Statistical Learning Literature

Annotated

Compiled by Edoardo Costantini

2021-01-14

Books

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1) (No. 10). Springer series in statistics New York.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC press.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Regularized Regression

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418–1429.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301–320.

Paper proposing Elastic Net (EN) Penalty.

Methods

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in medicine*, *25*(24), 4279–4292.

This source is fundamental to plan simulation studies and justify decisions made for all its aspects. It contains great descriptions of performance measures, the idea behind them, their computation, and assesment. In particular, it is helpful for 'Confidence Interval Coverage' providing a helpful simulation specific rule of thumb.

Ensemble Methods

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

Multi-View Data Research

Li, Y., Wu, F.-X., & Ngom, A. (2018). A review on machine learning principles for multi-view biological data integration. *Briefings in bioinformatics*, *19*(2), 325–340.

Good source to review how multiple sources of data (multi-view) on the same subjects are analysed in medical research.

van Loon, W., Fokkema, M., Szabo, B., & de Rooij, M. (2020). Stacked penalized logistic regression for selecting views in multi-view learning. *Information Fusion*.

Interesting approach to analyse multi-view data.

Software

pandas development team, T. (2020, February). *pandas-dev/pandas: Pandas*. Zenodo. Retrieved from https://doi.org/10.5281/zenodo.3509134 doi: 10.5281/zenodo.3509134

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/