References

Abudawood, T. (**2011**). Multi-class subgroup discovery: Heuristics, algorithms and predictiveness. Ph.D. thesis, University of Bristol, Department of Computer Science, Faculty of Engineering. 357

Abudawood, T. and Flach, P.A. (**2009**). Evaluation measures for multi-class subgroup discovery. In W.L. Buntine, M. Grobelnik, D. Mladenić and J. Shawe-Taylor (eds.), *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD 2009), Part I*, *LNCS*, volume 5781, pp. 35–50. Springer. 193

Agrawal, R., Imielinski, T. and Swami, A.N. (**1993**). Mining association rules between sets of items in large databases. In P. Buneman and S. Jajodia (eds.), *Proceedings of the ACM International Conference on Management of Data (SIGMOD 1993)*, pp. 207–216. ACM Press. 103

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. and Verkamo, A.I. (**1996**). Fast discovery of association rules. In *Advances in Knowledge Discovery and Data Mining*, pp. 307–328. AAAI/MIT Press. 193

Allwein, E.L., Schapire, R.E. and Singer, Y. (**2000**). Reducing multiclass to binary: A unifying approach for margin classifiers. In P. Langley (ed.), *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)*, pp. 9–16. Morgan Kaufmann. 102

Amit, **Y. and Geman**, **D.** (**1997**). Shape quantization and recognition with randomized trees. *Neural Computation* 9(7):1545–1588. 341

Angluin, **D.**, **Frazier**, **M. and Pitt**, **L.** (**1992**). Learning conjunctions of Horn clauses. *Machine Learning* 9:147–164. 128

Bakir, **G.**, **Hofmann**, **T.**, **Schölkopf**, **B.**, **Smola**, **A.J.**, **Taskar**, **B. and Vishwanathan**, **S.V.N.** (**2007**). *Predicting Structured Data*. MIT Press. 361

Banerji, **R.B.** (**1980**). *Artificial Intelligence: A Theoretical Approach*. Elsevier Science. 127

Bengio, **Y.** (**2009**). Learning deep architectures for AI. *Foundations and Trends in Machine Learning* 2(1):1–127. 361

Best, **M.J. and Chakravarti**, **N.** (**1990**). Active set algorithms for isotonic regression; a unifying framework. *Mathematical Programming* 47(1):425–439. 80, 229

Blockeel, **H.** (**2010***a*). Hypothesis language. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 507–511. Springer. 127

Blockeel, **H.** (**2010***b*). Hypothesis space. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 511–513. Springer. 127

Blockeel, **H.**, **De Raedt**, **L. and Ramon**, **J.** (**1998**). Top-down induction of clustering trees. In J.W. Shavlik (ed.), *Proceedings of the Fifteenth International Conference on Machine Learning (ICML 1998)*, pp. 55–63. Morgan Kaufmann. 103, 156

Blumer, **A.**, **Ehrenfeucht**, **A.**, **Haussler**, **D. and Warmuth**, **M.K.** (**1989**). Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM* 36(4):929–965. 128

Boser, **B.E.**, **Guyon**, **I. and Vapnik**, **V.** (**1992**). A training algorithm for optimal margin classifiers. In *Proceedings of the International Conference on Computational Learning Theory (COLT 1992)*, pp. 144–152. 229

Bouckaert, **R. and Frank**, **E.** (**2004**). Evaluating the replicability of significance tests for comparing learning algorithms. In H. Dai, R. Srikant and C. Zhang (eds.), *Advances in Knowledge Discovery and Data Mining, LNCS*, volume 3056, pp. 3–12. Springer. 358

Boullé, **M.** (**2004**). Khiops: A statistical discretization method of continuous attributes. *Machine Learning* 55(1):53–69. 328

Boullé, **M.** (**2006**). MODL: A Bayes optimal discretization method for continuous attributes. *Machine Learning* 65(1):131–165. 328

Bourke, C., Deng, K., Scott, S.D., Schapire, R.E. and Vinodchandran, N.V. (**2008**). On reoptimizing multi-class classifiers. *Machine Learning* 71(2-3):219–242. 102

