# Bibliography


[1] R. Agrawal, T. Imielinski, and A.N. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207–216, Washington, D.C., 1993.

[2] T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons Ltd., 1958.

[3] S. Baraty, D.A. Simovici, and C. Zara. The impact of triangular inequality violations on medoid-based clustering. In *Foundations of Intelligent Systems*, volume 6804 of *Lecture Notes in Computer Science*, pages 280–289. Springer Berlin / Heidelberg, 2011.

[4] H. Blockeel and L. De Raedt. Lookahead and discretization in ILP. In *Proceedings of the Seventh International Workshop on Inductive Logic Programming*, pages 77–84, Berlin, 1997. Springer.

[5] H. Blockeel and L. De Raedt. Top-down induction of first order logical decision trees. *Artificial Intelligence*, 101(1-2):285–297, 1998.

[6] R.E. Bohn and J.E. Short. How much information? 2009 report on American consumers. Technical report, Global Information Industry Center, University of California, San Diego, 2009.

[7] U. Bohnebeck, T. Horváth, and S. Wrobel. Term comparisons in first-order similarity measures. In *Proceedings of the Eighth International Conference on Inductive Logic Programming*, pages 65–79, Berlin, 1998. Springer.

[8] D. Boley. Principal direction divisive partitioning. *Data Mining and Knowledge Discovery*, 2(4):325–344, 1998.

[9] L. Breiman, J.H. Friedman, R. Olshen, and C.J. Stone. *Classification and Regression Trees*. Chapman and Hall, New York, 1984.

[10] S. Brin, R. Motwani, J. Ullman, and S. Tsur. Dynamic itemset counting and implication rules for market basket data. In *Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data*, pages 255–264, Tucson, AZ, 1997.

[11] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30(1-7):107–117, 1998.

[12] E.F. Codd. A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387, 1970.

[13] W.W. Cohen. Fast effective rule induction. In *Proceedings of the 12th International Conference on Machine Learning*, pages 115–123, Tahoe City, CA, 1995.

[14] T.F. Coleman and J. Moré. Estimation of sparse jacobian matrices and graph coloring problems. *SIAM Journal on Numerical Analysis*, 20(1):187–209, 1983.

[15] T.M. Connolly and C.E. Begg. *Database Systems: A Practical Approach to Design, Implementation and Management*. Addison-Wesley Publishing Company, 5th edition, 2009.

[16] D.J. Cook and L.B. Holder. Substructure discovery using minimum description length and background knowledge. *Journal of Artificial Intelligence Research*, 1:231–255, 1994.

[17] D.J. Cook and L.B. Holder. Graph-based data mining. *IEEE Intelligent Systems*, 15(2):32–41, 2000.

[18] T.M. Cover and P.E. Hart. Nearest neighbor pattern classification. *Knowledge Based Systems*, 8:373–389, 1995.

[19] L. Dehaspe and H. Toivonen. Discovery of frequent datalog patterns. *Data Mining and Knowledge Discovery*, 3(1):7–36, 1999.

[20] S. Džeroski. Data mining in a nutshell. In *Relational Data Mining*, chapter 1. Springer-Verlag, 2001.

[21] C. Elkan. Using the triangle inequality to accelerate $k$-means. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML)*, pages 147–153, Washington DC, US, 2003.

[22] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pages 226–231, Portland, OR, 1996. AAAI Press.

[23] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*, pages 1–34, Cambridge, MA, 1996. MIT Press.

[24] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

[25] E. Fix and J.L. Hodges. Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

[26] K. Florek, J. Lukaszewicz, J. Perkal, and S. Zubrzycki. Sur la liaison et la division des points d'un ensemble fini. *Colloquium Mathematicae*, 2:282–285, 1951.

[27] T. Gärtner, P.A. Flach, and S. Wrobel. On graph kernels: Hardness results and efficient alternatives. In *Proceedings of the 16th Annual Conference on Computational Learning Theory and the 7th Kernel Workshop*, 2003.

[28] S. Geisser. *Predictive Inference*. Chapman and Hall, New York, 1993.

[29] M. Girvan and M. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[30] C. Goutte, L.K. Hansen, M.G. Liptrot, and E. Rostrup. Feature-space clustering for fMRI meta-analysis. *Human Brain Mapping*, 13(3):165–183, 2001.

[31] E.-H. Han, G. Karypis, and V. Kumar. Text categorization using weight adjusted $k$-nearest neighbor classification. In *Proceedings of the 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2001.

[32] J. Han, M. Kamber, and A. Tung. Spatial clustering methods in data mining: A survey. In *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis, 2001.

[33] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pages 1–12, Dallas, TX, 2000.

[34] D. Heckerman. Bayesian networks for data mining. *Data Mining and Knowledge Discovery*, 1:79–119, 1097.

