ELE 538B: Sparsity, Structure and Inference	Spring 2017

Homework 2

Due date: Wednesday, Mar. 29, 2017 (at the beginning of class)

You are allowed to drop 1 subproblem without penalty. In addition, up to 1 bonus point will be awarded to each subproblem for clean, well-organized, and elegant solutions.

1. Prediction error for the MMSE estimator (*20 points*)

Suppose that the training data is $(\boldsymbol{y} \in \mathbb{R}^n, \boldsymbol{X} \in \mathbb{R}^{n\times n})$ obeying

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \qquad \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

(a) Suppose that we know a priori that $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. What is the Bayes minimum mean square error (MMSE) estimate $\hat{\boldsymbol{\beta}}^{\mathsf{mmse}}$, i.e. the estimate that minimizes $\mathbb{E}\left[\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\|^2\right]$?

(b) We now want to use $\boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathsf{mmse}}$ to predict new data $\tilde{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta} + \tilde{\boldsymbol{\eta}}$ with $\tilde{\boldsymbol{\eta}} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ independent of $\boldsymbol{\eta}$ (note that $\tilde{\boldsymbol{y}}$ is unobserved). Can you come up with an unbiased estimate for the prediction error $\mathsf{PE} := \mathbb{E}[\|\tilde{\boldsymbol{y}} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathsf{mmse}}\|^2]$?

2. Lasso with a single parameter (*10 points*)

Consider the single parameter setting $\boldsymbol{y} = \beta \boldsymbol{z} + \boldsymbol{\eta}$ with $\beta \in \mathbb{R}$. In this case, the Lasso estimator is given by

$$\text{minimize}_{\hat{\beta}\in\mathbb{R}} \quad \frac{1}{2}\|\boldsymbol{y} - \hat{\beta}\boldsymbol{z}\|^2 + \lambda|\hat{\beta}|.$$

Show that $\hat{\beta} = \psi_{\mathsf{st}}\left(\frac{\boldsymbol{z}^\top \boldsymbol{y}}{\|\boldsymbol{z}\|^2}; \frac{\lambda}{\|\boldsymbol{z}\|^2}\right)$ is a closed-form solution to the above program, where $\psi_{\mathsf{st}}(x;\lambda) = \text{sign}(x)\max\{|x| - \lambda, 0\}$ is the soft-thresholding operator. You should use the optimality condition based on subgradients.

3. Convexity of the SLOPE estimator (*30 points*)

For any $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_p]^\top \in \mathbb{R}^p$, let $|\beta|_{(1)} \geq |\beta|_{(2)} \geq \cdots \geq |\beta|_{(p)}$ denote the order statistics of $\{|\beta_1|, \cdots, |\beta_p|\}$, i.e. $|\beta|_{(i)}$ is the ith largest in $\{|\beta_1|, \cdots, |\beta_p|\}$.

(a) Suppose that $p = 2$. Show that the function

$$f(\boldsymbol{\beta}) := \lambda_1|\beta|_{(1)} + \lambda_2|\beta|_{(2)}$$

is convex if $\lambda_1 \geq \lambda_2 \geq 0$.

(b) Show that the function

$$g_k(\boldsymbol{\beta}) = \sum_{i=1}^{k} |\beta|_{(i)}$$

is convex for any $1 \leq k \leq p$.

(c) Show that the function

$$f(\boldsymbol{\beta}) := \sum_{i=1}^{p} \lambda_i |\beta|_{(i)}$$

is convex for any $p \geq 3$, as long as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. This justifies that SLOPE (Sorted L-One Penalized Estimation)

$$\text{minimize}_{\boldsymbol{\beta}} \quad \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \sum_{i=1}^{p} \lambda_i |\beta|_{(i)}$$

is a convex program.

4. Gaussian graphical models (*20 points*)

(a) Consider a p-dimensional Gaussian vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. For any $1 \leq u, v \leq p$, show that

$$x_u \perp\!\!\!\perp x_v \mid \boldsymbol{x}_{\mathcal{V}\setminus\{u,v\}}$$

(namely, x_u and x_v are conditionally independent given all other variables) if and only if $\Theta_{u,v} = 0$. Here, $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$.

(b) In graphical lasso, the objective function includes a term $\log\det\boldsymbol{\Theta}$. Show that $g(\boldsymbol{\Theta}) := \log\det(\boldsymbol{\Theta})$ ($\boldsymbol{\Theta} \succ \boldsymbol{0}$) is a concave function.

Hint: A function $g(\boldsymbol{\Theta})$ is concave if $h(t) := g(\boldsymbol{\Theta} + t\boldsymbol{V})$ is concave for all t and $\boldsymbol{V}$ obeying $\boldsymbol{\Theta} + t\boldsymbol{V} \succ \boldsymbol{0}$.