STAT 542 HW2

Xiruo Li(xiruoli2)

Question 1

(a)(b)(c)

Question 1

(1) $$\frac{\partial f(\beta, \beta_0)}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\left\{\frac{1}{2n}\|y - X\beta - \beta_0 \cdot \mathbf{1}\|^2 + \lambda \sum_{j=1}^{p}|\beta_j|\right\}$$

$$= \frac{1}{2n}\frac{\partial}{\partial \beta_0}\langle y - X\beta - \beta_0 \cdot \mathbf{1},\ y - X\beta - \beta_0 \mathbf{1}\rangle$$

$$= \frac{1}{2n}\left(-\sum_{i=1}^{n} y_i + \sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}\beta_j + n\beta_0\right) = 0$$

$$\Rightarrow \beta_0 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)$$

(2) $$\frac{\partial f(\beta, \beta_0)}{\partial \beta_j} = \frac{\partial}{\partial \beta_j}\left\{\frac{1}{2n}\|y - X\beta - \beta_0 \cdot \mathbf{1}\|^2 + \lambda \sum_{j=1}^{p}|\beta_j|\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(-x_{ij})\left(y_i - \beta_0 - \sum_{k=1}^{p} x_{ik}\beta_k\right) + \begin{cases}\lambda, & \beta_j > 0 \\ -\lambda, & \beta_j < 0.\end{cases}$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}y_i - \beta_0 x_{ij} - \sum_{k=1}^{p} x_{ij}x_{ik}\beta_k\right) \pm \lambda$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}y_i - \beta_0 x_{ij} + \sum_{\substack{k=1 \\ k \neq j}}^{p} x_{ij}x_{ik}\beta_k\right) \pm \lambda$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}^2\beta_j\right) = 0$$

$$\Rightarrow \beta_j = \frac{1}{\sum_{i=1}^{n} x_{ij}^2}\left\{\sum_{i=1}^{n}\left(x_{ij}y_i - \beta_0 x_{ij} - \sum_{k=1, k\neq j}^{p} x_{ij}x_{ik}\beta_k\right) \begin{matrix} -\lambda n, & \beta_j > 0 \\ +\lambda n, & \beta_j < 0 \end{matrix}\right\}$$

let $A = \sum_{i=1}^{n}\left(x_{ij}y_i - \beta_0 x_{ij} - \sum_{k=1, k\neq j}^{p} x_{ij}x_{ik}\beta_k\right)$, $B = \sum_{i=1}^{n} x_{ij}^2$.

$$\beta_k = \begin{cases}\frac{A - \lambda n}{B}, & A > \lambda n \\ 0, & -\lambda n \le A \le \lambda n \\ \frac{A + \lambda n}{B}, & A < -\lambda n\end{cases}$$

(3) Let $\beta_1 \cdots \beta_p = 0$, $\beta_0 = \bar{y}$

$$|\lambda \cdot n| = A$$

$$\lambda_{max} = \max_{j=1,\cdots,p} \left| \frac{\sum_{i=1}^{n} (x_{ij} y_i - x_{ij} \bar{y})}{n} \right|$$

(d)

I use the beta of different lambda from the training data to predict Y in the test data.

Then I got the prediction errors of different lambda. To be the best lambda, the error should be the minimum. Thus, I got the best lambda=0.05372383.

Question 2

(a)

Number of nonzero parameters: 61(including intercept)

Best lambda: 0.04173215

(b)

Estimated degree of freedom: 60.31393, which is close to the number of non-zero parameter in part a.

(c)

Number of nonzero parameters: 201(including intercept)

Best lambda: 29.291

theoretical value of the degrees of freedom: 170.1506, from the following formula:

$$\mathrm{df}(\lambda) = \mathrm{Trace}\left(\mathbf{X}(\mathbf{X}^{\mathsf{T}}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^{\mathsf{T}}\right) = \sum_{j=1}^{p} \frac{d_j^2}{d_j^2 + \lambda}$$

Estimated degree of freedom: 173.9836, which matches the theoretical value