House Sales Subjective Assignment Questions

Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Optimal value for alpha

- Ridge = 4
- Lasso = .0001

```
########## Ridge R2 score #########
#        alpha     X_train     X_test
          4          .958         .878
########## Lasso R2 score #########
#         alpha     X_train    X_test
        .0001       .962       .879
```

<u>Top 5 coefficients of Ridge regression</u>

Neighborhood_Crawfor	0.115189
OverallCond_9	0.101665
OverallQual_9	0.094293
Exterior1st_BrkFace	0.083142
OverallCond_8	0.083087

<u>Top 5 coefficients of Lasso regression</u>

Neighborhood_Crawfor	0.156330
OverallQual_9	0.144760
Exterior1st_Stucco	0.141526
OverallCond_9	0.138120
OverallQual_10	0.116798

After double the alpha value for Ridge and Lasso Models

```
########## Ridge R2 score #########

#         alpha     X_train   X_test

           8           .948      .879
```

```
########## Lasso R2 score #########

#          alpha     X_train   X_test

#          .0002      .957      .877
```

<u>Top 5 coefficients of Ridge regression</u>

Neighborhood_Crawfor	0.098973
OverallCond_9	0.077265
OverallQual_9	0.076181
OverallCond_8	0.071630
SaleCondition_Normal	0.069565

<u>Top 5 coefficients of Lasso regression</u>

OverallQual_9	0.152613
Neighborhood_Crawfor	0.151378
OverallCond_9	0.134988
OverallQual_10	0.122588
Exterior1st_Stucco	0.102448

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer:

Both Ridge and Lasso have very close R2 score for train and test data. But when

we have a greater number of Betas then it is good to choose Lasso because it eliminates the betas have very less impact on the model.

Question 3

After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables.

Which are the five most important predictor variables now?

<u>Removed Top 5 predictor's which are not present in new incoming data.</u>

Neighborhood_Crawfor	0.156330
OverallQual_9	0.144760
Exterior1st_Stucco	0.141526
OverallCond_9	0.138120
OverallQual_10	0.116798

<u>After removing, New top 5 predictors with their coef values.</u>

SaleType_CWD	0.106182
Exterior1st_BrkFace	0.085790
SaleCondition_Partial	0.081654
SaleCondition_Normal	0.078215
BsmtFullBath_2	0.072939

Question 4

How can you make sure that a model is robust and generalisable?

What are the implications of the same for the accuracy of the model and why?

1. Simple models are more robust and generalised.

2. Overfit models have high variance; such models fail to predict the unseen data.

3. Complex models have high accuracy but to make it robust and generalized we reduce the variance which Leads to bias. By adding Bias accuracy decreases.

4. Lasso and Ridge models helps to find a balance between Accuracy and complexity.