Spring 2022

Quantitative Analysis
Chung-Ming Kuan
Jui-Chung Yang

Problem Set 8

Due: 5/2

Part One: Hand-Written Exercise

1. Suppose we estimate the regression coefficients in a linear regression model by minimizing

$$\min_{\beta} \text{RSS} = \min_{\beta} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 + \lambda \beta_1^2.$$

   Please derive the estimator of coefficients, $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively.

2. For a data set $(y_i, \mathbf{x}_i)_{i=1}^n$, where y_i is a scalar and $\mathbf{x}_i$ a $p \times 1$ column vector. That is, the regression model is $y_i = \sum_{j=1}^p \beta_j x_{ij} + u_i$. Please show that the Ridge Regression estimator $\hat{\beta}_R$ is given by:

$$\hat{\beta}_R = \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' + \lambda \mathbf{I} \right)^{-1} \left(\sum_{i=1}^{n} \mathbf{x}_i y_i \right),$$

   where λ is the tuning parameter and $\mathbf{I}$ is the p-dimensional identity matrix.

3. We perform best subset, forward stepwise, and backward stepwise selection on a single data set. For each approach, we obtain $p + 1$ models, containing 0, 1, 2, ..., p predictors. For $k = 1, .., p$, please answer the following questions and justify your answers:

   (a) Which of the three models with k predictors has the smallest training RSS?

   (b) Which of the three models with k predictors has the smallest testing MSE?

   (c) (True or False) The predictors in the k-variable model identified by forward stepwise are a subset of the predictors in the $(k + 1)$-variable model identified by forward stepwise selection.

   (d) (True or False) The predictors in the k-variable model identified by backward stepwise are a subset of the predictors in the $(k+1)$-variable model identified by backward stepwise selection.

Part Two: Computer Exercise

Consider the simulation settings on slide 24, Lecture 7. The data generating process (DGPs) are

$$\text{DGP 1}: y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i, \quad \beta_0 = 2,\ \beta_1 = \beta_2 = 1,$$
$$\text{DGP 2}: y_i = \beta_0 + \sum_{j=1}^{20} \beta_j x_{ij} + \epsilon_i, \qquad \beta_0 = 2,\ \beta_1 = \cdots = \beta_{20} = 1,$$

where all x_{ij} and ϵ_i are i.i.d. $\mathcal{N}(0,1)$, and the sample size is $n = 55$. For each DGP, we estimate two models with 25 and 50 regressors, including the "signal" variables in the DGP and other "noise" variables (generated as i.i.d. $\mathcal{N}(0,1)$). Consider 100 values in $[0,2]$, with the optimal determined by 10-fold CV.

1. Plot the lines of ridge estimates averaged over 100 replications. Combine the 4 graphs on a single plot and place them as 2×2.

2. Attach legends properly on the 4 graphs respectively.

Note that this exercise replicates the figures on slides 25 - 28, Lecture 7. Your numbers should be close to those on the slides but not necessarily the same.