Assignment Subjective 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 double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

- The optimal values of Alpha for ridge and lasso are as follows:
    - Ridge : 3
    - Lasso : 0.001
- Doubling alpha, resulted in slight increase of mean_test and mean_train scores.
- There was no change in the top predictor variables post changes in ridge, but slight changes in Lasso.

```
=====Ridge====
OverallQual                0.267711
TotRmsAbvGrd               0.165807
GarageCars                 0.164700
GrLivArea                  0.156031
OverallCond                0.148093
1stFlrSF                   0.143651
FullBath                   0.135031
Neighborhood_Crawfor       0.131167
Fireplaces                 0.115313
2ndFlrSF                   0.113643
```

```
Name: Ridge, dtype: float64
=====Lasso====
GrLivArea                 0.569924
OverallQual               0.563113
GarageCars                0.217409
TotRmsAbvGrd              0.170288
OverallCond               0.166312
FullBath                  0.157586
Neighborhood_Crawfor      0.134434
Neighborhood_NridgHt      0.110435
Neighborhood_StoneBr      0.098168
Fireplaces                0.095445
=====RidgeNew====
OverallQual               0.267711
TotRmsAbvGrd              0.165807
GarageCars                0.164700
GrLivArea                 0.156031
OverallCond               0.148093
1stFlrSF                  0.143651
FullBath                  0.135031
Neighborhood_Crawfor      0.131167
Fireplaces                0.115313
2ndFlrSF                  0.113643
Name: RidgeNew, dtype: float64
=====LassoNew====
OverallQual               0.644553
GrLivArea                 0.440108
```

```
GarageCars                0.219438
TotRmsAbvGrd              0.195550
Fireplaces                0.119018
OverallCond               0.118670
FullBath                  0.116617
Neighborhood_Crawfor      0.102532
Neighborhood_NridgHt      0.088842
YearRemodAdd              0.079009
```

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?

- Based on the performance, I chose Lasso for its simplicity and better r-squared and mean square error values.

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?

```
Top features post dropping 5 features are:


                   Coefficient
RoofMatl_Membran      2.083879
RoofMatl_WdShngl      1.867802
```

```
RoofMatl_CompShg     1.819287
RoofMatl_Tar&Grv     1.663115
RoofMatl_WdShake     1.437711
```

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?

- Keeping the model simple.
- The model should not explain 100% of training data but 0% of test data.
- Select the right features to keep the model simple but effective.