Brazdil, P., Giraud-Carrier, C.G., Soares, C. and Vilalta, R. (**2009**). *Metalearning – Applications to Data Mining*. Springer. 342

Brazdil, P., Vilalta, R., Giraud-Carrier, C.G. and Soares, C. (**2010**). Metalearning. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 662–666. Springer. 342

Breiman, L. (**1996***a*). Bagging predictors. *Machine Learning* 24(2):123–140. 341

Breiman, L. (**1996***b*). Stacked regressions. *Machine Learning* 24(1):49–64. 342

Breiman, L. (**2001**). Random forests. *Machine Learning* 45(1):5–32. 341

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (**1984**). *Classification and Regression Trees*. Wadsworth. 156

Brier, G.W. (**1950**). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78(1):1–3. 80

Brown, G. (**2010**). Ensemble learning. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 312–320. Springer. 341

Bruner, J.S., Goodnow, J.J. and Austin, G.A. (**1956**). *A Study of Thinking*. Science Editions. 2nd edn 1986. 127

Cesa-Bianchi, N. and Lugosi, G. (**2006**). *Prediction, Learning, and Games*. Cambridge University Press. 361

Cestnik, B. (**1990**). Estimating probabilities: A crucial task in machine learning. In *Proceedings of the European Conference on Artificial Intelligence (ECAI 1990)*, pp. 147–149. 296

Clark, P. and Boswell, R. (**1991**). Rule induction with CN2: Some recent improvements. In Y. Kodratoff (ed.), *Proceedings of the European Working Session on Learning (EWSL 1991), LNCS*, volume 482, pp. 151–163. Springer. 192

Clark, P. and Niblett, T. (**1989**). The CN2 induction algorithm. *Machine Learning* 3:261–283. 192

Cohen, W.W. (**1995**). Fast effective rule induction. In A. Prieditis and S.J. Russell (eds.), *Proceedings of the Twelfth International Conference on Machine Learning (ICML 1995)*, pp. 115–123. Morgan Kaufmann. 192, 341

Cohen, **W.W. and Singer**, **Y.** (**1999**). A simple, fast, and effictive rule learner. In J. Hendler and D. Subramanian (eds.), *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI 1999)*, pp. 335–342. AAAI Press / MIT Press. 341

Cohn, **D.** (**2010**). Active learning. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 10–14. Springer. 128

Cortes, **C. and Vapnik**, **V.** (**1995**). Support-vector networks. *Machine Learning* 20(3):273–297. 229

Cover, **T. and Hart**, **P.** (**1967**). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13(1):21–27. 260

Cristianini, **N. and Shawe-Taylor**, **J.** (**2000**). *An Introduction to Support Vector Machines*. Cambridge University Press. 229

Dasgupta, **S.** (**2010**). Active learning theory. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 14–19. Springer. 128

Davis, **J. and Goadrich**, **M.** (**2006**). The relationship between precision-recall and ROC curves. In W.W. Cohen and A. Moore (eds.), *Proceedings of the Twenty-Third International Conference on Machine Learning (ICML 2006)*, pp. 233–240. ACM Press. 358

De Raedt, **L.** (**1997**). Logical settings for concept-learning. *Artificial Intelligence* 95(1):187–201. 128

De Raedt, **L.** (**2008**). *Logical and Relational Learning*. Springer. 193

De Raedt, **L.** (**2010**). Logic of generality. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 624–631. Springer. 128

De Raedt, **L. and Kersting**, **K.** (**2010**). Statistical relational learning. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 916–924. Springer. 193

Dempster, **A.P.**, **Laird**, **N.M. and Rubin**, **D.B.** (**1977**). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)* pp. 1–38. 296

Demšar, **J.** (**2008**). On the appropriateness of statistical tests in machine learning. In *Proceedings of the ICML'08 Workshop on Evaluation Methods for Machine Learning*. 359

Demšar, **J.** (**2006**). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* 7:1–30. 359

Dietterich, **T.G.** (**1998**). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation* 10(7):1895–1923. 358

Dietterich, **T.G. and Bakiri**, **G.** (**1995**). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research* 2:263–286. 102

Dietterich, **T.G.**, **Kearns**, **M.J. and Mansour**, **Y.** (**1996**). Applying the weak learning framework to understand and improve c4.5. In *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 96–104. 156

