pp. 907–915.

Meo, R., Psaila, G., and Ceri, S. (1996) A new SQL-like operator for mining association rules. *Proceedings of the 22nd International Conference on Very Large Data Bases (VLDB'96),* San Mateo, CA: Morgan Kaufmann.

Michell, J. (1986) Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin,* 100, pp. 398–407.

Michell, J. (1990) *An Introduction to the Logic of Psychological Measurement*. Hillsdale: Lawrence Erlbaum.

Mitchell, M. (1997) *An Introduction to Genetic Algorithms.* Cambridge, MA: MIT Press.

Mitchell, T. (1997) *Machine Learning,* New York: McGraw-Hill.

Moore, A. (1999) Very fast EM-based mixture model clustering using multiresolution kd-trees. In *Advances in Neural Information Processing Systems 12,* San Francisco, CA: Morgan Kaufmann.

Moore, A.W. (1999) Cached sufficient statistics for automated discovery and data mining from massive data sources. Online white paper, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Moore, A.W., and Lee, M. (1998) Cached sufficient statistics for efficient machine learning with large data sets. *Journal of Artificial Intelligence Research,* 8, pp. 67–91.

Morgan, B.J.T. (1981) Three applications of methods of cluster analysis. *The Statistician,* 30, pp. 205-223.

Morgan, J.N., and Sonquist, J.A. (1963) Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association,* 58, pp. 415–434.

Mosteller, F. (1968) Nonsampling errors. In *International Encyclopedia of the Social Sciences*, 5, D.L. Sills (ed.), New York: MacMillan and Free Press, pp. 113–132.

Muggleton, S. (1995) *Foundations of Inductive Logic Programming,* Englewood Cliffs, NJ: Prentice Hall.

Murthy, S.K. (1998) Automatic construction of decision trees from data: A multi-disciplinary survey. *Data Mining and Knowledge Discovery,* 2, pp. 345–389.

Myles, J.P., and Hand, D.J. (1990) The multi-class metric problem in nearest neighbour discrimination rules. *Pattern Recognition,* 23, pp. 1291–1297.

Nakhaeizadeh, G., and Taylor, C.C. (eds.) (1997) *Machine Learning and Statistics*. New York: Wiley.

Neal, R. (1996) *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics 118, New York: Springer.

Neal, R., and Hinton, G. (1998) A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models*, Jordan, M.I. (ed.), Cambridge, MA: MIT Press, pp. 355–371.

Nering, E.D., and Tucker, A.W. (1993) *Linear Programs and Related Problems*. Academic Press Inc.

Newcomb, S. (1886) A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, 8, pp. 343–366.

Nightingale, F. (1858) *Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army, founded chiefly on the Experience of the Late War*. London: Harrison.

Oliver, J.J., and Hand, D.J. (1996) Averaging over decision trees. *Journal of Classification*, 13, pp. 281–297.

Papadimitriou, C.H., and Steiglitz, K (1982) *Combinatorial Optimization—Algorithms and Complexity*. Englewood Cliffs, NJ: Prentice-Hall.

Park, J.S., Chen, M.S., and Yu, P.S. (1995) An effective hash-based algorithm for mining association rules. *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD'95)*, New York: ACM Press, pp. 175–186.

Pearl, J. (1984) *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. Reading, MA: Addison-Wesley.

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

Peixoto, J.L. (1990) A property of well-formulated polynomial regression models. *American Statistician*, 44, pp. 26–30.

Pentland, A., Picard, R.W., and Sclaroff, S. (1994) Photobook: Tools for content-based manipulation of image databases. *International Journal of Computer Vision*, 18, pp. 233–254.

Piatetsky-Shapiro, G. (1991) Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases*. G. Piatetsky-Shapiro and W. Frawley (eds.), Menlo Park, CA: AAAI Press.

Piatetsky-Shapiro, G. (1999) The data-mining industry coming of age. *IEEE Expert,* 14(6), pp. 32–34.

Platt, J. (1999) Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods—Support Vector Learning,* B. Scholkopf, C.J.C. Burges, and A.J. Smola (eds.), Cambridge, MA: MIT Press, pp. 185–208.

Poulsen, C.S. (1990) Mixed Markov and latent Markov modelling applied to brand choice behavior. *International Journal of Research in Marketing,* 7, pp. 5–19.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1988) *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, UK: Cambridge University Press.

