

Homework1_Q3

Regression vs KNN

Extract two trims(350 and 65 AMG) from all trim levels.

1.Split the data into training and testing set.

```
# define training and testing data
D1_train = sclass350[train_1,]
D1_test = sclass350[-train_1,]

D2_train = sclass65AMG[train_2,]
D2_test = sclass65AMG[-train_2,]

D1_test = arrange(D1_test, mileage)
D2_test = arrange(D2_test, mileage)
```

2.Run K-nearest-neighbors, for many different values of K, starting at K=3. For each value of K, fit the model to the training set and make predictions on your test set.

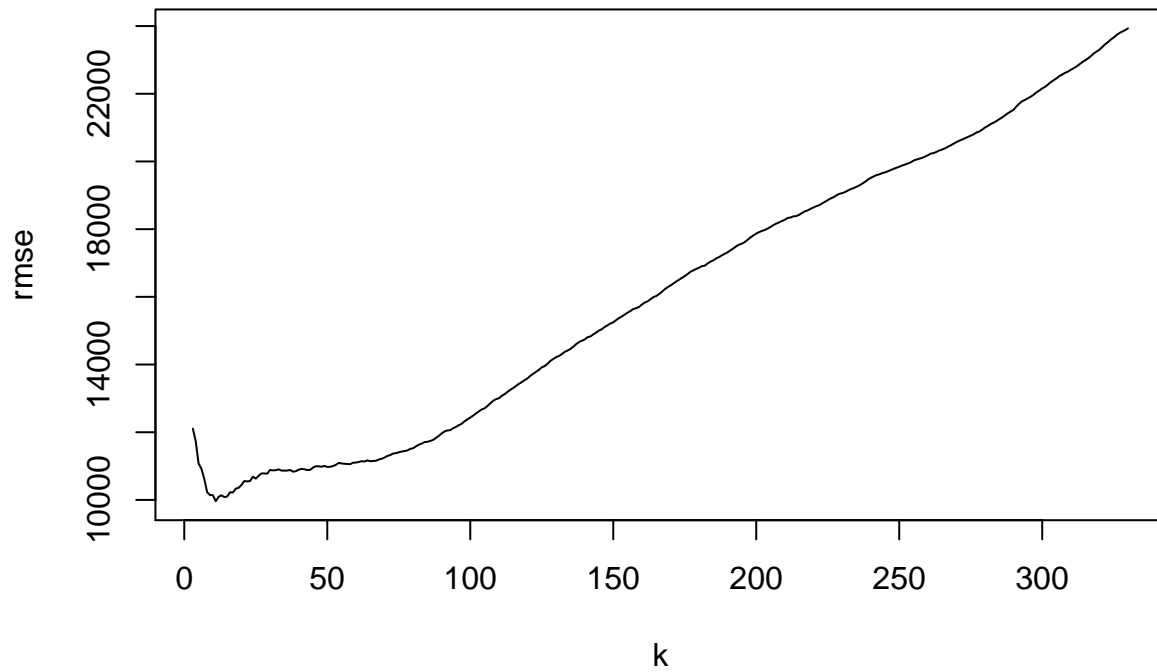
3.Calculate the out-of-sample root mean-squared error (RMSE) for each value of K.

```
rmse1.matrix= matrix(NA,nrow = 328, ncol = 2)
for (i in 3:330) {
  knn=knn.reg(train = X1_train, test = X1_test, y= y1_train, k=i)
  ypred_knn = knn$pred
  rmse(y1_test, ypred_knn)
  rmse1.matrix[i-2,2]=rmse(y1_test, ypred_knn)
  rmse1.matrix[i-2,1]=i
}

rmse2.matrix= matrix(NA,nrow = 228, ncol = 2)
for (i in 3:230) {
  knn=knn.reg(train = X2_train, test = X2_test, y= y2_train, k=i)
  ypred_knn = knn$pred
  rmse(y1_test, ypred_knn)
  rmse2.matrix[i-2,2]=rmse(y2_test, ypred_knn)
  rmse2.matrix[i-2,1]=i
}
```