Ding, **C.H.Q. and He**, **X.** (**2004**). *K*-means clustering via principal component analysis. In C.E. Brodley (ed.), *Proceedings of the Twenty-First International Conference on Machine Learning (ICML 2004)*. ACM Press. 329

Domingos, **P. and Pazzani**, **M.** (**1997**). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning* 29(2):103–130. 296

Donoho, **S.K. and Rendell**, **L.A.** (**1995**). Rerepresenting and restructuring domain theories: A constructive induction approach. *Journal of Artificial Intelligence Research* 2:411–446. 328

Drummond, **C.** (**2006**). Machine learning as an experimental science (revisited). In *Proceedings of the AAAI'06 Workshop on Evaluation Methods for Machine Learning*. 359

Drummond, **C. and Holte**, **R.C.** (**2000**). Exploiting the cost (in)sensitivity of decision tree splitting criteria. In P. Langley (ed.), *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)*, pp. 239–246. Morgan Kaufmann. 156

Egan, **J.P.** (**1975**). *Signal Detection Theory and ROC Analysis*. Academic Press. 80

Fawcett, **T.** (**2006**). An introduction to ROC analysis. *Pattern Recognition Letters* 27(8):861–874. 80, 358

Fawcett, **T. and Niculescu-Mizil**, **A.** (**2007**). PAV and the ROC convex hull. *Machine Learning* 68(1):97–106. 80, 229

Fayyad, **U.M. and Irani**, **K.B.** (**1993**). Multi-interval discretization of continuous-valued attributes for classification learning. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 1993)*, pp. 1022–1029. 328

Ferri, **C.**, **Flach**, **P.A. and Hernández-Orallo**, **J.** (**2002**). Learning decision trees using the area under the ROC curve. In C. Sammut and A.G. Hoffmann (eds.), *Proceedings of the Nineteenth International Conference on Machine Learning (ICML 2002)*, pp. 139–146. Morgan Kaufmann. 156

Ferri, **C.**, **Flach**, **P.A. and Hernández-Orallo**, **J.** (**2003**). Improving the AUC of probabilistic estimation trees. In N. Lavrač, D. Gamberger, L. Todorovski and H. Blockeel (eds.), *Proceedings of the European Conference on Machine Learning (ECML 2003)*, *LNCS*, volume 2837, pp. 121–132. Springer. 156

Fix, **E. and Hodges**, **J.L.** (**1951**). Discriminatory analysis. Nonparametric discrimination: Consistency properties. Technical report, USAF School of Aviation Medicine, Texas: Randolph Field. Report Number 4, Project Number 21-49-004. 260

Flach, **P.A.** (**1994**). *Simply Logical – Intelligent Reasoning by Example*. Wiley. 193

Flach, **P.A.** (**2003**). The geometry of ROC space: Understanding machine learning metrics through ROC isometrics. In T. Fawcett and N. Mishra (eds.), *Proceedings of the Twentieth International Conference on Machine Learning (ICML 2003)*, pp. 194–201. AAAI Press. 156

Flach, **P.A.** (**2010***a*). First-order logic. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 410–415. Springer. 128

Flach, **P.A.** (**2010***b*). ROC analysis. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 869–875. Springer. 80

Flach, **P.A. and Lachiche**, **N.** (**2001**). Confirmation-guided discovery of first-order rules with Tertius. *Machine Learning* 42(1/2):61–95. 193

Flach, **P.A. and Matsubara**, **E.T.** (**2007**). A simple lexicographic ranker and probability estimator. In J.N. Kok, J. Koronacki, R.L. de Mántaras, S. Matwin, D. Mladenic and A. Skowron (eds.), *Proceedings of the Eighteenth European Conference on Machine Learning (ECML 2007)*, *LNCS*, volume 4701, pp. 575–582. Springer. 80, 229

Freund, **Y.**, **Iyer**, **R.D.**, **Schapire**, **R.E. and Singer**, **Y.** (**2003**). An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research* 4:933–969. 341

Freund, **Y. and Schapire**, **R.E.** (**1997**). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* 55(1):119–139. 341

