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?

Answer:

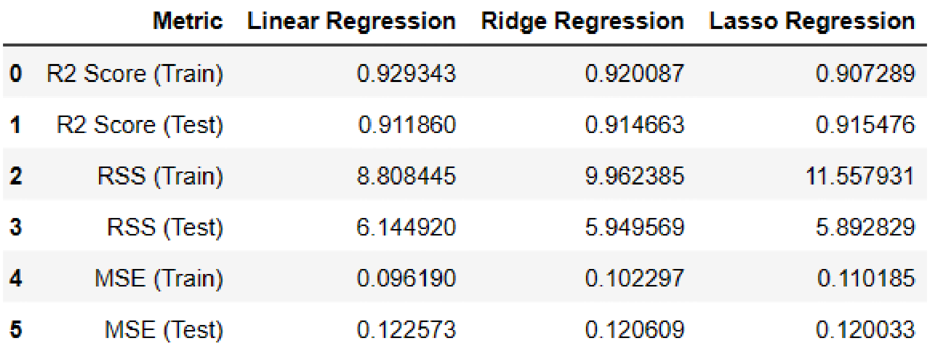
Optimal Value of alpha for ridge and lasso regression are:

- Alpha for Ridge: 10
- Alpha for Lasso: 0.001

After doubling the value of alphas for ridge and lasso:

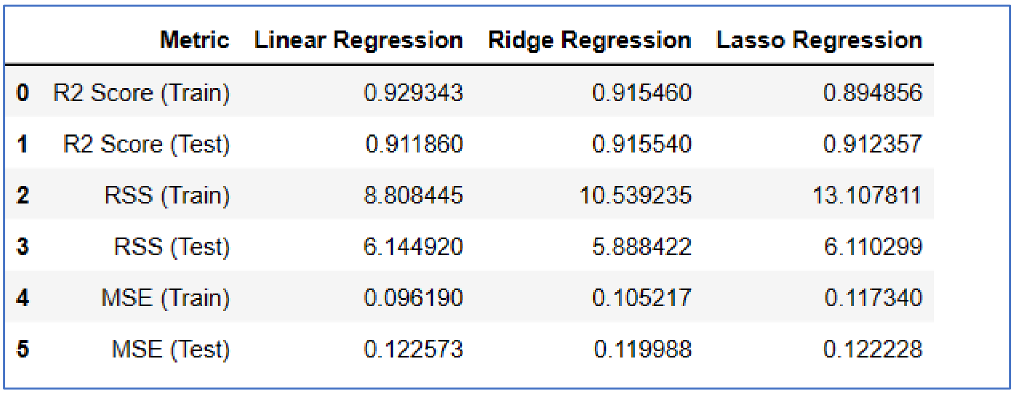
- The coefficients are reduced in case of Ridge and there are more coefficients becoming 0 in case of Lasso (eliminating more features)
- Slight decrease in R2 scores is observed for both the models
- Very slight increase in MSE values
- Slight increase in RSS values
- Overall model accuracy has not been impacted hugely after the alpha value is doubled.

Before doubling the alpha value:
Alpha for Ridge: 10
Alpha for Lasso: 0.001



	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.929343	0.920087	0.907289
1	R2 Score (Test)	0.911860	0.914663	0.915476
2	RSS (Train)	8.808445	9.962385	11.557931
3	RSS (Test)	6.144920	5.949569	5.892829
4	MSE (Train)	0.096190	0.102297	0.110185
5	MSE (Test)	0.122573	0.120609	0.120033

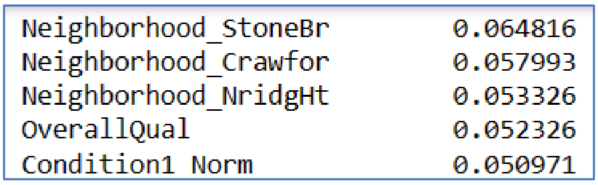
After doubling the alpha value:
Alpha for Ridge: 20
Alpha for Lasso: 0.002



	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.929343	0.915460	0.894856
1	R2 Score (Test)	0.911860	0.915540	0.912357
2	RSS (Train)	8.808445	10.539235	13.107811
3	RSS (Test)	6.144920	5.888422	6.110299
4	MSE (Train)	0.096190	0.105217	0.117340
5	MSE (Test)	0.122573	0.119988	0.122228

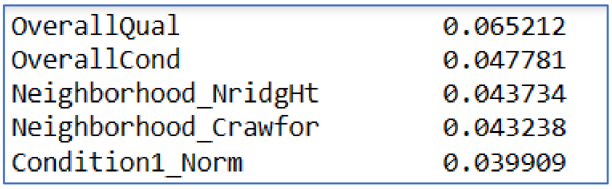
Below is the most important predictor after change is implemented:

For Ridge:



```
Neighborhood_StoneBr        0.064816
Neighborhood_Crawfor        0.057993
Neighborhood_NridgHt        0.053326
OverallQual                 0.052326
Condition1 Norm             0.050971
```

For Lasso:



```
OverallQual                    0.065212
OverallCond                    0.047781
Neighborhood_NridgHt           0.043734
Neighborhood_Crawfor           0.043238
Condition1_Norm                0.039909
```

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:

- Ridge's stats are:
  - Train R2_Scrore: 0.9200872068397271
  - Test R2_Scrore: 0.9146625075981272
  - RSS Train: 9.962385457307084
  - RSS Test 5.949568592163249
  - MSE Train 0.010464690606415004
  - MSE Test 0.01454662247472677

- Lasso's stats are:
  - Train R2 Score: 0.9072886157819771
  - Test R2 Score: 0.915476347689298
  - Train RSS: 11.557930956161405
  - Test RSS: 5.89282920002514
  - Train MSE: 0.012140683777480468
  - Test MSE: 0.014407895354584695

As per the above observed stats, it looks like both the models has achieved a good accuracy, as the R2 score is in the range of 90-92 for the both the models. We can choose Lasso in this case, as it helps in feature elimination by decreasing the model complexity

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?

