Math 867 Spring 2015 Topics and Materials for Projects

Each team is free to choose one topic from either the theory and methods topics
(TM1–TM6) or the data analysis topics (DA1–DA3).

TM1. Error control in variable selection

Barber, R. F. and Candès, E. J. (2015). Controlling the false discovery rate via knockoffs. *Annals of Statistics*, to appear.

Shah, R. D. and Samworth, R. J. (2013). Variable selection with error control: Another look at stability selection. *Journal of the Royal Statistical Society, Series B*, 75, 55–80.

TM2. Model selection criteria and degrees of freedom

Fan, Y. and Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society, Series B*, 75, 531–552.

Kaufman, S. and Rosset, S. (2014). When does more regularization imply fewer degrees of freedom? Sufficient conditions and counterexamples. *Biometrika*, 101, 771–784.

TM3. Sparse discriminant analysis

Mai, Q., Zou, H. and Yuan, M. (2012). A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika*, 99, 29–42.

Witten, D. M. and Tibshirani, R. (2011). Penalized classification using Fisher's linear discriminant. *Journal of the Royal Statistical Society, Series B*, 73, 753–772.

TM4. Exponential family graphical models

Loh, P.-L. and Wainwright, M. J. (2013). Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *Annals of Statistics*, 41, 3022–3049.

Ravikumar, P., Wainwright, M. J. and Lafferty, J. D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Annals of Statistics*, 38, 1287–1319.

TM5. Sparse sufficient dimension reduction

Chen, X., Zou, C. and Cook, R. D. (2010). Coordinate-independent sparse sufficient dimension reduction and variable selection. *Annals of Statistics*, 38, 3696–3723.

Li, L. (2007). Sparse sufficient dimension reduction. *Biometrika*, 94, 603–613.

TM6. Proximal algorithms

Chen, X., Lin, Q., Kim, S., Carbonell, J. G. and Xing, E. P. (2012). Smoothing proximal gradient method for general structured sparse regression. *Annals of Applied Statistics*, 6, 719–752.

Parikh, N. and Boyd, S. (2014). Proximal algorithms. *Foundations and Trends in Optimization*, 1, 127–239.

DA1. LSI Data Sets

Data sets accompanying Efron's book are available at http://statweb.stanford.edu/~ckirby/brad/LSI/datasets-and-programs/datasets.html.

DA2. ASAIP Data Sets

The Astrostatistics and Astroinformatics Portal (ASAIP) summarizes some astronomical data sets at https://asaip.psu.edu/resources/datasets.

DA3. EPA Data Sets

Air quality data sets are available at the U.S. Environmental Protection Agency (EPA)'s website: http://www.epa.gov/airdata/ad_data.html.