Bibliography

1. Ash, T. (1989). Dynamic node creation in backpropagation neural networks. *Connection Science: Journal of Neural Computing, Artificial Intelligence, and Cognitive Research, 1*, 365–375.
2. Ball, G. H. & Hall, D. J. (1965). ISODATA, a novel method of data analysis and clasification. *Technical Report of the Standford University*, Stanford, CA
3. Bellman, R. E. (1956). A problem in the sequential design of experiments. *Sankhya, 16*, 221–229.
4. Bellman, R. E. (1957). *Dynamic programming*. Princeton: Princeton University Press.
5. Blake, C. L. & Merz, C. J. (1998). Repository of machine learning databases. Department of Information and Computer Science, University of California at Irvine. www.ics.uci.edu/~mlearn/MLRepository.html.
6. Blumer, W., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1989). Learnability and the Vapnik-Chervonenkis dimension. *Journals of the ACM, 36*, 929–965.
7. Boutell, M. R., Luo, J., Shen, X., & Brown, C. M. (2004). Learning multi-label scene classification. *Pattern Recognition, 37*, 1757–1771
8. Bower, G. H. & Hilgard, E. R. (1981). *Theories of learning*. Englewood Cliffs: Prentice-Hall.
9. Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*, 123–140.
10. Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.
11. Breiman, L., Friedman, J., Olshen, R., & Stone, C. J. (1984). *Classification and regression trees*. Belmont: Wadsworth International Group.
12. Broomhead, D. S. & Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems, 2*, 321–355.
13. Bryson, A. E. & Ho, Y.-C. (1969). *Applied optimal control*. New York: Blaisdell.
14. Chow, C. K. (1957). An optimum character recognition system using decision functions. *IRE Transactions on Computers, EC-6*, 247–254.
15. Clare, A. & King, R. D. (2001). Knowledge discovery in multi-label phenotype data. In *Proceedings of the 5th European conference on principles of data mining and knowledge discovery, PKDD'01*, Freiburg, Germany (pp. 42–53)
16. Clark, P. & Niblett, R. (1989). The CN2 induction algorithm. *Machine Learning, 3*, 261–284.
17. Coppin, B. (2004). *Artificial intelligence illuminated*. Sudbury: Jones and Bartlett.
18. Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers, EC-14*, 326–334.
19. Cover, T. M. (1968). Estimation by the nearest neighbor rule. *IEEE Transactions on Information Theory, IT-14*, 50–55.

M. Kubat, *An Introduction to Machine Learning*,
DOI 10.1007/978-3-319-63913-0

20. Cover, T. M. & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, IT-13*, 21–27.
21. Dasarathy, B. V. (1991). *Nearest-neighbor classification techniques*. Los Alomitos: IEEE Computer Society Press.
22. Dietterich, T. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation, 10*, 1895–1923.
23. Dudani, S. A. (1975). The distance-weighted *k*-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-6*, 325–327.
24. Fayyad, U. M. & Irani, K. B. (1992). On the handling of continuous-valued attributes in decision tree generation. *Machine Learning, 8*, 87–102.
25. Fisher, R. A. (1936). The use of multiple measurement in taxonomic problems. *Annals of Eugenics, 7*, 111–132.
26. Fisher, D. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning, 2*, 139–172.
27. Fix, E. & Hodges, J. L. (1951). Discriminatory analysis, non-parametric discrimination. USAF School of Aviation Medicine, Randolph Field, TX, Project 21-49-004, Report 4, Contract AF41(128)-3
28. Freund, Y. & Schapire, R. E. (1996). Experiments with a new boosting algorithm. In *Machine Learning: Proceedings of the Thirteenth International Conference*, Bari (pp. 148–156).
29. Fogel, L. J., Owens, A. J., & Walsh, M. J. (1966). *Artificial intelligence through simulated evolution*. New York: Wiley.
30. Friedman, J. H., Bentley, J. L., & Finkel, R. A. (1977). An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software, 3*(3), 209–226.
31. Gennari, J. H., Langley, P., & Fisher, D. (1990). Models of incremental concept formation. *Artificial Intelligence, 40*, 11–61.
32. Godbole, S. & Sarawagi, S. (2004). Discriminative methods for multi-label classification. In H. Dai, R. Srikant, & C. Zhang (Eds.), *Lecture Notes in Artificial Intelligence* (Vol. 3056, pp. 22–30). Berlin/Heidelberg: Springer.
33. Good, I. J. (1965). *The estimation of probabilities: An essay on modern Bayesian methods*. Cambridge: MIT.
34. Gordon, D. F. & desJardin, M. (1995). Evaluation and selection of biases in machine learning. *Machine Learning, 20*, 5–22.
35. Hart, P. E. (1968). The condensed nearest neighbor rule. *IEEE Transactions on Information Theory, IT-14*, 515–516.
36. Hellman, M. E. (1970). The nearest neighbor classification rule with the reject option. *IEEE Transactions on Systems Science and Cybernetics, 6*, 179–185.
37. Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.
38. Holte, R. C. (1993). Very simple classification rules perform well on most commonly used databases. *Machine Learning, 11*, 63–90.
39. Hunt, E. B., Marin, J., & Stone, P. J. (1966). *Experiments in induction*. New York: Academic Press.
40. Katz, A. J., Gately, M. T., & Collins, D. R. (1990). Robust classifiers without robust features. *Neural Computation, 2*, 472–479.
41. Kearns, M. J. & Vazirani, U. V. (1994). *An introduction to computational learning theory*. Cambridge, MA: MIT Press.
42. Kodratoff, Y. (1988). *Introduction to machine learning*. London: Pitman.
43. Kodratoff, Y. & Michalski, R. S. (1990). *Machine learning: An artificial intelligence approach* (Vol. 3). San Mateo: Morgan Kaufmann.
44. Kohavi, R. (1997). Wrappers for feature selection. *Artificial Intelligence, 97*(1–2), 273–324.
45. Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*, 59–69
46. Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE, 78*(9), 1464–1480.