Fürnkranz, **J.** (**1999**). Separate-and-conquer rule learning. *Artificial Intelligence Review* 13(1):3–54. 192

Fürnkranz, **J.** (**2010**). Rule learning. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 875–879. Springer. 192

Fürnkranz, **J. and Flach**, **P.A.** (**2003**). An analysis of rule evaluation metrics. In T. Fawcett and N. Mishra (eds.), *Proceedings of the Twentieth International Conference on Machine Learning (ICML 2003)*, pp. 202–209. AAAI Press. 79

https://doi.org/10.1017/CBO9780511973000.017 Published online by Cambridge University Press

Fürnkranz, J. and Flach, P.A. (**2005**). ROC 'n' Rule learning – towards a better understanding of covering algorithms. *Machine Learning* 58(1):39–77. 192

Fürnkranz, J., Gamberger, D. and Lavrač, N. (**2012**). *Foundations of Rule Learning*. Springer. 192

Fürnkranz, J. and Hüllermeier, E. (eds.) (**2010**). *Preference Learning*. Springer. 361

Fürnkranz, J. and Widmer, G. (**1994**). Incremental reduced error pruning. In *Proceedings of the Eleventh International Conference on Machine Learning (ICML 1994)*, pp. 70–77. 192

Gama, J. and Gaber, M.M. (eds.) (**2007**). *Learning from Data Streams: Processing Techniques in Sensor Networks*. Springer. 361

Ganter, B. and Wille, R. (**1999**). *Formal Concept Analysis: Mathematical Foundations*. Springer. 127

Garriga, G.C., Kralj, P. and Lavrač, N. (**2008**). Closed sets for labeled data. *Journal of Machine Learning Research* 9:559–580. 127

Gärtner, T. (**2009**). *Kernels for Structured Data*. World Scientific. 230

Grünwald, P.D. (**2007**). *The Minimum Description Length Principle*. MIT Press. 297

Guyon, I. and Elisseeff, A. (**2003**). An introduction to variable and feature selection. *Journal of Machine Learning Research* 3:1157–1182. 328

Hall, M.A. (**1999**). Correlation-based feature selection for machine learning. Ph.D. thesis, University of Waikato. 328

Han, J., Cheng, H., Xin, D. and Yan, X. (**2007**). Frequent pattern mining: Current status and future directions. *Data Mining and Knowledge Discovery* 15(1):55–86. 193

Hand, D.J. and Till, R.J. (**2001**). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning* 45(2):171–186. 102

Haussler, D. (**1988**). Quantifying inductive bias: AI learning algorithms and Valiant's learning framework. *Artificial Intelligence* 36(2):177–221. 128

Hernández-Orallo, J., Flach, P.A. and Ferri, C. (**2011**). Threshold choice methods: The missing link. Available online at `http://arxiv.org/abs/1112.2640`. 358

Ho, T.K. (**1995**). Random decision forests. In *Proceedings of the International Conference on Document Analysis and Recognition*, p. 278. IEEE Computer Society, Los Alamitos, CA, USA. 341

Hoerl, **A.E. and Kennard**, **R.W.** (**1970**). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* pp. 55–67. 228

Hofmann, **T.** (**1999**). Probabilistic latent semantic indexing. In *Proceedings of the Twenty-Second Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR 1999)*, pp. 50–57. ACM Press. 329

Hunt, **E.B.**, **Marin**, **J. and Stone**, **P.J.** (**1966**). *Experiments in Induction*. Academic Press. 127, 156

Jain, **A.K.**, **Murty**, **M.N. and Flynn**, **P.J.** (**1999**). Data clustering: A review. *ACM Computing Surveys* 31(3):264–323. 261

Japkowicz, **N. and Shah**, **M.** (**2011**). *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press. 357

Jebara, **T.** (**2004**). *Machine Learning: Discriminative and Generative*. Springer. 296

John, **G.H. and Langley**, **P.** (**1995**). Estimating continuous distributions in Bayesian classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI 1995)*, pp. 338–345. Morgan Kaufmann. 295

Kaufman, **L. and Rousseeuw**, **P.J.** (**1990**). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley. 261

