Spring 2022

Quantitative Analysis
Chung-Ming Kuan
Jui-Chung Yang

Problem Set 7

Due: 4/25

Part One: Hand-Written Exercise

1. Please verify the fact that

$$\left(\mathbf{A}-\mathbf{x}_i\mathbf{x}_i'\right)\left(\mathbf{A}^{-1}+\frac{\mathbf{A}^{-1}\mathbf{x}_i\mathbf{x}_i'\mathbf{A}^{-1}}{1-\mathbf{x}_i'\mathbf{A}^{-1}\mathbf{x}_i}\right)=\mathbf{I}$$

for any invertible matrix $\mathbf{A}$ that has the same dimension as $\mathbf{x}_i\mathbf{x}_i'$.

2. Suppose that we obtain a bootstrap sample from a set of N observations.

   (a) For $i = 1, ..., N$ and $j = 1, ..., N$, what is the probability that the ith bootstrap observation is *not* the jth observation from the original samples? Does your answer depend on i or j? Justify your answer.

   (b) What is the probability that the jth observation from the original samples is *not* in the N bootstrap samples? Justify your answer.

   (c) Continue with part (b), calculate the probability for $N = 5$ and $N = 5000$.

   (d) Continue with part (b), calculate the probability when $N \to \infty$.

Part Two: Computer Exercise

1. Please load the data set `Auto` from the package `ISLR`. `Auto` contains gas mileage, horsepower, and other information for 392 vehicles. Suppose we have three competing models:

Model 1: $\texttt{mpg} = \beta_0 + \beta_1\texttt{horsepower} + u$

Model 2: $\texttt{mpg} = \beta_0 + \beta_1\texttt{horsepower} + \beta_2\texttt{weight} + u$

Model 3: $\texttt{mpg} = \beta_0 + \beta_1\texttt{horsepower} + \beta_2\texttt{weight} + \beta_3\texttt{acceleration} + u.$

Complete the following questions by setting the random seed to `set.seed(1)`:

   (a) Please choose the best model using the validation set approach and estimate its testing MSE.

   (b) Please choose the best model using LOOCV and estimate its testing MSE.

   (c) Please choose the best model using 10-fold CV and estimate its testing MSE.

2. For the simple linear regression model:

$$\texttt{mpg} = \beta_0 + \beta_1 \texttt{horsepower} + u,$$

please obtain the OLS estimator $\hat{\beta}_1$, and construct the following bootstrap estimators of $\mathrm{SD}(\hat{\beta}_1)$ using $B = 1000$ simulations.

   (a) Compute the "Paired Bootstrap" estimator of $\mathrm{SD}(\hat{\beta}_1)$ without the function `boot()`.

   (b) Compute the "Residual Bootstrap" estimator of $\mathrm{SD}(\hat{\beta}_1)$.