47. Koller, D. & Sahami, M. (1997). Hierarchically classifying documents using very few words. In *Proceedings of the 14th International conference on machine learning, ICML'07*, San Francisco, USA (pp. 170–178)
48. Kononenko, I., Bratko, I., & Kukar, M. (1998). Application of machine learning to medical diagnosis. In R. Michalski, I. Bratko, & M. Kubat (Eds.), *Machine learning and data mining: Methods and applications*. Chichester: Wiley.
49. Kubat, M. (1989). Floating approximation in time-varying knowledge bases. *Pattern Recognition Letters, 10*, 223–227.
50. Kubat, M., Pfurtscheller, G., & Flotzinger D. (1994). AI-based approach to automatic sleep classification. *Biological Cybernetics, 79*, 443–448.
51. Kubat, M., Holte, R., & Matwin, S. (1997). Learning when negatives examples abound. In *Proceedings of the European conference on machine learning (ECML'97)*, Apr 1997, Prague (pp. 146–153).
52. Kubat, M., Holte, R., & Matwin, S. (1998). Detection of oil-spills in radar images of sea surface. *Machine Learning, 30*, 195–215.
53. Kubat, M., Koprinska, I., & Pfurtscheller, G. (1998). Learning to classify medical signals. In R. Michalski, I. Bratko, & M. Kubat (Eds.), *Machine learning and data mining: Methods and applications*. Chichester: Wiley.
54. Littlestone, N. (1987). Learning quickly when irrelevant attributes abound: A new linear threshold algorithm. *Machine Learning, 2*, 285–318.
55. Lewis, D. D. & Gale, W. A. (1994). A sequential algorithm for training text classifiers. In *Proceedings of the 17th annual international ACM SIGIR conference on research and development in information retrieval (SIGIR'94)*, Dublin (pp. 3–12).
56. Louizou, G. & Maybank, S. J. (1987). The nearest neighbor and the bayes error rates. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 9*, 254–262.
57. McCallum, A. (1999). Multi-label text classification with a mixture model trained by EM. In *Proceedings of the workshop on text learning (AAAI'99)* (pp. 1–7).
58. McQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley symposium on mathematical statistics and probability*, Berkeley (pp. 281–297).
59. Michalski, R. S. (1969). On the quasi-minimal solution of the general covering problem. In *Proceedings of the 5th international symposium on information processing (FCIP'69)*, Bled, Yugoslavia (Vol. A3, pp. 125–128).
60. Michalski, R. S. & Tecuci, G. (1994). *Machine learning: A multistrategy approach*. Palo Alto: Morgan Kaufmann.
61. Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1983). *Machine learning: An artificial intelligence approach*. Palo Alto: Tioga Publishing Company.
62. Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1986). *Machine learning: An artificial intelligence approach* (Vol. 2). Palo Alto: Tioga Publishing Company.
63. Michalski, R., Bratko, I., & Kubat, M. (1998). *Machine learning and data mining: Methods and applications*. New York: Wiley.
64. Michell, M. (1998). *An introduction to genetic algorithm*. Cambridge, MA: MIT.
65. Mill, J. S. (1865). *A system of logic*. London: Longmans.
66. Minsky, M. & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT.
67. Mitchell, T. M. (1982). Generalization as search. *Artificial Intelligence, 18*, 203–226.
68. Mitchell, T. M. (1997). *Machine learning*. New York: McGraw-Hill.
69. Mori, S, Suen, C. Y., & Yamamoto, K. (1992). Historical overview of OCR research and development. *Proceedings of IEEE, 80*, 1029–1058.
70. Muggleton, S. & Buntine, W. (1988). Machine invention of first-order predicates by inverting resolution. In *Proceedings of the 5th international machine learning conference*, Ann Arbor, Michigan (pp. 339–352)
71. Murty, M. N. & Krishna, G. (1980). A computationally efficient technique for data clustering. *Pattern Recognition, 12*, 153–158.