Kearns, **M.J. and Valiant**, **L.G.** (**1989**). Cryptographic limitations on learning Boolean formulae and finite automata. In D.S. Johnson (ed.), *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing (STOC 1989)*, pp. 433–444. ACM Press. 341

Kearns, **M.J. and Valiant**, **L.G.** (**1994**). Cryptographic limitations on learning Boolean formulae and finite automata. *Journal of the ACM* 41(1):67–95. 341

Kerber, **R.** (**1992**). Chimerge: Discretization of numeric attributes. In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI 1992)*, pp. 123–128. AAAI Press. 328

Kibler, **D.F. and Langley**, **P.** (**1988**). Machine learning as an experimental science. In *Proceedings of the European Working Session on Learning (EWSL 1988)*, pp. 81–92. 359

King, **R.D.**, **Srinivasan**, **A. and Dehaspe**, **L.** (**2001**). Warmr: A data mining tool for chemical data. *Journal of Computer-Aided Molecular Design* 15(2):173–181. 193

Kira, **K. and Rendell**, **L.A.** (**1992**). The feature selection problem: Traditional methods and a new algorithm. In W.R. Swartout (ed.), *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI 1992)*, pp. 129–134. AAAI Press / MIT Press. 328

Klösgen, **W.** (**1996**). Explora: A multipattern and multistrategy discovery assistant. In *Advances in Knowledge Discovery and Data Mining*, pp. 249–271. MIT Press. 103

Kohavi, **R. and John**, **G.H.** (**1997**). Wrappers for feature subset selection. *Artificial Intelligence* 97(1-2):273–324. 328

Koren, **Y.**, **Bell**, **R. and Volinsky**, **C.** (**2009**). Matrix factorization techniques for recommender systems. *IEEE Computer* 42(8):30–37. 328

Kramer, **S.** (**1996**). Structural regression trees. In *Proceedings of the National Conference on Artificial Intelligence (AAAI 1996)*, pp. 812–819. 156

Kramer, **S.**, **Lavrač**, **N. and Flach**, **P.A.** (**2000**). Propositionalization approaches to relational data mining. In S. Džeroski and N. Lavrač (eds.), *Relational Data Mining*, pp. 262–286. Springer. 328

Krogel, **M.A.**, **Rawles**, **S.**, **Zelezný**, **F.**, **Flach**, **P.A.**, **Lavrač**, **N. and Wrobel**, **S.** (**2003**). Comparative evaluation of approaches to propositionalization. In T. Horváth (ed.), *Proceedings of the Thirteenth International Conference on Inductive Logic Programming (ILP 2003)*, *LNCS*, volume 2835, pp. 197–214. Springer. 328

Kuncheva, **L.I.** (**2004**). *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley and Sons. 341

Lachiche, **N.** (**2010**). Propositionalization. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 812–817. Springer. 328

Lachiche, **N. and Flach**, **P.A.** (**2003**). Improving accuracy and cost of two-class and multi-class probabilistic classifiers using ROC curves. In T. Fawcett and N. Mishra (eds.), *Proceedings of the Twentieth International Conference on Machine Learning (ICML 2003)*, pp. 416–423. AAAI Press. 102

Lafferty, **J.D.**, **McCallum**, **A. and Pereira**, **F.C.N.** (**2001**). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In C.E. Brodley and A.P. Danyluk (eds.), *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001)*, pp. 282–289. Morgan Kaufmann. 296

Langley, **P.** (**1988**). Machine learning as an experimental science. *Machine Learning* 3:5–8. 359

Langley, **P.** (**1994**). *Elements of Machine Learning*. Morgan Kaufmann. 156

Langley, **P.** (**2011**). The changing science of machine learning. *Machine Learning* 82(3):275–279. 359

Lavrač, **N.**, **Kavšek**, **B.**, **Flach**, **P.A. and Todorovski**, **L.** (**2004**). Subgroup discovery with CN2-SD. *Journal of Machine Learning Research* 5:153–188. 193

Lee, **D.D.**, **Seung**, **H.S.** *et al.* (**1999**). Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755):788–791. 328

