DSA211 Statistical Learning with R

Homework 8

Use R functions and data file (Credit in ISLR Package) to solve the following problem:

1. Based on the information in the Credit file,

   (a) Fit the multiple regression equation to predict Balance with all 10 independent variables (note that ID is not a variable) by using the Best Subset Selection with BIC criterion on the training data set. Plot the graph to show the number of variables versus BIC in each selection step.

   (b) Fit the multiple regression equation to predict Balance with all 10 independent variables (note that ID is not a variable) by using the Forward Selection with C_p criterion on the training data set. Plot the graph to show the number of variables versus C_p in each selection step.

   (c) Fit the multiple regression equation to predict Balance with all 10 independent variables (note that ID is not a variable) by using the Backward Selection with adjusted R^2 criterion on the training data set. Plot the graph to show the number of variables versus adjusted R^2 in each selection step.

   (d) Under the validation approach, fit the multiple regression equation to predict Balance with all 10 independent variables (note that ID is not a variable) by using the Best Subset Selection with mean square error criterion on the validation data set. Plot the graph to show the number of variables versus mean square error in each selection step. Use the set.seed(121).

-END-