Provost, F., and Kolluri, V. (1999) A survey of methods for scaling up inductive algorithms. *Data Mining and Knowledge Discovery,* 3, pp. 131–169.

Provost, F., Jensen, D., and Oates, T. (1999) Efficient progressive sampling. *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* New York: ACM Press, pp. 23–32.

Quandt, R.E., and Ramsey, J.B. (1978) Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, 73(364), 730–738.

Quinlan, J.R. (1986) Induction of decision trees. *Machine Learning,* 1, pp. 81–106.

Quinlan, J.R. (1987) Generating production rules from decision trees. *Proceedings of the Tenth International Joint Conference on Artificial Intelligence,* San Mateo, CA: Morgan Kaufmann, pp. 304–307.

Quinlan, J.R. (1990) Learning logical definitions from relations, *Machine Learning,* 5, pp. 239–266.

Quinlan, J.R. (1993) *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.

Raghavan, V.V., and Wong, S.K.M. (1986) A critical analysis of the vector space model for information retrieval. *Journal of the American Society for Information Science,* 37(5), pp. 100–124.

Ramakrishnan, R., and Gehrke, J. (1999) *Database Management Systems*, Second Edition. New York: McGraw Hill.

Ramsey, J.O., and Silverman, B.W. (1996) *Functional Data Analysis*. New York: Springer-Verlag.

Randles, R.H., and Wolfe, D.A. (1979) *Introduction to the Theory of Nonparametric Statistics*. New York: Wiley.

Rao, M.R. (1971) Cluster analysis and mathematical programming. *Journal of the American Statistical Association,* 66, pp. 622–626.

Rastogi, R., and Shim, K. (1998) PUBLIC: A decision tree classifier that integrates building and pruning. *Proceedings of the 24th International Conference on Very Large Databases (VLDB'98),* pp. 405–415.

Redner, R.A., and Walker, H.F. (1984) Mixture densities, maximum likelihood, and the EM algorithm. *SIAM Review,* 26, pp. 195–239.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994) GroupLens: An open architecture for collaborative filtering of netnews. *Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work,* Chapel Hill, N.C.: ACM Press, pp. 175-186.

Reyment, R., and Jöreskog K.G. (1993) *Applied Factor Analysis in the Natural Sciences*, Cambridge: Cambridge University Press.

Ridgeway, G. (1997) Finite discrete Markov process clustering. Technical Report TR 97-24, Microsoft Research, Redmond, WA.

Ripley, B.D. (1996) *Pattern Recognition and Neural Networks*. Cambridge, U.K.: Cambridge University Press.

Rissanen, J. (1987) Stochastic complexity (with discussion). *Journal of the Royal Statistical Society, Series B,* 49, pp. 223–239 and pp. 253–265.

Robbins, H., and Monro, S. (1951) A stochastic approximation method. *Annals of Mathematical Statistics*, 22, pp. 400–407.

Roberts, F.S. (1979) *Measurement Theory with Applications to Decision-making, Utility, and the Social Sciences*. Reading: Addison-Wesley.

Rocchio, J.J. (1971) Relevance feedback in information retrieval. *The SMART Retrieval System: Experiments in Automatic Document Processing,* Salton, G. (ed.). Englewood Cliffs, N.J.: Prentice Hall, pp. 313–323.

Ross, S.M. (1997) *Introduction to Probability Models*. San Diego, CA: Academic Press, 6th ed.

Rui, Y., Huang, T.S., Ortega, M., and Mehrotra, S. (1997) Relevance feedback: a power tool in interactive content-based image retrieval. *Proceedings of the IEEE Transactions on Circuits and Systems for Video Technology*, 8(5), pp. 644–655.

RuleQuest Research (2000) http://www.rulequest.com/cubist-info.html.

Russek, E., Kronmal, R.A., and Fisher, L.D. (1983) The effect of assuming independence in applying Bayes' theorem to risk estimation and classification in diagnosis. *Computers and Biomedical Research,* 16, pp. 537–552.

Salton, G. (ed.) (1971) *The SMART Retrieval System: Experiments in Automatic Document Processing*. Englewood Cliffs, N.J.: Prentice Hall.

Salton, G., and Buckley, C. (1988) Term-weighting approaches in automatic text retrieval. *Information Processing and Management,* 24:513-523.