Leman, **D.**, **Feelders**, **A. and Knobbe**, **A.J.** (**2008**). Exceptional model mining. In W. Daelemans, B. Goethals and K. Morik (eds.), *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD 2008), Part II*, *LNCS*, volume 5212, pp. 1–16. Springer. 103

Lewis, **D.** (**1998**). Naive Bayes at forty: The independence assumption in information retrieval. In *Proceedings of the Tenth European Conference on Machine Learning (ECML 1998)*, pp. 4–15. Springer. 295

Li, **W.**, **Han**, **J. and Pei**, **J.** (**2001**). CMAR: Accurate and efficient classification based on multiple class-association rules. In N. Cercone, T.Y. Lin and X. Wu (eds.), *Proceedings of the IEEE International Conference on Data Mining (ICDM 2001)*, pp. 369–376. IEEE Computer Society. 193

Little, **R.J.A. and Rubin**, **D.B.** (**1987**). *Statistical Analysis with Missing Data*. Wiley. 296

Liu, **B.**, **Hsu**, **W. and Ma**, **Y.** (**1998**). Integrating classification and association rule mining. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD 1998)*, pp. 80–86. AAAI Press. 193

Lloyd, **J.W.** (**2003**). *Logic for Learning – Learning Comprehensible Theories from Structured Data*. Springer. 193

Lloyd, **S.** (**1982**). Least squares quantization in PCM. *IEEE Transactions on Information Theory* 28(2):129–137. 261

Mahalanobis, **P.C.** (**1936**). On the generalised distance in statistics. *Proceedings of the National Institute of Science, India* 2(1):49–55. 260

Mahoney, **M.W. and Drineas**, **P.** (**2009**). CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences* 106(3):697. 329

McCallum, **A. and Nigam**, **K.** (**1998**). A comparison of event models for naive Bayes text classification. In *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization*, pp. 41–48. 295

Michalski, **R.S.** (**1973**). Discovering classification rules using variable-valued logic system VL_1. In *Proceedings of the Third International Joint Conference on Artificial Intelligence*, pp. 162–172. Morgan Kaufmann Publishers. 127

Michalski, **R.S.** (**1975**). Synthesis of optimal and quasi-optimal variable-valued logic formulas. In *Proceedings of the 1975 International Symposium on Multiple-Valued Logic*, pp. 76–87. 192

Michie, **D.**, **Spiegelhalter**, **D.J. and Taylor**, **C.C.** (**1994**). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood. 342

Miettinen, **P.** (**2009**). Matrix decomposition methods for data mining: Computational complexity and algorithms. Ph.D. thesis, University of Helsinki. 329

Minsky, **M. and Papert**, **S.** (**1969**). *Perceptrons: An Introduction to Computational Geometry*. MIT Press. 228

Mitchell, **T.M.** (**1977**). Version spaces: A candidate elimination approach to rule learning. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pp. 305–310. Morgan Kaufmann Publishers. 127

Mitchell, **T.M.** (**1997**). *Machine Learning*. McGraw-Hill. 128

Muggleton, **S.** (**1995**). Inverse entailment and Progol. *New Generation Computing* 13(3&4):245–286. 193

Muggleton, **S.**, **De Raedt**, **L.**, **Poole**, **D.**, **Bratko**, **I.**, **Flach**, **P.A.**, **Inoue**, **K. and Srinivasan**, **A.** (**2012**). ILP turns 20 – biography and future challenges. *Machine Learning* 86(1):3–23. 193

Muggleton, **S. and Feng**, **C.** (**1990**). Efficient induction of logic programs. In *Proceedings of the International Conference on Algorithmic Learning Theory (ALT 1990)*, pp. 368–381. 193

Murphy, **A.H. and Winkler**, **R.L.** (**1984**). Probability forecasting in meteorology. *Journal of the American Statistical Association* pp. 489–500. 80

Nelder, **J.A. and Wedderburn**, **R.W.M.** (**1972**). Generalized linear models. *Journal of the Royal Statistical Society, Series A (General)* pp. 370–384. 296

Novikoff, **A.B.** (**1962**). On convergence proofs on perceptrons. In *Proceedings of the Symposium on the Mathematical Theory of Automata*, volume 12, pp. 615–622. Polytechnic Institute of Brooklyn, New York. 228