[35] J. Heller. *Catch-22*. Vintage Books, 1994. First published in 1962 by Jonathan Cape Ltd.

[36] R.C. Holte. Very simple classification rules perform well on most commonly used data sets. *Machine Learning*, 11:63–91, 1993.

[37] M Honarkhah and J Caers. Stochastic simulation of patterns using distance-based pattern modeling. *Mathematical Geosciences*, 42:487–517, 2010.

[38] T. Horváth, T. Gärtner, and S. Wrobel. Cyclic pattern kernels for predictive graph mining. In *Proceedings of the 10th ACM SIGKDD Internatonal Conference on Knowledge Discovery and Data Mining*, pages 158–167, 2004.

[39] W.M. Huang and R.P. Lippman. Neural net and traditional classifiers. In *Neural Information Processing Systems*, pages 387–396, 1988.

[40] A. Inokuchi, T. Washio, and H. Motoda. An apriori-based algorithm for mining frequent substructures from graph data. In *Proceedings of the 4th European Conference of Principles and Practice of Knowledge Discovery in Databases*, 2000.

[41] S.C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3), 1967.

[42] L.P. Kaelbling, M.L. Littman, and A.W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

[43] D.R. Karger. Global min-cuts in RNC, and other ramifications of a simple min-cut algorithm. In *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 21–30, 1993.

[44] G.V. Kass. An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29:119–127, 1980.

[45] L. Kaufman and P. J. Rousseeuw. Clustering by means of medoids. In *Statistical Data Analysis Based on the $L_1$-Norm and Related Methods*, pages 405–416. Elsevier Science, 1987.

[46] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data, An Introduction to Cluster Analysis*. John Wiley and Sons Ltd., New York, 1990.

[47] M. Kirsten and S. Wrobel. Relational distance-based clustering. In *Proceedings of the Eighth International Conference on Inductive Logic Programming*, pages 261–270, Berlin, 1998. Springer.

[48] M. Kirsten and S. Wrobel. Extending $k$-means clustering to first-order representations. In *Proceedings of the Tenth International Conference on Inductive Logic Programming*, pages 112–129, Berlin, 2000. Springer.

[49] I.R. Kondor and J.D. Lafferty. Diffusion kernels on graphs and other discrete structures. In *Proceedings of the Nineteenth International Conference on Machine Learning*, pages 315–322. Morgan Kaufmann Publishers Inc., 2002.

[50] S. Kramer. Structural regression trees. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, Cambridge, MA, 1996. AAAI Press/MIT Press.

[51] S. Kramer. *Relational Learning vs. Propositionalization: Investigations in Inductive Logic Programming and Propositional Machine Learning*. PhD thesis, Vienna University of Technology, 1999.

[52] M. Kryszkiewicz and P. Lasek. TI-DBSCAN: Clustering with DBSCAN by means of the triangle inequality. In *Proceedings of the Seventh International Conference on Rough Sets and Current Trends in Computing*, pages 60–69, 2010.

[53] M. Kuramochi and G. Karypis. Frequent subgraph discovery. In *Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM)*, pages 313–320, San Jose, CA, 2001.

[54] P. Langley, W. Iba, and K. Thompson. An analyses of Bayesian classifiers. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 223–228, San Jose, CA, 1992.

[55] A.N. Langville and C.D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, 2006.

[56] B. Larsen and C. Aone. Fast and effective text mining using linear-time document clustering. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 16–22, 1999.

[57] H. Lei, L.R. Tang, J.R. Iglesias, S. Mukherjee, and S. Mohanty. S-means: Similarity driven clustering and its application in gravitational-wave astronomy data mining. In *Proceedings of the International Workshop on Knowledge Discovery from Ubiquitous Data Streams (IWKDUDS 2007)*. Warsaw, Poland, 2007.

[58] D.D. Lewis. Naive (Bayes) at forty: The independance assumption in information retrieval. In *Proceedings of the Tenth European Conference on Machine Learning: ECML-98*, pages 4–15, 1998.

[59] Y. Lin, X. Shi, and Y. Wei. On computing pagerank via lumping the Google matrix. *Journal of Computational and Applied Mathematics*, 224:702–708, 2009.

[60] P. Lyman and H.R. Varian. How much information, 2003. Available at http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/.

[61] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297. University of California Press, 1967.

[62] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press., 2008.

[63] K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Academic Press, 1979.

[64] A. McCallum and K. Nigam. A comparison of invent models for naive Bayes text classification. In *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization*, 1998.

[65] A. McCallum, K. Nigam, J. Rennie, and K. Seymore. Automating the construction of internet portals with machine learning. *Information Retrieval Journal*, 3:127–163, 2000.

[66] L.L. McQuitty. Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. *Educational and Psychological Measurement*, 17:207–229, 1957.