Salton, G., and Buckley, C. (1990) Improving retrieval performance by relevance feedback. *Journal of the American Society of Information Science,* 41(4), pp. 288-297.

Salton, G., and McGill, M. (1983) *Introduction to Modern Information Retrieval,* New York: McGraw Hill.

Salzberg, S. (1997) On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery*, 1(3), pp. 317–327.

Salzberg, S. (1999) Gene discovery in DNA sequences. *IEEE Expert,* 14(6), pp. 44–48.

Sarawagi, S., Thomas, S., and Agrawal, R. (1998) Integrating mining with relational database systems: Alternatives and implications. *Proceedings of the ACM SIGMOD Conference on Mangement of Data (SIGMOD 1998),* New York: ACM Press, pp. 343–354.

Sarawagi, S., Thomas, S., and Agrawal, R. (2000) Integrating association rule mining with relational database systems: alternatives and implications. *Data Mining and Knowledge Discovery,* 4, pp. 89–125.

Savasere, A., Omiecinski, E., and Navathe, S. (1995) An efficient algorithm for mining association rules. *Proceedings of the 21st International Conference on Very Large Data Bases (VLDB'95),* San Mateo, CA: Morgan Kaufmann, pp. 432–444.

Schafer, J.B., Konstan, J., and Riedl, J. (in press) Electronic commerce recommender applications. *Data Mining and Knowledge Discovery*.

Schaffer, C. (1994) Cross-validation, stacking and bi-level stacking: Meta-methods for classification and learning. In *Selecting Models from Data: AI and Statistics IV*, P. Cheeseman and R.W. Oldford (eds.), New York: Springer-Verlag.

Schapire, R.E., Freund, Y., Bartlett, P., and Lee, W.S. (1998) Boosting the margin: A new explanation for the effectiveness of voting methods. *The Annals of Statistics,* 26(5), pp. 1651–1686.

Schervish, M.J. (1995) *Theory of Statistics.* New York: Springer-Verlag.

Schiavo, R., and Hand, D.J. (2000) Ten more years of error rate research. *International Statistical Review*, 68, pp. 295-310.

Scholkopf, B., Burges, C.J.C., and Smola, A.J. (eds.) (1999) *Advances in Kernel Methods—Support Vector Learning.* Cambridge, MA: MIT Press.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics,* 6, pp. 461–464.

Scott, D.F. (1992) *Multivariate Density Estimation: Theory and Visualization*. New York: Wiley.

Segal, R., and Etzioni, O. (1994) Learning decision lists using homogenous rules. *Proceedings of the Twelfth National Conference on Artificial Intelligence,* Menlo Park, CA: AAAI Press, pp. 619–625.

Shafer, G., and Pearl, J. (1990) *Readings in Uncertain Reasoning*. San Mateo: CA, Morgan Kaufman.

Shafer, J.C., Agrawal, R., and Mehta, M. (1996), SPRINT: A scalable parallel classifier for data mining. *Proceedings of the 22nd International Conference on Very Large Databases (VLDB'96),* San Francisco, CA:Morgan Kaufmann, pp. 544–555.

Shardanand, U., and Maes, P., (1995) Social information filtering: Algorithms for automating "word of mouth."*Proceedings of CHI'95–Human Factors in Computing Systems,* pp. 210-217.

Shaw, S.W., Defigueiredo, R.J.P. (1990) Structural processing of waveforms as trees. *IEEE Transactions on Acoustic, Speech, and Signal Processing,* 38(2), pp. 328–338.

Shepard, R.N., and Arabie, P. (1979) Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review,* 86, pp. 87–123.

Short, R.D., and Fukunaga, K. (1981) The optimal distance measure for nearest neighbor classification. *IEEE Transactions on Information Theory,* 27, pp. 622–627.

Shoshani, A. (1997) OLAP and statistical databases: Similarities and differences. *Proceedings of the Sixteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, New York: ACM Press, pp. 185–196.

Sibson, R. (1973) SLINK: An optimally efficient algorithm for the single link method. *Computer Journal,* 16, pp. 30–34.

Silberschatz, A., and Tuzhilin, A. (1996) What makes patterns interesting in knowledge-discovery systems, *IEEE Transactions on Knowledge and Data Engineering,* 8(6), pp. 970–974.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Simpson, C.H. (1951) The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B,* 13, pp. 238–241.

