Books

M. Mohri, A. Rostamizadeh, and A. Talwalkar 2012. *Foundations of Machine Learning*. MIT Press (Chp. 8, 11, 13)

S. Shalev-Shwartz, S. Ben-David (2014). *Understanding Machine Learning*. Cambridge (Chp. 1, 14)

C. M. Bishop (2006). *Pattern Recognition and Machine Learning*. Springer (Chp. 6, 7, 12, 15)

T. Hastie, R. Tibshirani, and J. Friedman 2009. *The Elements of Statistical Learning*. Springer (Chp. 3, 4, 5, 9)

L. Györfi, M. Kohler, A. Krzyzak, and H. Walk (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer (Chp. 1, 2, 14)

C. Giraud (2015). *Introduction to High-Dimensional Statistics*. CRC (Chp. 4)

L. Devroye, L. Györfi, and G. Lugosi (1996). *A Probabilistic Theory of Pattern Recognition*. Springer (Chp. 1, 2)

Articles

L. Breiman (2001). Statistical Modeling: The Two Cultures. *Statist. Sci.*, 16(3):199–231 (Chp. 1)

D. Donoho. *50 years of data science* (Chp. 1)

P. Domingos (2000). A Unified Bias-Variance Decomposition for Zero-One and Squared Loss. *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence* (Chp. 1)

R. Tibshirani (1996). Regression Shrinkage and Selection via the Lasso. *Journal Royal Stat. Society: Series B*, 58(1):267–288 (Chp. 4)

R. Tibshirani (1996). The Lasso Problem and Uniqueness. *Electron. J. Statist.*, 7:1456–1490(Chp. 4)

H. Zou & T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal Royal Stat. Society: Series B*, 67(2):301–320 (Chp. 4)

J. Friedman, T. Hastie, H. Hofling, and R. Tibshirani (2007). Pathwise Coordinate Optimization. *The Annals of Appl.Stat.*, 2(1):302–332 (Chp. 4)

H. Zou, T. Hastie, and R. Tibshirani (2007). On the Degrees of Freedom of the Lasso. *The Annals of Statistics*, 35(5):2173–2192 (Chp. 4)

J. Friedman, T. Hastie, R. Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1):1–22 (Chp. 6)

T. Zhang (2004). Statistical Behaviour and Consistency of Classification Methods Based on Convex Risk Minimization. *Ann. of Stat.*, 32(1):56–134 (Chp. 10)

P. Bartlett, M. Jordan, and J. Mcauliffe(2006). Convexity, Classification, and Risk Bounds. *Statistical Science*, 21(3):341–346 (Chp. 10)

Y. Freund, and R. Shapire (1996). Experiments with a New Boosting Algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference*, 148–156 (Chp. 11)

J. Zhu, H. Zou, S. Rosset, and T. Hastie (2009). Multi-class AdaBoost. *Statistics and its Interface*,2:349–360 (Chp. 11)

J. Friedman (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 29(5):1189–1232 (Chp. 11)

C. Cortes, and V. Vapnik (1995). Support Vector Networks. *Machine Learning*, 20:273–297 (Chp. 12)

J. Platt(1998). Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines.*Advances in Kernel Methods - Support Vector Learning* (Chp. 12)

S. Shalev-Schwartz, Y. Singer, N. Srebro, and A. Cotter (2011). Pegasos: Primal Estimated Sub-Gradient Solver for SVM. *Mathematical Programming*, 127(1):3-30.(Chp. 12 & 14)

B. Schlkopf, A.J. Smola, and R.C. Williamson (2000). New Support Vector Algorithms. *Neural Computation*, 12:1207111245 (Chp. 12)