DSA211 Statistical Learning with R

Homework 7

Use R functions and data files (Carseats in ISLR package and Boston in MASS package) to solve the following problems:

1. Based on the information of Sales and Price in the data file Carseats and set the random seed to 123,

   (a) By using the given data set as training set, fit the polynomial regression models for polynomials of order $i = 1$ to $i = 4$ that use Price to predict Sales.

   (b) Identify which polynomial regression model is the best.

   (c) Using the Leave-One-Out cross validation approach, fit the polynomial regression models for polynomials of order $i = 1$ to $i = 4$ that use Price to predict Sales and compute the cross-validation errors for each model.

   (d) Based on the results in part (c), determine which polynomial regression model should be used and the model estimates.

   (e) Using the 10-fold cross-validation, fit the polynomial regression models for polynomials of order $i = 1$ to $i = 4$ that use Price to predict Sales and compute the associated cross-validation errors for each model.

   (f) Based on the results in part (e), determine which polynomial regression model should be used and the model estimates.

2. Based on the information of medv in the data file Boston and set the random seed to 456,

   (a) Provide an estimate for the population mean of medv (denoted μ). Call this estimate $\hat{\mu}$.

   (b) Estimate of the standard error of $\hat{\mu}$ using classical inference approach.

   (c) Construct a 95% confidence interval for μ, based on the part (b).

   (d) Estimate μ and the standard error of $\hat{\mu}$ using the bootstrap with 10,000 replicates.

   (e) Construct a 95% Bootstrap Percentile confidence interval for μ.

   (f) Using the bootstrap with 10,000 replicates, provide a bootstrap estimate, $\hat{\theta}$, for the inter-quantile range θ of medv in the population.

   (g) Estimate the standard error of $\hat{\theta}$.

   (h) Construct a 95% Bootstrap Percentile confidence interval for θ.

-END-