Tutorial 2

1. If $\mathbb{X}$ is orthonormal, then $\mathbb{X}^\top\mathbb{X} = nI_p$, the optimal choice of λ minimizing $E\|\hat{\beta}_{ridge} - \beta\|^2$ and the expected prediction error is:

$$\lambda^* = \frac{p\sigma^2}{n\sum_{j=1}^{p}\beta_j^2}$$

where $(\beta_1, \beta_2, ..., \beta_p)$ is the true coefficient vector

2. For the simple linear regression model,

$$y_i = \beta_0 + \beta_1\mathbf{x}_i + \varepsilon_i, \quad i = 1, ..., n$$

Suppose we use the weighted least square estimation and minimize

$$\sum_{i=1}^{n} w_i(Y_i - \beta_0 - \beta_1\mathbf{x}_i)^2$$

Find the weighted estimator of β_0 and β_1.

3. Suppose $\sum_{i=1}^{n} \mathbf{x}_{ik}^2 = c_k, k = 1, ..., p$ and $\sum_{i=1}^{n} \mathbf{x}_{ik}\mathbf{x}_{i\ell} = 0$ if $k \neq \ell$. For model

$$Y_i = \beta_1 \mathbf{x}_{i1} + ... + \beta_k \mathbf{x}_{ip} + \varepsilon_i,$$

prove that the ridge regression estimator and the least squares estimator have the following relation

$$\hat{\beta}_{ridge} = \begin{pmatrix} c_1/(c_1+\lambda) & 0 & 0 & ... & 0 \\ 0 & c_2/(c_2+\lambda) & 0 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & ... & c_p/(c_p+\lambda) \end{pmatrix} \beta_{LSE}$$

4. For a fixed λ, please find the solution to

$$\min_{\beta} \left\{ \sum_{i=1}^{n} w_i (Y_i - \beta_0 - \beta_1 \mathbf{x}_{i1} - ... - \beta_p \mathbf{x}_{ip})^2 + \lambda(\beta_1^2 + ... + \beta_p^2) \right\}$$

5. For Data A, the first 50 columns are the predictors and the last the response.

Please fit a linear regression model to the data and use the ridge regression method to estimate the model.

(a) Please calculate the estimators of the coefficients with $\lambda = 0.1$, 1, 10, 100, 1000 respectively, and plot how the estimators change with λ.

(b) Use leave-one-out CV to choose the tuning parameter λ.

(c) Use the λ selected based on the above CV to estimate the model, predict the response in Data B, and calculate the mean prediction error square.

(d) Select λ by AIC and BIC, and 5-fold CV respectively [Hint: when AIC and BIC are used, please replace p in the penalties by $trace\{\mathbb{X}(\mathbb{X}^\top\mathbb{X} + \lambda I)^{-1}\mathbb{X}^\top\}$]