Books

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep learning*. MIT press. URL: https://www.deeplearningbook.org/contents/regularization.html.

Articles

Theobald, Chris M (1974). "Generalizations of mean square error applied to ridge regression". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 36.1, pp. 103–106. URL: https://www.jstor.org/stable/2984775.

Farebrother, Richard William (1976). "Further results on the mean square error of ridge regression". In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 248–250. URL: https://www.jstor.org/stable/2984971.

Hanson, Stephen and Lorien Pratt (1988). "Comparing biases for minimal network construction with back-propagation". In: *Advances in neural information processing systems* 1. URL: https://proceedings.neurips.cc/paper_files/paper/1988/file/1c9ac0159c94d8d0cbedc973445af2da-Paper.pdf.

Tibshirani, Robert (1996). "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1, pp. 267–288. URL: https://www.jstor.org/stable/2346178.

Fu, Wenjiang and Keith Knight (2000). "Asymptotics for lasso-type estimators". In: *The Annals of Statistics* 28.5, pp. 1356–1378. URL: https://doi.org/10.1214/aos/1015957397.

Fan, Jianqing and Runze Li (2001). "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties". In: *Journal of the American Statistical Association* 96.456, pp. 1348–1360. URL: https://doi.org/10.1198/016214501753382273.

Zou, Hui and Trevor Hastie (2005). "Regularization and variable selection via the elastic net". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67.2, pp. 301–320. URL: https://academic.oup.com/jrsssb/article/67/2/301/7109482?login=false.

Yuan, Ming and Yi Lin (2006). "Model selection and estimation in regression with grouped variables". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 68.1, pp. 49–67. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00532.x.

Zhao, Peng and Bin Yu (2006). "On model selection consistency of Lasso". In: *The Journal of Machine Learning Research* 7, pp. 2541–2563. URL: https://www.jmlr.org/papers/volume7/zhao06a/zhao06a.pdf.

Meinshausen, Nicolai and Bin Yu (2009). "LASSO-TYPE RECOVERY OF SPARSE REPRESENTATIONS FOR HIGH-DIMENSIONAL DATA". In: *The Annals of Statistics* 37.1, pp. 246–270. URL: https://projecteuclid.

org/journals/annals-of-statistics/volume-37/issue-1/Lasso-type-recovery-of-sparse-representations-for-high-dimensional-data/10.1214/07-AOS582.full.

Zhang, Cun-Hui (2010). "Nearly unbiased variable selection under minimax concave penalty". In: *The Annals of Statistics* 38.2, pp. 894–942. URL: https://doi.org/10.1214/09-AOS729.

Hans, Chris (2011). "Elastic net regression modeling with the orthant normal prior". In: *Journal of the American Statistical Association* 106.496, pp. 1383–1393. URL: https://www.jstor.org/stable/23239545.

Neyshabur, Behnam, Ryota Tomioka, and Nathan Srebro (2015). "In Search of the Real Inductive Bias: On the Role of Implicit Regularization in Deep Learning". In: arXiv: 1412.6614 [cs.LG]. URL: https://arxiv.org/abs/1412.6614.

Loshchilov, Ilya and Frank Hutter (2019). "Decoupled Weight Decay Regularization". In: *International Conference on Learning Representations*. URL: https://arxiv.org/pdf/1711.05101.

Others

Ali, Alnur, Edgar Dobriban, and Ryan Tibshirani (2020). "The implicit regularization of stochastic gradient flow for least squares". In: *International conference on machine learning*. PMLR, pp. 233–244. URL: https://proceedings.mlr.press/v119/ali20a/ali20a.pdf.