72. Neyman, J. & Pearson E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrica, 20A*, 175–240.
73. Ogden, C. K. & Richards, I. A. (1923). *The meaning of meaning* (8th ed., 1946). New York: Harcourt, Brace, and World.
74. Parzen E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics, 33*, 1065–1076.
75. Quinlan, J. R. (1979). Discovering rules by induction from large collections of examples. In D. Michie (Ed.) *Expert systems in the micro electronic age*. Edinburgh: Edinburgh University Press.
76. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*, 81–106.
77. Quinlan, R. (1990). Learning logical definitions from relations. *Machine Learning, 5*, 239–266
78. Quinlan, J. R. (1993). *C4.5: Programms for machine learning*. San Mateo: Morgan Kaufmann.
79. Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2011). Classifier chains for multi-label classification. *Machine Learning, 85*, 333–359.
80. Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer Systeme nach Principien der biologischen Evolution*. Stuttgart: Frommann-Holzboog.
81. Rosenblatt, M. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*, 386–408.
82. Rozsypal, A. & Kubat, M. (2001). Using the genetic algorithm to reduce the size of a nearest-neighbor classifier and to select relevant attributes. In *Proceedings of the 18th international conference on machine learning*, Williamstown (pp. 449–456).
83. Rumelhart, D. E. & McClelland, J. L. (1986). *Parallel distributed processing*. Cambridge: MIT Bradford Press.
84. Russell, S. & Norvig, P. (2003). *Artificial intelligence, a modern approach* (2nd ed.). Englewood Cliffs: Prentice Hall.
85. Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning, 5*, 197–227.
86. Shawe-Taylor, J., Anthony, M., & Biggs, N. (1993). Bounding sample size with the Vapnik-Chervonenkis dimension. *Discrete Applied Mthematics, 42*(1), 65–73.
87. Sutton, R. S. (1984). *Temporal credit assignment in reinforcement learning*. PhD Dissertation, University of Massachusetts, Amherst.
88. Thrun, S. B. & Mitchell, T. M. (1995). Lifelong robot learning. *Robotics and Automonous Systems, 15*, pp. 24–46.
89. Tomek, I. (1976). Two modifications of CNN. *IEEE Transactions on Systems, Man and Communications, SMC-6*, 769–772.
90. Turney, P. D. (1993). Robust classification with context-sensitive features. In *Proceedings of the sixth international conference of industrial and engineering applications of artificial intelligence and expert systems*, Edinburgh (pp. 268–276).
91. Valiant, L. G. (1984). A theory of the learnable. *Communications of the ACM, 27*, 1134–1142.
92. Vapnik, V. N. (1992). *Estimation of dependences based on empirical data*. New York: Springer.
93. Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer.
94. Vapnik, V. N. & Chervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications, 16*, 264–280.
95. Werbos, P. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. PhD thesis, Harvard University.
96. Whewel, W. (1858). *History of scientific ideas*. London: J.W. Parker.
97. Widmer, G. & Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine Learning, 23*, 69–101.

98. Widrow, B. & Hoff, M. E. (1960). Adaptive switching circuits. In *IRE WESCON convention record*, New York (pp. 96–104).
99. Wolpert, D. (1992). Stacked generalization. *Neural Networks, 5*, 241–259.
100. Wolpert, D. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation, 8*, 1341–1390.
101. Zhang, M.-L. & Zhou, Z.-H. (2007). ML-KNN: A lazy learing approach to multi-label learning. *Pattern Recognition, 40*, 2038–2048.

Index

M. Kubat, *An Introduction to Machine Learning*,
DOI 10.1007/978-3-319-63913-0