Stat 6021: Guided Question Set 11

For this guided question set, we will use the "nfl.txt" data set that we used in module 7. As a reminder, the data are on NFL team performance from the 1976 season. The variables are:

- y: Games won (14-game season)
- x_1: Rushing yards (season)
- x_2: Passing yards (season)
- x_3: Punting average (yards/punt)
- x_4: Field goal percentage (FGs made/FGs attempted)
- x_5: Turnover differential (turnovers acquired - turnovers lost)
- x_6: Penalty yards (season)
- x_7: Percent rushing (rushing plays/total plays)
- x_8: Opponents' rushing yards (season)
- x_9: Opponents' passing yards (season)

We will use the predictors x_2, x_7, and x_8, which we found to have the lowest BIC from module 7. For this guided question set, your group is to write your own code to carry out Leave-One-Out Cross-Validation and to provide an estimate for the test MSE for this model. Be sure to submit your R script containing the code (comments should be added in the code when appropriate), as well as the estimated test MSE for this model for HW 11.

While there exist functions to carry out LOOCV, you are tasked to write your own code for this activity. Writing your code is a good way to test if you really understand all the steps in the process involved with a statistical method or algorithm. You can then compare your code with existing functions to verify if you wrote your code correctly.

The `cv.glm()` function from the `boot` library can carry out cross-validation. For example,

```
glm.fit<-glm(y~x2+x7+x8, data=data)
cv.err<-cv.glm(data, glm.fit)
cv.err$delta[1] ##the output for the LOOCV should match your own
cv.glm(data,glm.fit, K=10)$delta[1] ##k fold CV with k=10
```