Smith, J.R., and Chang, S. (1997) Querying by color regions using VisualSEEk content-based visual query system. *Intelligent Multimedia Information Retrieval,* Maybury, M.T. (ed.). Menlo Park, CA: AAAI Press, pp. 23–41.

Smoliar, S., and Zhang, H. (1994) Content-based video indexing and retrieval. *IEEE Multimedia,* 1, pp. 62–72.

Smyth, P. (1994) Hidden Markov models for fault detection in dynamic systems. *Pattern Recognition*, 27(1), pp.149–164.

Smyth, P. (1997) Clustering sequences using hidden Markov models. In *Advances in Neural Information Processing 9,* M.C. Mozer, M.I. Jordan, and T. Petsche (eds.), Cambridge, MA: MIT Press, pp. 648–654.

Smyth, P. (1999) Probabilistic model-based clustering of multivariate and sequential data. In *Proceedings of the Seventh International Workshop on AI and Statistics,* D. Heckerman, and J. Whittaker eds., San Francisco, CA: Morgan Kaufman, pp. 299–304.

Smyth, P. (2000) Data mining: Data analysis on a grand scale? *Statistical Methods in Medical Research*. 9, pp. 309–327.

Smyth, P. (2000) Model selection for probabilistic clustering using cross-validated likelihood. *Statistics and Computing*, 9, pp. 63–72.

Smyth, P., and Goodman, R. (1992) An information-theoretic approach to rule induction from databases. *IEEE Transactions on Knowledge and Data Engineering,* 4(4), pp. 301–306.

Smyth, P., Ide, K., and Ghil, M. (1999) Multiple regimes in northern hemisphere height fields via mixture model clustering. *Journal of the Atmospheric Sciences,* 56(21), pp. 3704–3723.

Sparck Jones, K., and Willett, P. (1997) *Readings in Information Retrieval.* San Francisco: Morgan Kaufmann.

Späth, H. (1979) Clusterwise linear regression. *Computing,* 22(4), pp. 367–73.

Späth, H. (1985) *Cluster Analysis and Dissection*. Chichester, U.K.: Ellis Horwood.

Srikant, R., and Agrawal, R. (1995) Mining generalized association rules. *Proceedings of the 21st International Conference on Very Large Data Bases (VLDB'95),* San Mateo, CA: Morgan Kaufmann, pp. 407–419.

Srikant, R., and Agrawal, R. (1996) Mining quantitative association rules in large relational tables. *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD'96),* New York: ACM Press, pp. 1–12.

Srikant, R., Vu, Q., and Agrawal, R. (1997) Mining association rules with item constraints. *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD'97),* Heckerman, D., Mannila, H., and Pregibon, D. (eds.). Menlo Park, CA: AAAI Press, pp. 67–73.

Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions (with Discussion). *Journal of the Royal Statistical Society, Series B,* 36, pp. 111–147.

Streiner, D.L., and Norman, G.R. (1995) *Health Measurement Scales*, second edition. Oxford: Oxford University Press.

Swanson, D.R. (1987) Two medical literatures that are logically but not bibliographically connected. *Journal of the American Society for Information Retrieval*, 38(4), pp. 228–233.

Swanson, D.R., and Smalheiser, N.R. (1994) Assessing a gap in the biomedical literature: Magnesium deficiency and neurologic disease. *Neuroscience Research Communications*, 15, pp. 1–9.

Swanson, D.R., and Smalheiser N.R. (1997) An interactive system for finding complementary literatures: A stimulus to scientific discovery. *Artificial Intelligence*, 91, pp. 183–203.

Suppes, P., Krantz, D.H., Luce, R.D., and Tversky, A. (1989) *Foundations of Measurement, Volume 2: Geometrical, Threshold, and Probabilistic Representations*. San Diego, CA: Academic Press.

Szalay, A.S., Kunszt, P., Thakar, A., and Gray, J. (1999) Designing and mining multi-terabyte astronomy archives: The Sloan Digital Sky Survey. Technical Report MS-TR-99-30, San Francisco, CA: Microsoft Research.

Thall, P.F., and Vail, S.C. (1990) Some covariance models for longitudinal count data with overdispersion. *Biometrics,* 46, pp. 657-671.

