1. **Cross Validation**

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

   You are asked to use Cross Validation approach to estimate the test error rates that results from fitting various linear models on the Boston data set. You will want to use boot::cv.glm() function.

   (a) Compute the LOOCV mean squared error (MSE) for polynomial of any degree m. (write a function using for loop). And show the results m from 1 to 10. What is the model LOOCV chosen?

   (b) Compute now 10-folds MSE for the polynomial degree of m. And show the results m from 1 to 10. What is the model 10-folds CV chosen?

   (c) Compare the results between (a) and (b) above.

2. **Shrinkage Method - Ridge Regression**
   Please refer to data in the "Hitters" is included in ISLR package, where *Salary* is the response and the rest are predictors.

   You decide to randomly sample the data to form a training set, and fit a **Ridge regression** on the training set. The tuning parameter in the Ridge regression is determined by a 10-fold cross-validation. When estimating the tuning parameter, let's set.seed(1) for cv.glmnet() so that we are on the same page. **Ridge regression** is minimizing

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2 \tag{1}$$

   (a) Please vary the size of the training set to be from 50% to 100% (in 1/10 increments) of the dataset. Let's set.seed(1) to form our training set through random sampling so we are on the same page.

      i. For each of these 9 training sets of various sizes, please estimate the optimal tuning parameter for the ridge regression via a 10-fold cross-validation and then use the optimal lambda.1se to predict in the test data set and then to calculate the mse. When estimating the tuning parameter, let's set.seed(1) for cv.glmnet() so that we are on the same page.

      ii. Plot the optimal tuning parameter values against the training set size (as a % of the dataset).

      iii. plot the mse against the train set size (as a % of the dataset). What is the training set that minimizes the mse?

   (b) Please make the training set to be 85% of the dataset. However, when performing random sampling to for our training set, we vary the seeds from 1 to 10.

      i. For each of these 10 training sets of the same sizes, please estimate the optimal tuning parameter for the ridge regression via a 10-fold cross-validation. When estimating the tuning parameter, let's set.seed(1) for cv.glmnet() so that we are on the same page.

ii. Plot the optimal tuning parameter values against the seed values (1 to 10).

(c) What are your takeaways / observations?

3. **Shrinkage Method**

Please refer to data in the "Hitters" again, where *Salary* is the response and the rest are predictors.

You decide to randomly sample 2/3 of the data as training set, and use different alpha to fit linear regressions on the training set. You are asked to obtain the mean squared error (mse) error corresponding to the optimal lambda.1se for varies α.

The tuning parameter in those regression are determined by a 10-fold cross-validation. When estimating the tuning parameter, let's set.seed(1) for cv.glmnet().

Elastic-net is minimizing

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\left[\alpha\sum_{j=1}^{p}|\beta_j| + (1-\alpha)\sum_{j=1}^{p}\beta_j^2\right] \tag{2}$$

(a) Please find the mse corresponding to $\alpha = i/20, i = 0, 1, \ldots, 20$.

(b) Please plot the fit of cv.glmnet when $\alpha = 1$.

(c) Which regression (or alpha) has the smallest mse?