

BIBLIOGRAPHY

- [1] S. C. Ahalt, A. K. Krishnamurthy, P. Chen, and D. E. Melton. Competitive learning algorithms for vector quantization. *Neural Networks*, 3:277–290, 1990.
- [2] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
- [3] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.
- [4] D. Arthur and S. Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1027–1035, Philadelphia, 2007. Society for Industrial and Applied Mathematics.
- [5] S. Asmussen and P. W. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer, New York, 2007.
- [6] R. G. Bartle. *The Elements of Integration and Lebesgue Measure*. John Wiley & Sons, Hoboken, 1995.
- [7] D. Bates and D. Watts. *Nonlinear Regression Analysis and Its Applications*. John Wiley & Sons, Hoboken, 1988.
- [8] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, second edition, 1985.
- [9] J. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.
- [10] P. J. Bickel and K. A. Doksum. *Mathematical Statistics*, volume I. Pearson Prentice Hall, Upper Saddle River, second edition, 2007.

- [11] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, third edition, 1995.
- [12] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2006.
- [13] P. T. Boggs and R. H. Byrd. Adaptive, limited-memory BFGS algorithms for unconstrained optimization. *SIAM Journal on Optimization*, 29(2):1282–1299, 2019.
- [14] Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *Annals of Statistics*, 38(5):2916–2957, 2010.
- [15] Z. I. Botev and D. P. Kroese. Global likelihood optimization via the cross-entropy method, with an application to mixture models. In R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, editors, *Proceedings of the 2004 Winter Simulation Conference*, pages 529–535, Washington, DC, December 2004.
- [16] Z. I. Botev, D. P. Kroese, R. Y. Rubinstein, and P. L'Ecuyer. The cross-entropy method for optimization. In V. Govindaraju and C.R. Rao, editors, *Machine Learning: Theory and Applications*, volume 31 of *Handbook of Statistics*, pages 35–59. Elsevier, 2013.
- [17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3:1–122, 2010.
- [18] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004. Seventh printing with corrections, 2009.
- [19] R. A. Boyles. On the convergence of the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 45(1):47–50, 1983.
- [20] L. Breiman. *Classification and Regression Trees*. CRC Press, Boca Raton, 1987.
- [21] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
- [22] L. Breiman. Heuristics of instability and stabilization in model selection. *Annals of Statistics*, 24(6):2350–2383, 12 1996.
- [23] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [24] F. Cao, D.-Z. Du, B. Gao, P.-J. Wan, and P. M. Pardalos. Minimax problems in combinatorial optimization. In D.-Z. Du and P. M. Pardalos, editors, *Minimax and Applications*, pages 269–292. Kluwer, Dordrecht, 1995.
- [25] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Press, Pacific Grove, second edition, 2001.
- [26] K. L. Chung. *A Course in Probability Theory*. Academic Press, New York, second edition, 1974.

- [27] E. Cinlar. *Introduction to Stochastic Processes*. Prentice Hall, Englewood Cliffs, 1975.
- [28] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.
- [29] J. W. Daniel, W. B. Gragg, L. Kaufman, and G. W. Stewart. Reorthogonalization and stable algorithms for updating the Gram-Schmidt QR factorization. *Mathematics of Computation*, 30(136):772–795, 1976.
- [30] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein. A tutorial on the cross-entropy method. *Annals of Operations Research*, 134(1):19–67, 2005.
- [31] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1 – 38, 1977.
- [32] L. Devroye. *Non-Uniform Random Variate Generation*. Springer, New York, 1986.
- [33] N. R. Draper and H. Smith. *Applied Regression Analysis*. John Wiley & Sons, New York, third edition, 1998.
- [34] Q. Duan and D. P. Kroese. Splitting for optimization. *Computers & Operations Research*, 73:119–131, 2016.
- [35] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, New York, 2001.
- [36] B. Efron and T. J. Hastie. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press, Cambridge, 2016.
- [37] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1994.
- [38] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, June 2006.
- [39] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume I. John Wiley & Sons, Hoboken, second edition, 1970.
- [40] J. C. Ferreira and V. A. Menegatto. Eigenvalues of integral operators defined by smooth positive definite kernels. *Integral Equations and Operator Theory*, 64:61–81, 2009.
- [41] N. I. Fisher and P. K. Sen, editors. *The Collected Works of Wassily Hoeffding*. Springer, New York, 1994.
- [42] G. S. Fishman. *Monte Carlo: Concepts, Algorithms and Applications*. Springer, New York, 1996.
- [43] R. Fletcher. *Practical Methods of Optimization*. John Wiley & Sons, New York, 1987.

