1. **Model Selection**

   Please refer to data in the "Boston" is included in MASS package, where *medv* is the response and the rest as predictors.

   You are asked to use stepwise regression algorithm to help you select a model. You will want to use leaps::regsubsets() to perform the task.

   (a) Perform the best subset selection. Show the first $1, 2, \ldots, 6$ variables have been selected by the best subset selection respectively. What is the model the c_p, BIC, and R^2_{adj} selected respectively.

   (b) Similar to (a), you are asked to perform forward and backward selection. Show also the first $1, 2, \ldots, 6$ variables have been selected by the forward and the backward selection respectively.

   (c) Compare (the variables and coefficients for) the selections (a) and (b) above.

   (d) You are asked to choose among models using cross validation approach. In this exercise, we consider the variables selected using the best subset selection algorithm. And use $10-$folds CV to choose the among the models selected with the best subset selection method.

      i First, you will create a vector that allocates each observation to one of $k = 10$ folds using sample() function, and also create a matrix in which you will store the results.

      ii Write a for() loop for the prediction and store the results in the matrix you created in (d.i). (Please reference James Lab page249-250 section 6.5.3 Choosing Among Models using Validation Set Approach.)

      iii Compute the mean squared error vector for each model by using apply() function. Plot the mse results. What is the best model chosen by the $10-$folds CV?

      iv Show the coefficients for the best model.

2. **Feature Selection and Model Selection**

   You are given SSR of eight models on 10 observations. They are based ten data point and total of three predictors. See Table 1 below for Question 2.

   Please perform the follow:

   (a) subset selection

      - What is the two variables that forward stepwise selected?
      - what is the one variables that the backward stepwise selected?
      - what is the two variables that the best subset selected?

   (b) Perform the following Model selection?

      - What is the model that Mallow's, C_p and Akaike Information Criterion, AIC selected under the best subset selection?

Model	Predictors	SSR
1	None	25.00
2	x1	9.50
3	x2	18.00
4	x3	15.00
5	x1,x2	8.25
6	x1,x3	6.25
7	x2,x3	5.06
8	x1,x2,x3	3.05

Table 1: SSR for eight models based on 10 observations

- What is the model that Bayesian Information Criterion, BIC selected under the forward stepwise selection?
- What is the model that Adjusted R^2, R^2_{Adj} selected under the backward stepwise selection?