Statistical Methods - Homework #3

1. Consider the multiple linear regression model:

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim (0, \sigma^2) \;\; \text{independently for } i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n, \boldsymbol{\beta} \in \mathbb{R}^p$ and $X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)^\top \in \mathbb{R}^{n \times p}$ is full rank. ($\text{rank}(X) = p$) Letting $\boldsymbol{y} = (y_1, \cdots, y_n)^\top$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, (1) can equivalently be written as

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \text{ and } \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 I_n.$$

(A) Find the ridge estimator of $\boldsymbol{\beta}$, denoted $\hat{\boldsymbol{\beta}}_\lambda$, that minimizes

$$\ell(\boldsymbol{\beta}) = (\boldsymbol{y} - X\boldsymbol{\beta})^\top (\boldsymbol{y} - X\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^\top \boldsymbol{\beta} = \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2 + \lambda \boldsymbol{\beta}^\top \boldsymbol{\beta}.$$

(B) Find $\text{E}(\hat{\boldsymbol{\beta}}_\lambda)$. What is the value of λ that makes $\text{E}(\hat{\boldsymbol{\beta}}_\lambda) = \boldsymbol{\beta}$?

(C) Find $\mathrm{Var}(\hat{\boldsymbol{\beta}}_\lambda)$. Also, show that

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_\lambda) \preceq \mathrm{Var}(\hat{\boldsymbol{\beta}}_0).$$

(Hint: You may use eigen-value decomposition on $X^\top X$: There exists a matrix $P \in \mathbb{R}^{p\times p}$ and a diagonal matrix $\Lambda \in \mathbb{R}^{p\times p}$ such that $P^\top P = PP^\top = I$ and $X^\top X = P\Lambda P^\top$).

(D) Show the bias-variance decomposition of mean-squared error(MSE):

$$\mathrm{E}\left[(\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta})^\top\right] = \mathrm{Var}(\hat{\boldsymbol{\beta}}_\lambda) + (\boldsymbol{\beta} - E[\hat{\boldsymbol{\beta}}_\lambda])(\boldsymbol{\beta} - E[\hat{\boldsymbol{\beta}}_\lambda])^\top.$$