Answer:

Dropped the top 5 most important predictor variables as per the Lasso model

```
1. OverallQual                   0.065212
2. OverallCond                   0.047781
3. Neighborhood_NridgHt          0.043734
4. Neighborhood_Crawfor          0.043238
5. Condition1_Norm               0.039909
```

The new 5 most important variable are as below after dropping the previous five most important predictor variables :

```
1. CentralAir_Y                  0.073385
2. Neighborhood_StoneBr          0.065794
3. SaleType_New                  0.059009
4. SaleCondition_Normal          0.050526
5. BsmtCond_Gd                   0.042979
```

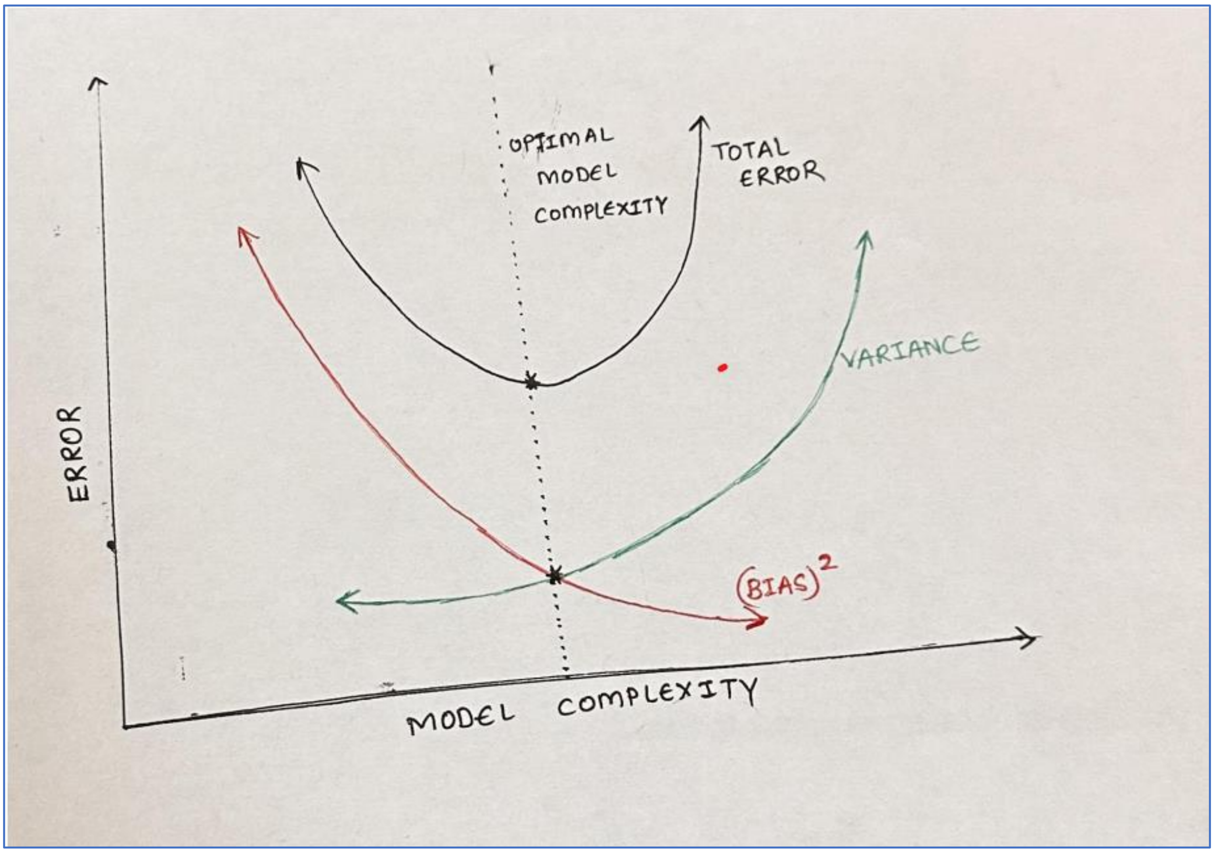
Question-4:

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?

Answer:

- Simpler models are more generic and robust, although we have to compromise on accuracy.
- Complex models tend to memorise the data because of which there is a huge variance even with small change in training data and might not perform well on test data.
- There is a Bias-Variance trade-off, Simple models have low variance, high bias and complex models have low bias, high variance.
- It is important to find the balance in Bias and Variance, so that we make the model simple but not simpler. This can be achieved using Regularization.
- Regularization helps in managing the model complexity by shrinking the coefficients towards 0. It penalizes the model if it becomes more complex and helps in achieving optimum simpler model, by compromising a bit on Bias to reach a position where we have optimum Bias, Variance and minimum total error. This is known as Bias-Variance trade off.
- This point is known as Optimum Model Complexity, where Model is sufficiently simple to be generalisable and also sufficiently complex to be robust.

  As shown in the below graph, accuracy of the model is maintained by keeping the balance between Bias and Variance, with minimum total error.


OPTIMAL
MODEL
COMPLEXITY
TOTAL
ERROR
VARIANCE
ERROR
(BIAS)²
MODEL COMPLEXITY