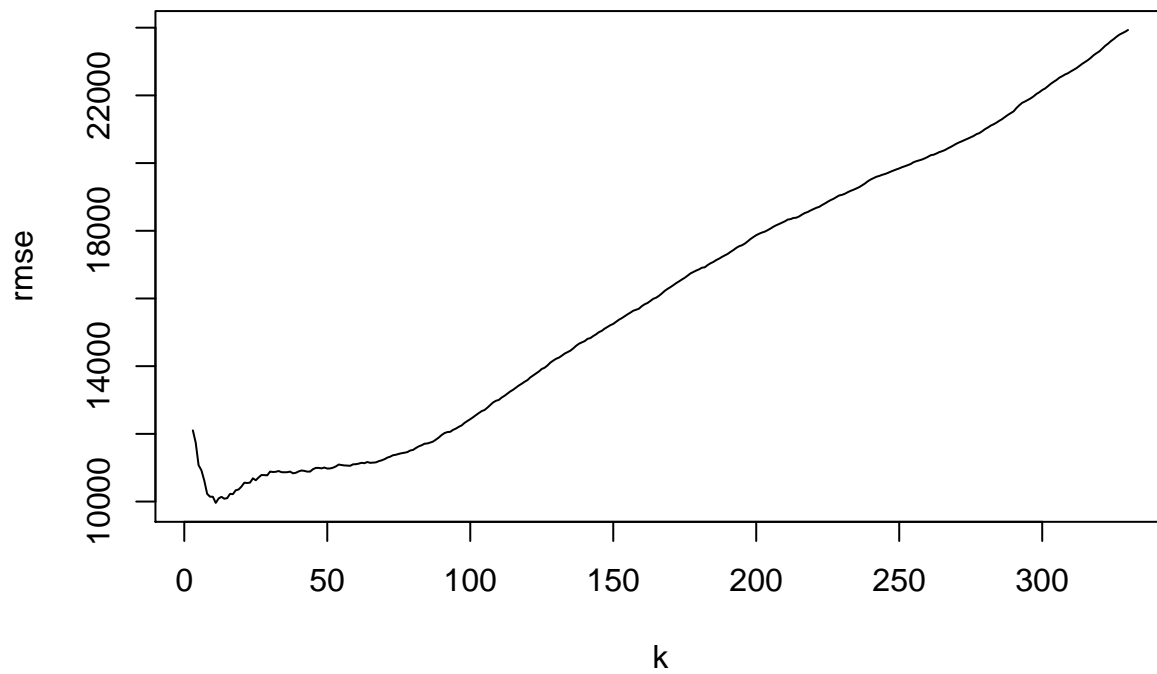
Make plot of RMSE versus K for trim 350

350



Make plot of RMSE versus K for trim 65AMG

65AMG



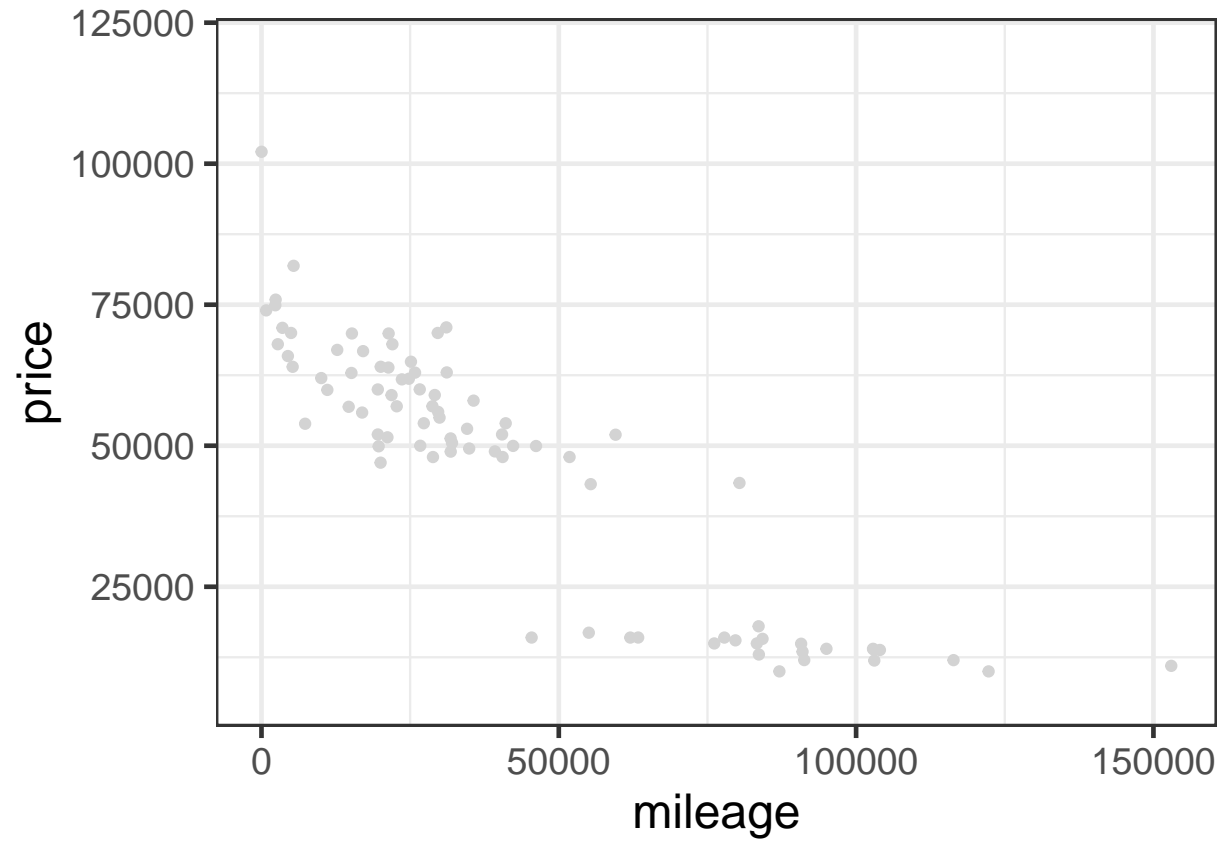
For the optimal value of K, the plot of fitted model for trim 350 The optimal value of K is below:

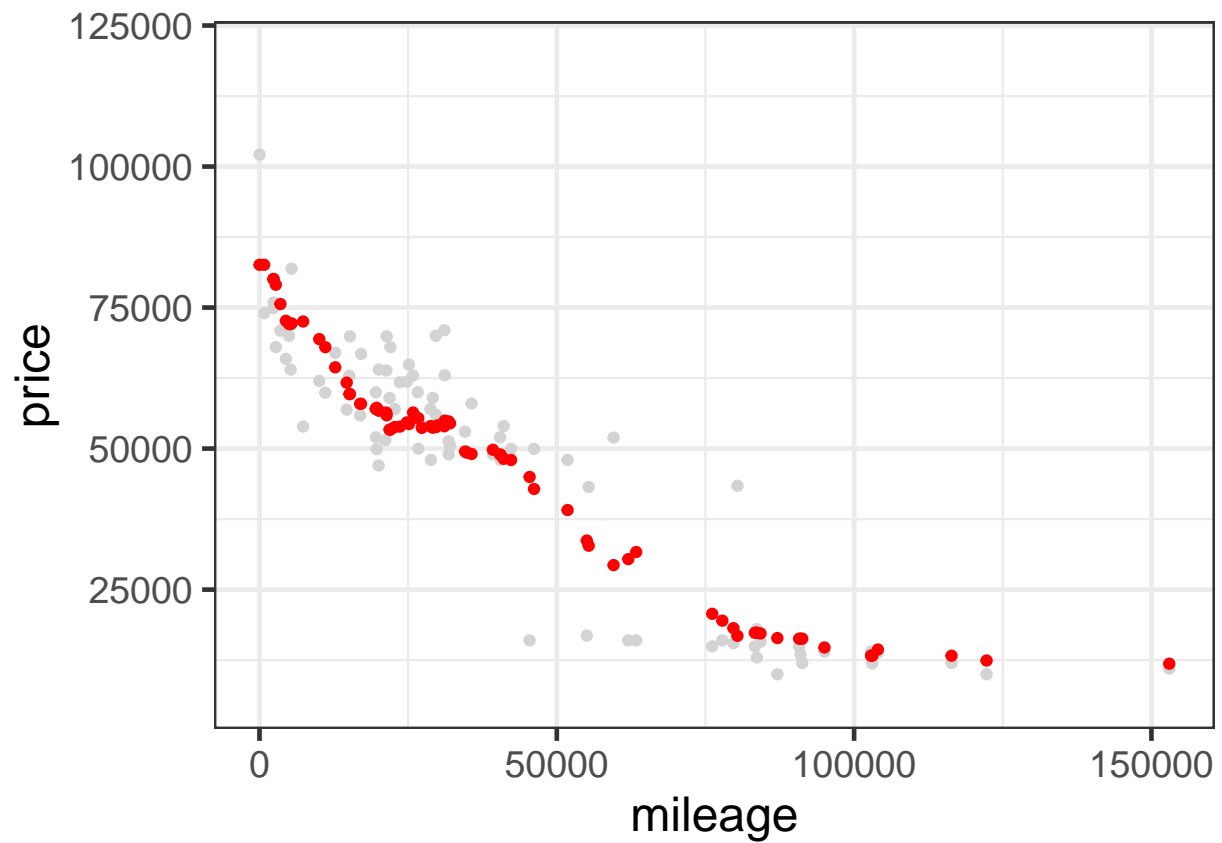
```
## [1] 24
```

The corresponding RMSE is below:

```
## [1] 8946.917
```

The plot of fitted model for trim 350





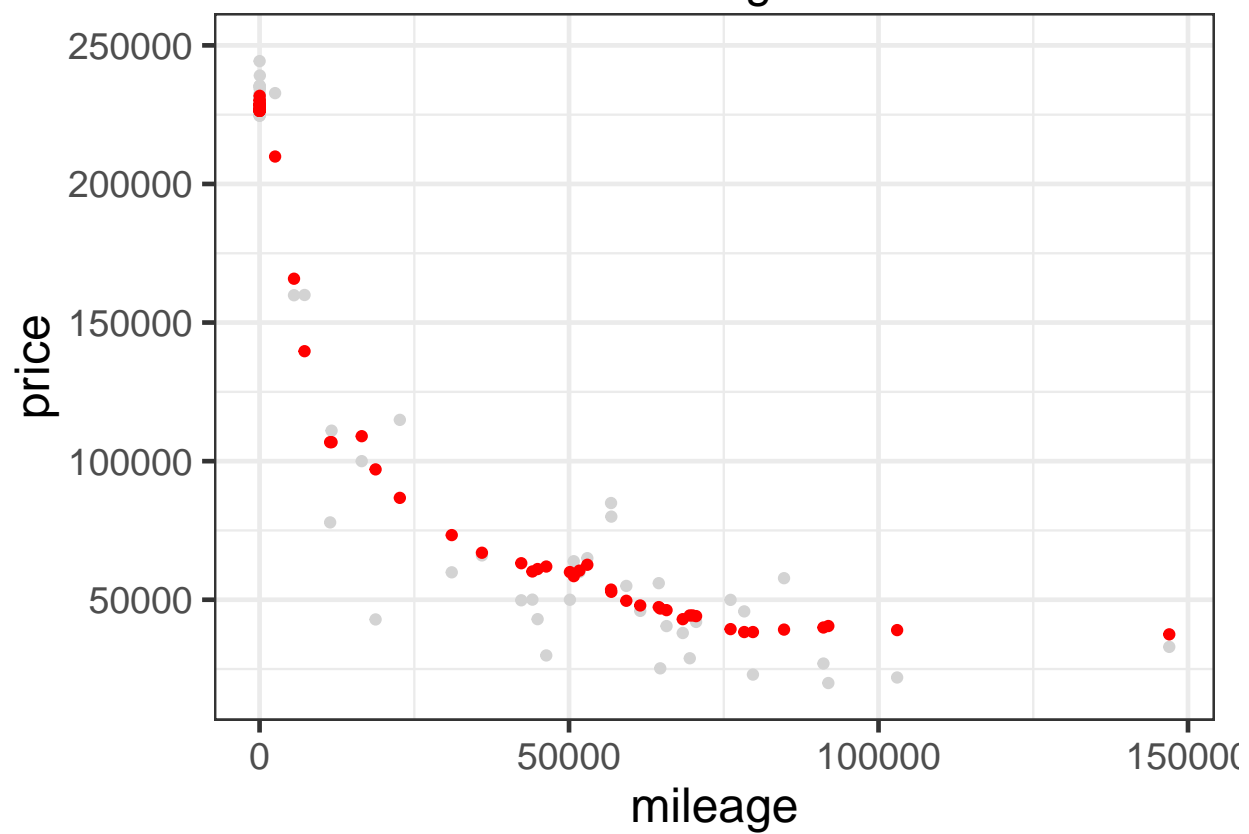
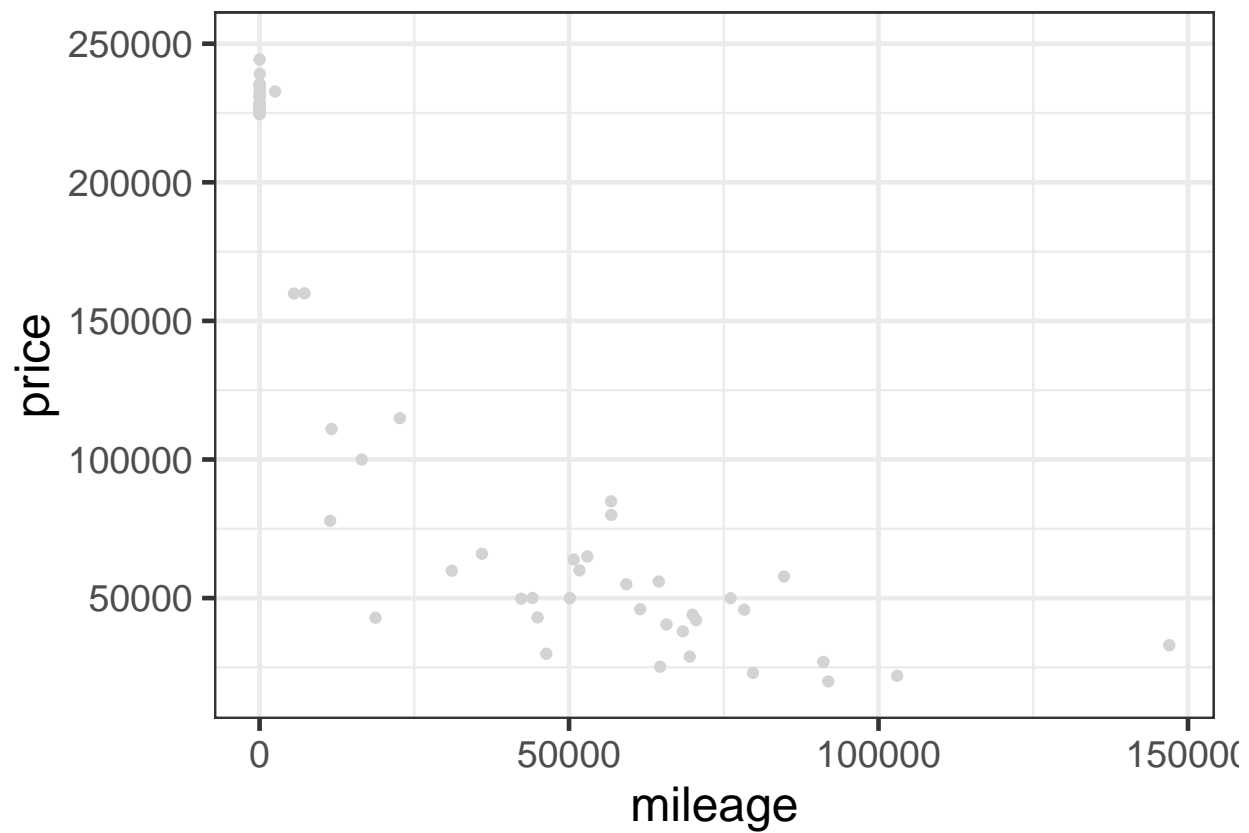
For the optimal value of K, the plot of fitted model for trim 65AMG The optimal value of K is below:

```
## [1] 17
```

The corresponding RMSE is below:

```
## [1] 14547.33
```

The plot of fitted model for trim 65AMG



Which trim yields a larger optimal value of K ? Why do you think this is?

Trim 350 yields larger optimal value of K . Because the sample size for trim 350 is larger than trim 65AMG. Then I guess the larger the sample size is, the larger the optimal K will be.