P9120- Homework # 1

Assigned: September 19, 2019 Due: October 3, 2019 (in class)

Maximum points that you can score in this Homework is 20.

Please include all R code you used to complete this homework.

1. Let $\mathbf{X}$ denote an $n \times p$ matrix with each row an input vector and $\mathbf{y}$ denote an n-dimensional vector of the output in the training set. For fixed $q \geq 1$, define

$$\text{Bridge}_\lambda(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j|^q$$

for $\lambda > 0$. Denote the minimal value of the penalty function over the least squares solution set by

$$t_0 = \min_{\boldsymbol{\beta}:\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}=\mathbf{X}^T\mathbf{y}} \sum_{j=1}^{p} |\beta_j|^q.$$

(a) Show that $\text{Bridge}_\lambda(\boldsymbol{\beta})$ for $\lambda > 0$ is a convex function in $\boldsymbol{\beta}$, which is strictly convex for $q > 1$.

(b) Show that for $q > 1$ there is a unique minimizer, $\hat{\boldsymbol{\beta}}(\lambda)$, with $\sum_{j=1}^{p} |\hat{\beta}_j(\lambda)|^q \leq t_0$.

(c) Show that for $q = 1$ there exists a minimizer and for all minimizers, $\hat{\boldsymbol{\beta}}(\lambda)$, the penalty function takes the same value

$$s(\lambda) \triangleq \sum_{j=1}^{p} |\hat{\beta}_j(\lambda)|^q \leq t_0.$$

Thus for $q \geq 1$, $s(\lambda)$ is well defined as a function of λ on the interval $(0, \infty)$.

(d) Show that minimizing $\text{Bridge}_\lambda(\boldsymbol{\beta})$ is equivalent to minimizing

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad \text{subject to } \sum_{j=1}^{p} |\beta_j|^q \leq s(\lambda).$$

2. Exercise 1 (Section 6.8) of [ISL] (p. 259).

3. Exercise 3.16 of [ESL] (p. 96).

4. The prostate data described in Chapter 3 of [ESL] have been divided into a training set of size 67 and a test set of size 30. Carry out the following analyses on the training set:

   (a) Best-subset linear regression with k chosen by 5-fold cross-validation.

   (b) Best-subset linear regression with k chosen by BIC.

   (c) Lasso regression with λ chosen by 5-fold cross-validation.

   (d) Lasso regression with λ chosen by BIC.

   (e) Principle component regression with q chosen by 5-fold cross-validation.

   For each analysis, compute and plot the cross-validation or BIC estimates of the prediction error as the model complexity increases as in Figure 3.7. Report the final estimated model as well as the test error and its standard error over the test set as in Table 3.3. Briefly discuss your results.