- [44] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, 1997.
- [45] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2000.
- [46] A. Gelman. *Bayesian Data Analysis*. Chapman & Hall, New York, second edition, 2004.
- [47] A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge, 2006.
- [48] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.
- [49] J. E. Gentle. *Random Number Generation and Monte Carlo Methods*. Springer, New York, second edition, 2003.
- [50] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, New York, 1996.
- [51] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.
- [52] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, fourth edition, 2013.
- [53] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, Cambridge, 2016.
- [54] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford University Press, third edition, 2001.
- [55] T. J. Hastie, R. J. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data mining, Inference, and Prediction*. Springer, New York, 2009.
- [56] T. J. Hastie, R. J. Tibshirani, and M. Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press, Boca Raton, 2015.
- [57] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Springer, New York, 2001.
- [58] W. Hock and K. Schittkowski. *Test Examples for Nonlinear Programming Codes*. Springer, New York, 1981.
- [59] J. E. Kelley, Jr. The cutting-plane method for solving convex programs. *Journal of the Society for Industrial and Applied Mathematics*, 8(4):703–712, 1960.
- [60] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, Englewood Cliffs, 1989.

- [61] O. Kallenberg. *Foundations of Modern Probability*. Springer, New York, second edition, 2002.
- [62] A. Karalic. Linear regression in regression tree leaves. In *Proceedings of ECAI-92*, pages 440–441, Hoboken, 1992. John Wiley & Sons.
- [63] C. Kaynak. Methods of combining multiple classifiers and their applications to handwritten digit recognition. Master's thesis, Institute of Graduate Studies in Science and Engineering, Bogazici University, 1995.
- [64] T. Keilath and A. H. Sayed, editors. *Fast Reliable Algorithms for Matrices with Structure*. SIAM, Pennsylvania, 1999.
- [65] C. Nussbaumer Knaflitz. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. John Wiley & Sons, Hoboken, 2015.
- [66] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, Cambridge, 2009.
- [67] A. N. Kolmogorov and S. V. Fomin. *Elements of the Theory of Functions and Functional Analysis*. Dover Publications, Mineola, 1999.
- [68] D. P. Kroese, T. Brereton, T. Taimre, and Z. I. Botev. Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(6):386–392, 2014.
- [69] D. P. Kroese and J. C. C. Chan. *Statistical Modeling and Computation*. Springer, 2014.
- [70] D. P. Kroese, S. Porotsky, and R. Y. Rubinstein. The cross-entropy method for continuous multi-extremal optimization. *Methodology and Computing in Applied Probability*, 8(3):383–407, 2006.
- [71] D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods*. John Wiley & Sons, New York, 2011.
- [72] H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, New York, second edition, 2003.
- [73] P. Lafaye de Micheaux, R. Drouilhet, and B. Liqueur. *The R Software: Fundamentals of Programming and Statistical Analysis*. Springer, New York, 2014.
- [74] R. J. Larsen and M. L. Marx. *An Introduction to Mathematical Statistics and Its Applications*. Prentice Hall, New York, third edition, 2001.
- [75] A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis*. McGraw-Hill, New York, third edition, 2000.
- [76] P. L'Ecuyer. A unified view of IPA, SF, and LR gradient estimation techniques. *Management Science*, 36:1364–1383, 1990.

- [77] P. L'Ecuyer. Good parameters and implementations for combined multiple recursive random number generators. *Operations Research*, 47(1):159 – 164, 1999.
- [78] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, New York, second edition, 1998.
- [79] T. G. Lewis and W. H. Payne. Generalized feedback shift register pseudorandom number algorithm. *Journal of the ACM*, 20(3):456–468, 1973.
- [80] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. John Wiley & Sons, Hoboken, second edition, 2002.
- [81] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, 1989.
- [82] M. Lutz. *Learning Python*. O'Reilly, fifth edition, 2013.
- [83] M. Matsumoto and T. Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, 1998.
- [84] W. McKinney. *Python for Data Analysis*. O'Reilly Media, Inc., second edition, 2017.
- [85] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons, Hoboken, second edition, 2008.
- [86] G. J. McLachlan and D. Peel. *Finite Mixture Models*. John Wiley & Sons, New York, 2000.
- [87] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.
- [88] C. A. Micchelli, Y. Xu, and H. Zhang. Universal kernels. *Journal of Machine Learning Research*, 7:2651–2667, 2006.
- [89] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, New York, third edition, 1996.
- [90] J. F. Monahan. *Numerical Methods of Statistics*. Cambridge University Press, London, 2010.
- [91] T. A. Mroz. The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 55(4):765–799, 1987.
- [92] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, Cambridge, 2012.
- [93] J. Neyman and E. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London, Series A*, 231:289–337, 1933.

