Stat/CS 5525 Homework 2

(Due on September 28th, 2023)

1. For the ridge regression, show that

$$\hat{\beta}^{ridge} = (\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}'(\boldsymbol{X}\boldsymbol{X}' + \lambda \boldsymbol{I})^{-1}\boldsymbol{y}.$$

2. Ex. 3.6 from the textbook.

3. **Grocery retailer.** A large, national grocery retailer tracks productivity and costs of its facilities closely. Data were obtained from a single distribution center for a one-year period. Each data point for each variable represents one week of activity. The variables included are the number of cases shipped (X_1), the indirect costs of the total labor hours as a percentage (X_2), a qualitative predictor called holiday that is coded 1 if the week has a holiday and 0 otherwise (X_3), and the total labor hours (Y). (Data file: Grocery.txt. The columns from left to right are Y, X_1, X_2, X_3.)

   (a) Fit the linear model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$. Show the table of the fitted model: coefficients estimation, their standard deviation, z-score, and p-values. Show the estimation $\hat{\sigma}^2$.

   (b) Use both best subset (C_p, forward variable selection, and backward variable selection to find a smaller model with the best fits. Show the results of the fits as in (a).

   (c) Use F-test to check whether the model returned from (b) significantly different ($\alpha = 0.05$) from the complete model in (a).

4. In Problem 3, add five more predictor variables $Z_1 = X_1X_2, Z_2 = X_1X_3, Z_3 = X_2X_3, Z_4 \sim N(30, 30), Z_5 \sim N(7, 1)$. Use the Lasso to analyze the data, and interpret your result.