Exercise 2.4 ; Extra Credit

Train RMSE

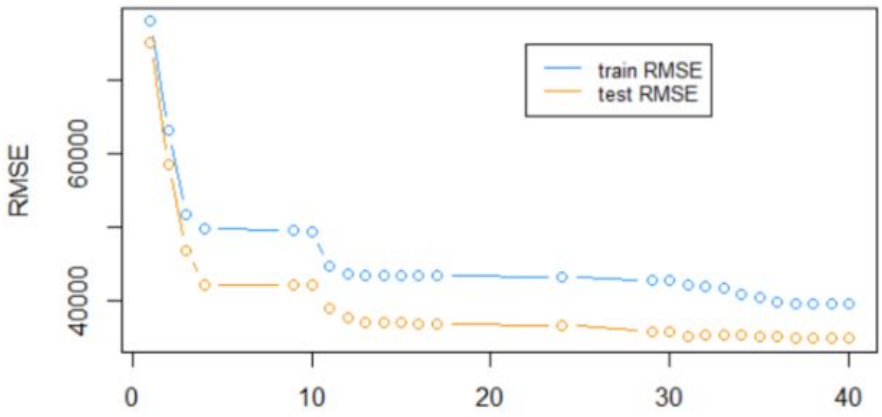
- Train data from initial data, i.e. training house data
- Data that used to make predictions and that is later compared to testing data for accuracy

Test RMSE

- Testing data from initial data, i.e. testing house data

Calculating RMSE

- Squaring the Residuals
- Finding the average of the residuals
- Taking the square root of the results



- Which number house the RMSE is being tested of

This model shows the RMSE based on the model size and how the train and test data compares to each other. In order to calculate the RMSE of each house, all the data of each house is collected ranging from how many car garage it may have and the Sale price of the home. Using that data, the residual is found. The residual is calculated by using the train data which is created by predictions and then subtracting that from the actual data(test data). You take that number and square it and find the averages of the residuals. After finding the averages, you square the the results to find the RMSE(Root Mean Square Deviation).

We wanted to get the lowest possible number for RMSE and you can see that on how the data slopes down more and more with the Model size increasing. We kept increasing the number of models that were being analyzed to decrease the RMSE. At the lowest RMSE, our coefficient of determination(R squared) equaled 0.78 which is 78%. Coefficient of determination explains how accurate our predictions are based on test data and tells how our variability of the prediction data matches up to the mean of the test data.