- [94] M. A. Nielsen. *Neural Networks and Deep Learning*, volume 25. Determination Press, 2015.
- [95] K. B. Petersen and M. S. Pedersen. *The Matrix Cookbook*. Technical University of Denmark, 2008.
- [96] J. R. Quinlan. Learning with continuous classes. In A. Adams and L. Sterling, editors, *Proceedings AI'92*, pages 343–348, Singapore, 1992. World Scientific.
- [97] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, 2006.
- [98] B. D. Ripley. *Stochastic Simulation*. John Wiley & Sons, New York, 1987.
- [99] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, second edition, 2004.
- [100] S. M. Ross. *Simulation*. Academic Press, New York, third edition, 2002.
- [101] S. M. Ross. *A First Course in Probability*. Prentice Hall, Englewood Cliffs, seventh edition, 2005.
- [102] R. Y. Rubinstein. The cross-entropy method for combinatorial and continuous optimization. *Methodology and Computing in Applied Probability*, 2:127–190, 1999.
- [103] R. Y. Rubinstein and D. P. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. Springer, New York, 2004.
- [104] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York, third edition, 2017.
- [105] S. Ruder. An overview of gradient descent optimization algorithms. *arXiv: 1609.04747*, 2016.
- [106] W. Rudin. *Functional Analysis*. McGraw–Hill, Singapore, second edition, 1991.
- [107] D. Salomon. *Data Compression: The Complete Reference*. Springer, New York, 2000.
- [108] G. A. F. Seber and A. J. Lee. *Linear Regression Analysis*. John Wiley & Sons, Hoboken, second edition, 2003.
- [109] S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, Cambridge, 2014.
- [110] Z. A. Shaw. *Learning Python 3 the Hard Way*. Addison–Wesley, Boston, 2017.
- [111] Y. Shen, S. Kiatsupaibul, Z. B. Zabinsky, and R. L. Smith. An analytically derived cooling schedule for simulated annealing. *Journal of Global Optimization*, 38(2):333–365, 2007.

- [112] N. Z. Shor. *Minimization Methods for Non-differentiable Functions*. Springer, Berlin, 1985.
- [113] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, New York, 1986.
- [114] J. S. Simonoff. *Smoothing Methods in Statistics*. Springer, New York, 2012.
- [115] I. Steinwart and A. Christmann. *Support Vector Machines*. Springer, New York, 2008.
- [116] G. Strang. *Introduction to Linear Algebra*. Wellesley–Cambridge Press, Cambridge, fifth edition, 2016.
- [117] G. Strang. *Linear Algebra and Learning from Data*. Wellesley–Cambridge Press, Cambridge, 2019.
- [118] W. N. Street, W. H. Wolberg, and O. L. Mangasarian. Nuclear feature extraction for breast tumor diagnosis. In *IS&T/SPIE 1993 International Symposium on Electronic Imaging: Science and Technology*, San Jose, CA, pages 861–870, 1993.
- [119] V. M. Tikhomirov. On the representation of continuous functions of several variables as superpositions of continuous functions of one variable and addition. In *Selected Works of A. N. Kolmogorov*, pages 383–387. Springer, Berlin, 1991.
- [120] S. van Buuren. *Flexible Imputation of Missing Data*. CRC Press, Boca Raton, second edition, 2018.
- [121] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
- [122] V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 16(2):264–280, 1971.
- [123] G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.
- [124] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2010.
- [125] A. Webb. *Statistical Pattern Recognition*. Arnold, London, 1999.
- [126] H. Wendland. *Scattered Data Approximation*. Cambridge University Press, Cambridge, 2005.
- [127] D. Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.
- [128] C. F. J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.