Statistics Prelim Exam, Linear Models part

12:00 pm - 2:30 pm, Wednesday, May 1, 2024

Q1 counts for 60 points and Q2, 40 points. Answer questions with showing all of your work.

1. (60 points) We condier a linear model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{y} = (y_1, y_2, y_3, y_4, y_5, y_6)'$, $\boldsymbol{\beta} = (\mu, \tau_1, \tau_2, \tau_3, \tau_4)'$,

$$\boldsymbol{X} = \begin{pmatrix} 1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & -1 \\ 1 & 1 & 0 & -1 & 0 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 1 & -1 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

   (a) Find the rank of $\boldsymbol{X}'$.

   (b) Show that $\boldsymbol{\lambda}'\boldsymbol{\beta} = -\tau_1 - \tau_2 + \tau_3 + \tau_4$ is an estimable function of $\boldsymbol{\beta}$.

   (c) Express the best linear unbiased estimator (BLUE) of $\boldsymbol{\lambda}'\boldsymbol{\beta}$ with regard to $\boldsymbol{y}$, $\boldsymbol{X}$ and $\boldsymbol{\lambda}$.

   (d) What is the rank of $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$? Justify your answer in details.

   (e) Prove that MSE $= \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$ is unbiased of σ^2.

   (f) Find the distribution of the BLUE of $\boldsymbol{\lambda}'\boldsymbol{\beta}$.

   (g) Show that $100(1-\alpha)\%$ confidence interval of $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is

$$\left(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - t_{\frac{\alpha}{2},2}\sqrt{MSE\ \boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\lambda}},\quad \boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} + t_{\frac{\alpha}{2},2}\sqrt{MSE\ \boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{\lambda}}\right).$$

   Show all steps in details.

2. (40 points) Consider the linear model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{X}$ is an $n \times p$ matrix with full rank and $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I})$. Let $\hat{\boldsymbol{\beta}}$ be the ordinary least squares estimator of $\boldsymbol{\beta}$.

   The ridge regression estimator $\hat{\boldsymbol{\beta}}_R(\lambda)$ is the vector-value of $\hat{\boldsymbol{\beta}}$ that minimizes

$$Q(\boldsymbol{\beta}) = (\mathbf{Y} - \boldsymbol{X\beta})'(\mathbf{Y} - \boldsymbol{X\beta}) + \lambda \boldsymbol{\beta}'\boldsymbol{\beta},$$

   where $\lambda > 0$ is a fixed real number.

   (a) Find an expression for $\hat{\boldsymbol{\beta}}_R(\lambda)$. Show that you can write $\hat{\boldsymbol{\beta}}_R(\lambda) = \mathbf{W}_\lambda \hat{\boldsymbol{\beta}}$ for some matrix $\mathbf{W}_\lambda$ that depends on λ and give the explicit expression for $\mathbf{W}_\lambda$.

   (b) Find the mean, bias and variance of $\hat{\boldsymbol{\beta}}_R(\lambda)$.

   (c) Show that $Var(\hat{\boldsymbol{\beta}}_R(\lambda)) < Var(\hat{\boldsymbol{\beta}})$ in the sense that $Var(\hat{\boldsymbol{\beta}}) - Var(\hat{\boldsymbol{\beta}}_R(\lambda))$ is positive definite.

   Hint: You may use the fact that $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{PDP}'$ where $\boldsymbol{D}$ is a diagonal matrix with elements $d_i > 0$ and $\boldsymbol{P}$ is an orthogonal matrix.

   (d) Show that the mean squared error of $\hat{\boldsymbol{\beta}}_R(\lambda)$ is expressed by

$$E\{(\hat{\boldsymbol{\beta}}_R(\lambda) - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}_R(\lambda) - \boldsymbol{\beta})\} = \boldsymbol{\beta}'\boldsymbol{P}'\boldsymbol{AP\beta} + \sigma^2 \boldsymbol{PBP}'$$

   where

   - $\boldsymbol{P}$ is an orthogonal matrix and d_i is the i-th diagonal element of the matrix $\boldsymbol{D}$ defined in part (c),
   - $\boldsymbol{A}$ is a diagonal matrix with elements $\left(\frac{\lambda}{d_i+\lambda}\right)^2$, and
   - $\boldsymbol{B}$ is a diagonal matrix with elements $\frac{d_i}{(d_i+\lambda)^2}$.

   (e) Assuming now that $\boldsymbol{X}'\boldsymbol{X} = d\boldsymbol{I}_p$ where d is a fixed constant, find the optimum value of λ that minimizes the mean squared error of $\hat{\boldsymbol{\beta}}_R(\lambda)$.