STAT 542: Homework 1

Due: Feb. 9 midnight on Canvas

Please make sure that your solutions are readable and the file size is reasonable. Typing the answers is highly encouraged.

Problem 1.

Suppose that the true observation model is given by

$$Y = X\beta + \epsilon \tag{1}$$

where $X \in \mathbb{R}^{n\times 2}$, and ϵ satisfies $\mathbb{E}[\epsilon] = 0$ and $\mathbb{E}[\epsilon\epsilon^\top] = \sigma^2 I$. Further assume that the $X_1, X_2 \in \mathbb{R}^n$ are the two columns of X, $\|X_1\|_2 = \|X_2\|_2 = 1$, and the inner product $\langle X_1, X_2\rangle = r$. Denote by

$$\hat{\beta} := (X^\top X)^{-1} X^\top Y \tag{2}$$

and OLS estimator using the full model, and

$$\hat{\beta}^{\mathsf{r}} := (X_1^\top X_1)^{-1} X_1^\top Y \tag{3}$$

the OLS estimator using the reduced model.

- [1pts] Suppose that we are only interested in estimating the first coordinate, β_1. Compute $\mathbb{E}[\hat{\beta}_1]$ and $\text{var}(\hat{\beta}_1)$ (express the answers using β, σ and r).
- [2pts] Compute $\mathbb{E}[\hat{\beta}_1^{\mathsf{r}}]$ and $\text{var}(\hat{\beta}_1^{\mathsf{r}})$.
- [2pts] Use the bias-variance tradeoffs to compute the mean square errors of $\hat{\beta}_1$ and $\hat{\beta}_1^{\mathsf{r}}$ (defined as $\mathbb{E}[|\hat{\beta}_1 - \beta_1|^2]$ and $\mathbb{E}[|\hat{\beta}_1^{\mathsf{r}} - \beta_1|^2]$). Find the range of β_2 for which the reduced model has a smaller mean square error than the full model. *Hint: note that when $|\beta_2|$ is large, you would expect that the full model is "more correct" and hence having a smaller error.*

Problem 2.

Use R or Python to perform the following experiment: you pick arbitrary numbers $\rho \in (0,1)$ and $r \in (0,1)$ satisfying

$$\frac{6\rho}{1+\rho^2} > \frac{1}{r} + 2r. \tag{4}$$

Set $X = \begin{pmatrix} 1 & \rho r \\ \rho & r \end{pmatrix}$ and $Y = X \begin{pmatrix} -1 \\ 2 \end{pmatrix}$. For any $\lambda > 0$, define

$$\hat{\beta}_\lambda := \arg\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\}. \tag{5}$$

Plot the coefficients of $\hat{\beta}_\lambda$ as a function of $\|\hat{\beta}_\lambda\|_1$, and repeat the experiments with different ρ and r satisfying (4). Include the plots in your solution. Do you find $\|\hat{\beta}_\lambda\|_0$ to be a monotonic function of the ℓ_1 norm or not? What is the implication of this phenomenon for implementing the LARS algorithm?

Hint: `lasso.R` *in Canvas contains most of the ingredients of the code. Note that using R code you can easily plot the lasso coefficients with the L_1 norm (see slides). Also, beware that the default options for the intercept and feature normalizations of the R function may not be what you want.*

Problem 3.

Generate a design matrix $X \in \mathbb{R}^{100 \times 200}$ and let $\beta \in \mathbb{R}^{200}$ be defined as

$$\beta_j = 1\{j \leq 30\}, \quad j = 1, \ldots, 200 \tag{6}$$

where $1\{\}$ denotes the indicator function. In the model $Y = X\beta + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, compute the optimal lasso regularization parameter λ_{opt} using cross-validation by R (Caution: no intercept and column normalization). Study the trend of λ_{opt} as σ varies, by plotting a figure showing their dependence.