Pasquier, **N.**, **Bastide**, **Y.**, **Taouil**, **R. and Lakhal**, **L.** (**1999**). Discovering frequent closed itemsets for association rules. In *Proceedings of the International Conference on Database Theory (ICDT 1999)*, pp. 398–416. Springer. 127

Peng, **Y.**, **Flach**, **P.A.**, **Soares**, **C. and Brazdil**, **P.** (**2002**). Improved dataset characterisation for meta-learning. In S. Lange, K. Satoh and C.H. Smith (eds.), *Proceedings of the Fifth International Conference on Discovery Science (DS 2002)*, *LNCS*, volume 2534, pp. 141–152. Springer. 342

Pfahringer, **B.**, **Bensusan**, **H. and Giraud-Carrier**, **C.G.** (**2000**). Meta-learning by landmarking various learning algorithms. In P. Langley (ed.), *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)*, pp. 743–750. Morgan Kaufmann. 342

Platt, **J.C.** (**1998**). Using analytic QP and sparseness to speed training of support vector machines. In M.J. Kearns, S.A. Solla and D.A. Cohn (eds.), *Advances in Neural Information Processing Systems 11 (NIPS 1998)*, pp. 557–563. MIT Press. 229

Plotkin, **G.D.** (**1971**). Automatic methods of inductive inference. Ph.D. thesis, University of Edinburgh. 127

Provost, **F.J. and Domingos**, **P.** (**2003**). Tree induction for probability-based ranking. *Machine Learning* 52(3):199–215. 156

Provost, **F.J. and Fawcett**, **T.** (**2001**). Robust classification for imprecise environments. *Machine Learning* 42(3):203–231. 79

Quinlan, **J.R.** (**1986**). Induction of decision trees. *Machine Learning* 1(1):81–106. 155

Quinlan, **J.R.** (**1990**). Learning logical definitions from relations. *Machine Learning* 5:239–266. 193

Quinlan, **J.R.** (**1993**). *C4.5: Programs for Machine Learning*. Morgan Kaufmann. 156

Ragavan, **H. and Rendell**, **L.A.** (**1993**). Lookahead feature construction for learning hard concepts. In *Proceedings of the Tenth International Conference on Machine Learning (ICML 1993)*, pp. 252–259. Morgan Kaufmann. 328

Rajnarayan, **D.G. and Wolpert**, **D.** (**2010**). Bias-variance trade-offs: Novel applications. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 101–110. Springer. 103

Rissanen, **J.** (**1978**). Modeling by shortest data description. *Automatica* 14(5):465–471. 297

Rivest, **R.L.** (**1987**). Learning decision lists. *Machine Learning* 2(3):229–246. 192

Robnik-Sikonja, **M. and Kononenko**, **I.** (**2003**). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning* 53(1-2):23–69. 328

Rosenblatt, **F.** (**1958**). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65(6):386. 228

Rousseeuw, **P.J.** (**1987**). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20(0):53–65. 261

Rumelhart, **D.E.**, **Hinton**, **G.E. and Williams**, **R.J.** (**1986**). Learning representations by back-propagating errors. *Nature* 323(6088):533–536. 229

Schapire, **R.E.** (**1990**). The strength of weak learnability. *Machine Learning* 5:197–227. 341

Schapire, **R.E.** (**2003**). The boosting approach to machine learning: An overview. In *Nonlinear Estimation and Classification*, pp. 149–172. Springer. 341

Schapire, **R.E.**, **Freund**, **Y.**, **Bartlett**, **P. and Lee**, **W.S.** (**1998**). Boosting the margin: A new explanation for the effectiveness of voting methods. *Annals of Statistics* 26(5):1651–1686. 341

Schapire, **R.E. and Singer**, **Y.** (**1999**). Improved boosting algorithms using confidence-rated predictions. *Machine Learning* 37(3):297–336. 341

Settles, **B.** (**2011**). *Active Learning*. Morgan & Claypool. 361

Shawe-Taylor, **J. and Cristianini**, **N.** (**2004**). *Kernel Methods for Pattern Analysis*. Cambridge University Press. 230

