STAT5030 Assignment 5

Due: April 21, 2023

1. 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

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

where $\lambda \geq 0$ is a tuning parameter, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$. The resulting estimate is denoted by $\hat{\boldsymbol{\beta}}^{\text{ridge}}$. Without assuming the design matrix $\boldsymbol{X} = (x_1, \dots, x_p)$ is of full rank,

   (a) prove that β^{ridge} is a biased estimator for $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ for given tuning parameter λ;

   (b) find the bias and the variance of $\hat{\beta}^{\text{ridge}}$ for given tuning parameter λ;

   (c) show that $\left\| \hat{\beta}^{\text{ridge}} \right\|$ increases as the tuning parameter $\lambda \to 0$.

2. For model (1), Zou and Hastie (2005) introduced the elastic-net penalty

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

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

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

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.

   (a) Show how the elastic-net optimization problem can turn this into a lasso problem.

   (b) Provide your own understanding about the effect of the elastic-net penalty on the parameter estimates.

3. Show that the smallest value of λ such that the regression coefficients estimated by the lasso are all equal to zero is given by

$$\lambda_{\max} = \max_j \left| \frac{1}{N} \langle \mathbf{x}_j, \mathbf{y} \rangle \right|.$$

4. We consider the model

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \gamma_i + \epsilon_i$$

with $\epsilon_i \sim N\left(0, \sigma^2\right)$ and $\gamma_i, i = 1, 2, \ldots, N$ are unknown constants. Let $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_N)$ and consider minimization of

$$\min_{\beta \in \mathbb{R}^p, \gamma \in \mathbb{R}^N} \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j - \gamma_i \right)^2 + \lambda \sum_{1}^{N} |\gamma_i| \,. \tag{2}$$

(a) Show this problem is jointly convex in β and γ.

(b) Consider Huber's loss function

$$\rho(t; \lambda) = \begin{cases} \lambda|t| - \lambda^2/2 & \text{if } |t| > \lambda \\ t^2/2 & \text{if } |t| \leq \lambda \end{cases} \tag{3}$$

This is a tapered squared-error loss; it is quadratic for $|t| \leq \lambda$ but linear outside of that range, to reduce the effect of outliers on the estimation of β. With the scale parameter σ fixed at one, Huber's robust regression method solves

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{N} \rho \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j ; \lambda \right) . \tag{4}$$

Show that problems (2) and (4) have the same solutions $\widehat{\beta}$.