[67] G.W. Milligan. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 45:325–342, 1980.

[68] A.E. Monge and C. Elkan. The field matching problem: Algorithms and applications. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 267–270, 1996.

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

[70] S. Muggleton. Inverse entailment and Progol. *New Generation Computing*, 13:245–286, 1995.

[71] S. Muggleton and W. Buntine. Machine invention of first-order predicates by inverting resolution. In *Proceedings of the Fifth International Conference on Machine Learning*, pages 339–352, San Mateo, CA, 1988. Morgan Kaufman.

[72] J. Neville, M. Adler, and D. Jensen. Clustering relational data using attribute and link information. In *Proceedings of the Text Mining and Link Analysis Workshop, Eighteenth International Joint Conference on Artificial Intelligence*, 2003.

[73] S. Nijssen, Y. Chi, R. Muntz, and J.N. Kok. Frequent tree mining — An overview. *Fundamenta Informaticae*, 66:161–198, 2005.

[74] J.S. Park, M.-S. Chen, and P.S. Yu. An effective hash-based algorithm for mining association rules. In *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data*, pages 175–186, San Jose, CA, 1995.

[75] J.R. Quinlan. Discovering rules by induction from large collection of examples. In *Expert Systems in the Micro Electronic Age*. Edinburgh University Press, 1979.

[76] J.R. Quinlan. Learning logical definitions from relations. *Machine Learning*, 5(3):239–266, 1990.

[77] J.R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan-Kaufman Publishers, San Mateo, CA, 1993.

[78] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77 (2), pages 257–286, 1989.

[79] L. De Raedt and M. Bruynooghe. A theory of clausal discovery. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 1058–1063, San Mateo, CA, 1993. Morgan Kaufmann.

[80] L. De Raedt and W. Van Laer. Inductive constraint logic. In *Proceedings of the Sixth International Workshop on Algorithmic Learning Theory*, pages 80–94, Berlin, 1995. Springer.

[81] C.R. Rao. *Advanced Statistical Data Analysis of Multivariate Observations*. John Wiley and Sons Ltd., 1952.

[82] S. Reeves and M. Clarke. *Logic for Computer Science*. Addison-Wesley, 2003.

[83] S.J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall International, Inc., Englewood Cliffs, NJ, third edition, 2010.

[84] J. Shawe-Taylor and N. Cristianini. *Kernel MEthods for Pattern Analysis*. Cambridge University Press, New York, 2004.

[85] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[86] P.H.A. Sneath. The application of computers to taxonomy. *Journal of General Microbiology*, 17:201–226, 1957.

[87] P.H.A. Sneath and R.R. Sokal. *Numerical Taxonomy — The Principles and Practice of Numerical Classification*. W.H. Freeman, San Francisco, CA, 1973.

[88] R. R. Sokal and C. D. Michener. A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*, 28:1409–1438, 1958.

[89] T. Sørensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biologiske Skrifter*, 5(4):1–34, 1948.

[90] A. Srinivasan. The Aleph manual. Technical report, Computing Laboratory, Oxford University, 2000. Available at `http://www.comlab.ox.ac.uk/activities/machinelearning/Aleph/`.

[91] B. Stein and O. Niggemann. On the nature of structure and its identification. In *25th Workshop on Graph Theory, Lecture Notes in Computer Science*. Springer-Verlag, 1999.

[92] H. Steinhaus. Sur la division des corps matériels en parties. *Bull. Acad. Polon. Sci.*, 4(12):801–804, 1957.

[93] D. Sullivan. comScore: US has most searches; China slowest growth; Google tops worldwide in 2009. Technical report, Search Engine Land, 2009. Available at `http://searchengineland.com/...`
`...comscore-us-most-searches-china-slowest-34217`.

[94] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Pearson Education Inc., Boston, MA, 2006.

[95] R.L. Thorndike. Who belong in the family? *Psychometrika*, 18(4), 1953.

[96] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, New York, 1995.

[97] V. Vapnik. *Statistical Learning Theory*. John Wiley and Sons Ltd., New York, 1998.

[98] J.H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58:236–244, 1963.

[99] T. Witsenburg and H. Blockeel. A method to extend existing document clustering procedures in order to include relational information. In *Proceedings of the Sixth International Workshop on Mining and Learning with Graphs*, Helsinki, Finland, 2008.

[100] D.H. Wolpert and W.G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, 1997.

[101] X. Yan and J. Han. gSpan: Graph-based substructure pattern mining. In *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM)*, pages 721–724, Maebashi City, Japan, 2002.

[102] Y. Zhou, H. Cheng, and J.X. Yu. Graph clustering based on structural/attribute similarities. In *Proceedings of the VLDB Endowment*, pages 718–729, Lyon, France, 2009.