STAT5030 Linear Models (Final Exam 2019-2020)

11 May 2020

1. (**30 marks**) Consider a linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{Y}$ is $n \times 1$, $\boldsymbol{X}$ is an $n \times p$ fixed design matrix, $\boldsymbol{\beta}$ is a p-vector of regression coefficient and $\boldsymbol{\epsilon}$ has mean $\mathbf{0}$ and known positive definite covariance matrix $\boldsymbol{V}$.

   (a) When $\boldsymbol{X}$ is of full rank, find the covariance of the best linear unbiased estimates (BLUE) of $\boldsymbol{p}^\top\boldsymbol{\beta}$ and $\boldsymbol{q}^\top\boldsymbol{\beta}$, where $\boldsymbol{p} \in \mathbb{R}^p$ and $\boldsymbol{p} \in \mathbb{R}^p$ are constant vectors.

   (b) When $\boldsymbol{X}$ is not of full rank, find a sufficient and necessary condition for $\boldsymbol{c}^\top\boldsymbol{\beta}$ to be estimable, where $\boldsymbol{c} \in \mathbb{R}^p$.

   (c) When $\boldsymbol{X}$ is not of full rank, find the covariance of the best linear unbiased estimates (BLUE) of two estimable functions $\boldsymbol{p}^\top\boldsymbol{\beta}$ and $\boldsymbol{q}^\top\boldsymbol{\beta}$, where $\boldsymbol{p} \in \mathbb{R}^p$ and $\boldsymbol{p} \in \mathbb{R}^p$ are constant vectors.

2. (**15 marks**) Let $\boldsymbol{x} = (X_1, \ldots, X_k)^\top \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where μ is a $k \times 1$ constant vector and $rank(\Sigma) = k$.

   (a) What is the distribution of $U = (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$?

   (b) Let $\boldsymbol{A} = \boldsymbol{\Sigma}^{-\mathbf{1}} - (\boldsymbol{\Sigma}^{-\mathbf{1}}\mathbf{1}_{\boldsymbol{k}}\mathbf{1}_{\boldsymbol{k}}^\top\boldsymbol{\Sigma}^{-\mathbf{1}})/(\mathbf{1}_{\boldsymbol{k}}^\top\boldsymbol{\Sigma}^{-\mathbf{1}}\mathbf{1}_{\boldsymbol{k}})$. Here $\mathbf{1}_k$ is a $k \times 1$ vector with all elements being 1. Find the distribution of $\boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{x}$.

3. (**20 marks**) Consider the model

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, 2, 3, 4, \quad j = 1, 2, 3, 4,$$

where ϵ_{ij} are independently distributed as $N(0, \sigma^2)$.

(a) Let $\boldsymbol{\beta} = (\mu, \tau_1, \tau_2, \tau_3, \tau_4)^\top$. Find a set of 4 linearly independent estimable functions of $\boldsymbol{\beta}$.

(b) Derive a test to test the null hypothesis $H_0 : \tau_1 - \tau_2 = \tau_3 - \tau_4$.

(c) Is $2\tau_1 + \tau_2$ estimable? Why?

4. (**30 marks**) Consider a linear regression model

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots x_{ip}\beta_p + \epsilon_i = \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i, \quad i = 1, \dots, n. \tag{1}$$

By convention, the response and covariates are centered and standardized. The ridge regression is to apply squared penalty on the least square estimate by minimizing

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2,$$

where $\lambda \geq 0$ is a tuning parameter, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$. The resulting estimate is denoted by $\hat{\boldsymbol{\beta}}^{\text{ridge}}$.

(a) Show that $\|\hat{\boldsymbol{\beta}}^{\text{ridge}}\|$ increases as the tuning parameter $\lambda \to 0$.

(b) Zou and Hastie (2005) introduced the *elastic-net penalty*

$$\lambda \sum_{j=1}^{p} (\alpha\beta_j^2 + (1 - \alpha)|\beta_j|),$$

a different compromise between ridge and lasso with $0 \leq \alpha \leq 1$. Consider the elastic-net optimization problem

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda[\alpha\|\boldsymbol{\beta}\|_2^2 + (1 - \alpha)\|\boldsymbol{\beta}\|_1],$$

where $\boldsymbol{y} = (y_1, y_2, \dots, y_n)^\top$, $\boldsymbol{X} = (x_1, x_2, \dots, x_p)_{n \times p}$, $\|\cdot\|_2$ and $\|\cdot\|_1$ are the L_2 norm and L_1 norm, respectively. For any $\alpha \in [0, 1]$, show how one can turn this into a lasso problem.

(c) The novel *fused lasso penalty* is of the following form

$$\lambda_F \sum_{j=2}^{p} |\beta_j - \beta_{j-1}|,$$

which is to penalize the sum of the absolute differences all pairs of successive regression coefficients. For model (1), one can estimate the regression parameters by minimizing

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda_F \sum_{j=2}^{p} |\beta_j - \beta_{j-1}|, \tag{2}$$

where λ_F is the tuning parameter. Show the fused lasso optimization problem in (2) can be turned into a lasso problem.