Thisted, R.A., (1988) *Elements of Statistical Computing*. London, Chapman and Hall.

Titterington, D.M., Smith, A.F.M., and Makov, U.E. (1985) *Statistical Analysis of Finite Mixture Distributions*. Chichester, U.K.: Wiley.

Toivonen, H. (1996) Sampling large databases for association rules, *Proceedings of the Twenty Second International Conference on Very Large Data Bases (VLDB'96),* San Mateo, CA: Morgan Kaufmann, pp. 134–145.

Toussaint, G.T. (1974) Bibliography on estimation of misclassification. *IEEE Transactions on Information Theory,* 20, pp. 472–479.

Tsur, D., Ullman, J.D., Abiteboul, S., Clifton, C., Motwani, R., Nestorov, S., and Rosenthal, A. (1998) QueryFlocks: A generalization of association rule mining. *Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD'98),* New York, NY: ACM Press, pp. 1–12.

Tufte, E.R. (1983) *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Tufte, E.R. (1990) *Envisioning Information*. Cheshire, CT: Graphics Press.

Tukey, J.W. (1977) *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Ullman, J.D. (1988) *Principles of Database and Knowledge-Base Systems*, vol. 1. Rockville, MD: Computer Science Press.

Ullman, J.D., and Widom, J. (1997) *A First Course in Database Systems*. Upper Saddle River, NJ: Prentice-Hall.

van Laarhoven, P.J.M., and Aarts, E.H.L. (1987) *Simulated Annealing: Theory and Applications*. Dordrecht, Netherlands: D. Reidel.

Van Rijsbergen, C.J. (1979) *Information Retrieval*. London: Butterworth Press.

Vapnik, V. (1995) *The Nature of Statistical Learning Theory*. Berlin: Springer-Verlag.

Vapnik, V. (1998) *Statistical Learning Theory*. Chichester, U.K.: Wiley.

Wand, M.P., and Jones, M.C. (1995) *Kernel Smoothing*. London: Chapman and Hall.

Wang, J.T., Zhang, K., Jeong, K., and Shasha, D. (1994) A system for approximate tree matching. *IEEE Transactions on Knowledge and Data Engineering,* 6(4), 559–571.

Webb, A. (1999) *Statistical Pattern Recognition*. London: Arnold.

Webb, G. (2000) Efficient search for association rules. *Proceedings of the ACM Seventh International Conference on Knowledge Discovery and Data Mining,* New York, NY: ACM Press, pp. 300–304.

Wedel, M., and Kamakura, W.A. (1998) *Market Segmentation: Conceptual and Methodological Foundations*. Boston, MA: Kluwer.

Wegman, E.J. (1990) Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association,* 85(411), pp. 664–675.

Weiss, S., and Indurkhya, N. (1993) Rule-based regression. *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI-93,* San Mateo, CA: Morgan Kaufmann, pp. 1072–1078.

Weiss, S., and Indurkhya, N. (1995) Rule-based machine learning methods for functional prediction. *Journal of Artificial Intelligence Research,* 3, pp. 383–403.

Weiss, S.M., and Indurkhya, N. (1998) *Predictive Data Mining: A Practical Guide*. San Francisco, CA: Morgan Kaufmann.

Whittaker, J. (1990) *Graphical Models in Applied Multivariate Statistics,* Chichester, U.K.: Wiley.

Wilkinson, L. (1999) *The Grammar of Graphics*. New York: Springer-Verlag.

Witten, I.H., and Franke, E. (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco, CA: Morgan Kaufmann.

Witten, I.H., Moffat, A., and Bell, T.C. (1999) *Managing Gigabytes: Compressing and Indexing Documents and Images,* 2nd ed. San Francisco, CA: Morgan Kaufmann.

Xu, L., Krzyzak, A., and Suen, C.Y. (1992) Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22, pp. 418–435.

Zamir, O., and Etzioni, O. (1998) Web document clustering: A feasibility demonstration. *Proceedings of the 21st International ACM SIGIR Conference*, New York: ACM Press, pp. 46–54.

Zhang, T., Ramakrishnan, R., and Livny, M. (1997) BIRCH: an efficient data clustering method for very large databases. *Data Mining and Knowledge Discovery,* 1(2), pp. 141–182.