EPPS6323 Knowledge Mining

Assignment 7

1. Lab09 in R
2. Review ISLR Chapters 6 and look up answers for the following questions
3. From the three methods (best subset, forward stepwise, and backward stepwise):
   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 test RSS?
4. Application exercise:

Generate simulated data, and then use this data to perform best subset selection.

1. Use the rnorm() function to generate a predictor X of length n = 100, as well as a noise vector ε of length n = 100.

   Hint:

   ```
   set.seed(1)
   X = rnorm(100)
   eps = rnorm(100)
   ```

2. Generate a response vector Y of length n = 100 according to the model:

   $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon,$

   where β_0, β_1, β_2, and β_3 are 4, 9, 2, 1 respectively.
3. Use the regsubsets() function to perform best subset selection in order to choose the best model containing the predictors X, $X^2 \ldots X^{10}$. What is the best model obtained according to Cp, BIC, and adjusted R^2? Show some plots to provide evidence for your answer, and report the coefficients of the best model obtained. Note you will need to use the data.frame() function to create a single data set containing both X and Y .
4. Repeat 3, using forward stepwise selection and also using back- wards stepwise selection. How does your answer compare to the results in 3?