Math 867 Spring 2015 Homework 1

Due: April 22, 2015

1. (Gaussian tail bound) Let $Z \sim N(0, \sigma^2)$. Show that
$$\sup_{t>0} \left(P(Z \geq t) e^{t^2/(2\sigma^2)} \right) = \frac{1}{2}.$$

2. (Irrepresentable condition) Consider the covariance matrix $\mathbf{\Sigma} = (\sigma_{ij})$ with an autoregressive Toeplitz structure: $\sigma_{ij} = \rho^{|i-j|}$ for all i and j with $0 < |\rho| < 1$. Show that the irrepresentable condition
$$\|\mathbf{\Sigma}_{S^c S}(\mathbf{\Sigma}_{SS})^{-1} \operatorname{sgn}(\boldsymbol{\beta}^*_S)\|_\infty \leq \alpha < 1$$
holds and identify the constant α.

3. (Model size of the Lasso) Assume what we have done in class. Prove inequality (7.9) in Bickel, Ritov and Tsybakov (2009): with probability tending to 1, the Lasso estimator $\widehat{\boldsymbol{\beta}}_L$ satisfies
$$\mathcal{M}(\widehat{\boldsymbol{\beta}}_L) \leq \frac{64\phi_{\max}}{\kappa^2(s,3)} s,$$
where $\mathcal{M}(\boldsymbol{\beta}) = \sum_{j=1}^p I(\beta_j \neq 0)$ and $\phi_{\max}$ is the maximal eigenvalue of $\mathbf{X}^T\mathbf{X}/n$.

4. (Residual of the Dantzig selector) For the linear regression model, show that, with probability tending to 1, the residual $\boldsymbol{\delta} = \widehat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}^*$ of the Dantzig selector $\widehat{\boldsymbol{\beta}}_D$ satisfies $\|\boldsymbol{\delta}_{S^c}\|_1 \leq \|\boldsymbol{\delta}_S\|_1$.

5. (Dirichlet–multinomial regression) Derive the log-likelihood function of the Dirichlet–multinomial regression model (eq. (7) in Chen and Li (2013)).

6. (Coordinate descent with nonconvex penalties) In the proof of Theorem 4 in Mazumder, Friedman and Hastie (2011), the authors arrived at inequality (A.13):
$$Q(\boldsymbol{\beta}^m) - Q(\boldsymbol{\beta}^{m+1}) \geq \theta \|\boldsymbol{\beta}^{m+1} - \boldsymbol{\beta}^m\|_2^2,$$
from which they concluded that the sequence $\{\boldsymbol{\beta}^k\}$ converges. Prove or disprove this claim.

7. (ADMM for group Lasso) Derive the ADMM algorithm in scaled form for the group Lasso problem
$$\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \sum_{g=1}^G \|\boldsymbol{\beta}_g\|_2 \right\},$$
where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_G^T)^T$.

8. (Bayesian elastic net) Recall that the elastic net penalty is $\lambda_1\|\boldsymbol{\beta}\|_1 + \lambda_2\|\boldsymbol{\beta}\|_2^2$. Derive a hierarchical representation of the Bayesian elastic net similar to eq. (5) in Park and Casella (2008).