Shotton, **J.**, **Fitzgibbon**, **A.W.**, **Cook**, **M.**, **Sharp**, **T.**, **Finocchio**, **M.**, **Moore**, **R.**, **Kipman**, **A. and Blake**, **A.** (**2011**). Real-time human pose recognition in parts from single depth images. In *Proceedings of the Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*, pp. 1297–1304. 155

Silver, **D. and Bennett**, **K.** (**2008**). Guest editor's introduction: special issue on inductive transfer learning. *Machine Learning* 73(3):215–220. 361

Solomonoff, **R.J.** (**1964***a*). A formal theory of inductive inference: Part I. *Information and Control* 7(1):1–22. 297

Solomonoff, **R.J.** (**1964***b*). A formal theory of inductive inference: Part II. *Information and Control* 7(2):224–254. 297

Srinivasan, **A.** (**2007**). The Aleph manual, version 4 and above. Available online at `www.cs.ox.ac.uk/activities/machlearn/Aleph/`. 193

https://doi.org/10.1017/CBO9780511973000.017 Published online by Cambridge University Press

Stevens, **S.S.** (**1946**). On the theory of scales of measurement. *Science* 103(2684):677–680. 327

Sutton, **R.S. and Barto**, **A.G.** (**1998**). *Reinforcement Learning: An Introduction*. MIT Press. 361

Tibshirani, **R.** (**1996**). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (Methodological)* pp. 267–288. 228

Todorovski, **L. and Dzeroski**, **S.** (**2003**). Combining classifiers with meta decision trees. *Machine Learning* 50(3):223–249. 342

Tsoumakas, **G.**, **Zhang**, **M.L. and Zhou**, **Z.H.** (**2012**). Introduction to the special issue on learning from multi-label data. *Machine Learning* 88(1-2):1–4. 361

Tukey, **J.W.** (**1977**). *Exploratory Data Analysis*. Addison-Wesley. 103

Valiant, **L.G.** (**1984**). A theory of the learnable. *Communications of the ACM* 27(11):1134–1142. 128

Vapnik, **V.N. and Chervonenkis**, **A.Y.** (**1971**). On uniform convergence of the frequencies of events to their probabilities. *Teoriya Veroyatnostei I Ee Primeneniya* 16(2):264–279. 128

Vere, **S.A.** (**1975**). Induction of concepts in the predicate calculus. In *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*, pp. 281–287. 127

von Hippel, **P.T.** (**2005**). Mean, median, and skew: Correcting a textbook rule. *Journal of Statistics Education* 13(2). 327

Wallace, **C.S. and Boulton**, **D.M.** (**1968**). An information measure for classification. *Computer Journal* 11(2):185–194. 297

Webb, **G.I.** (**1995**). Opus: An efficient admissible algorithm for unordered search. *Journal of Artificial Intelligence Research* 3:431–465. 192

Webb, **G.I.**, **Boughton**, **J.R. and Wang**, **Z.** (**2005**). Not so naive Bayes: Aggregating one-dependence estimators. *Machine Learning* 58(1):5–24. 295

Winston, **P.H.** (**1970**). Learning structural descriptions from examples. Technical report, MIT Artificial Intelligence Lab. AITR-231. 127

Wojtusiak, **J.**, **Michalski**, **R.S.**, **Kaufman**, **K.A. and Pietrzykowski**, **J.** (**2006**). The AQ21 natural induction program for pattern discovery: Initial version and its novel features. In *Proceedings of the Eighteenth IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2006)*, pp. 523–526. 192

Wolpert, **D.H.** (**1992**). Stacked generalization. *Neural Networks* 5(2):241–259. 342

Zadrozny, **B. and Elkan**, **C.** (**2002**). Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the Eighth ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD 2002)*, pp. 694–699. ACM Press. 80, 229

Zeugmann, **T.** (**2010**). PAC learning. In C. Sammut and G.I. Webb (eds.), *Encyclopedia of Machine Learning*, pp. 745–753. Springer. 128

Zhou, **Z.H.** (**2012**). *Ensemble Methods: Foundations and Algorithms.* Taylor & Francis. 341

https://doi.org/10.1017/CBO9780511973000.017 Published online by Cambridge University Press