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?

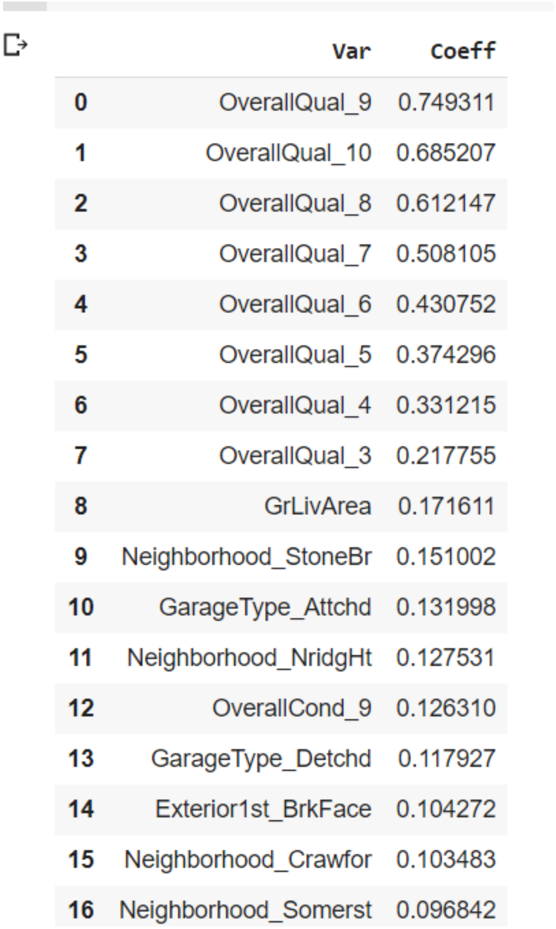
Answer:

For the models I have created the optimal value for:

Ridge Regression: 0.01

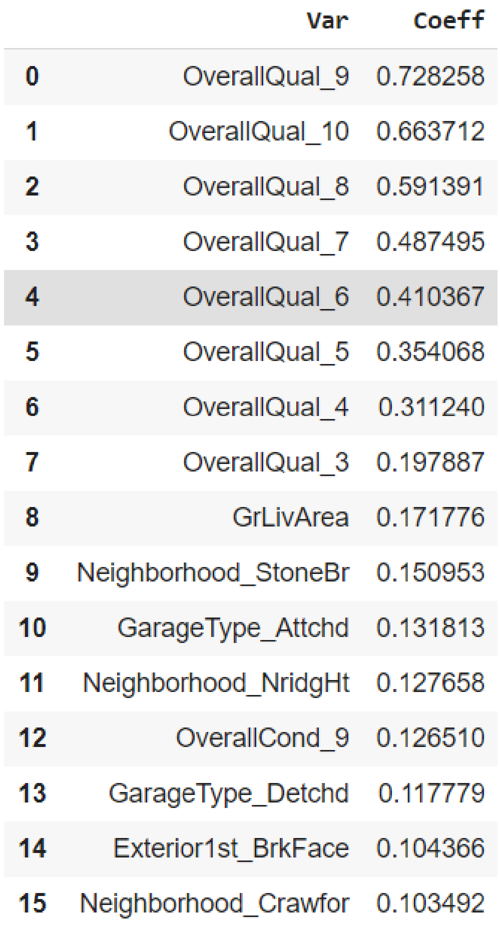
Lasso Regression: 0.00039.

Analysing the cost function for ridge regression, as lambda increases the squares of the sums of coefficients needs to decrease to reach minima for the cost function and as a result the value of coefficients for all the variables will decrease. Same is shown in the below figures:



	Var	Coeff
0	OverallQual_9	0.749311
1	OverallQual_10	0.685207
2	OverallQual_8	0.612147
3	OverallQual_7	0.508105
4	OverallQual_6	0.430752
5	OverallQual_5	0.374296
6	OverallQual_4	0.331215
7	OverallQual_3	0.217755
8	GrLivArea	0.171611
9	Neighborhood_StoneBr	0.151002
10	GarageType_Attchd	0.131998
11	Neighborhood_NridgHt	0.127531
12	OverallCond_9	0.126310
13	GarageType_Detchd	0.117927
14	Exterior1st_BrkFace	0.104272
15	Neighborhood_Crawfor	0.103483
16	Neighborhood_Somerst	0.096842

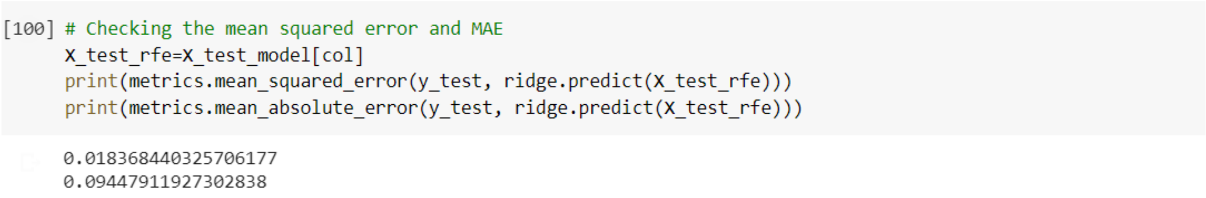
Figure 1/Optimum Lambda=0.01



	Var	Coeff
0	OverallQual_9	0.728258
1	OverallQual_10	0.663712
2	OverallQual_8	0.591391
3	OverallQual_7	0.487495
4	OverallQual_6	0.410367
5	OverallQual_5	0.354068
6	OverallQual_4	0.311240
7	OverallQual_3	0.197887
8	GrLivArea	0.171776
9	Neighborhood_StoneBr	0.150953
10	GarageType_Attchd	0.131813
11	Neighborhood_NridgHt	0.127658
12	OverallCond_9	0.126510
13	GarageType_Detchd	0.117779
14	Exterior1st_BrkFace	0.104366
15	Neighborhood_Crawfor	0.103492

Figure 2/Lambda=0.02

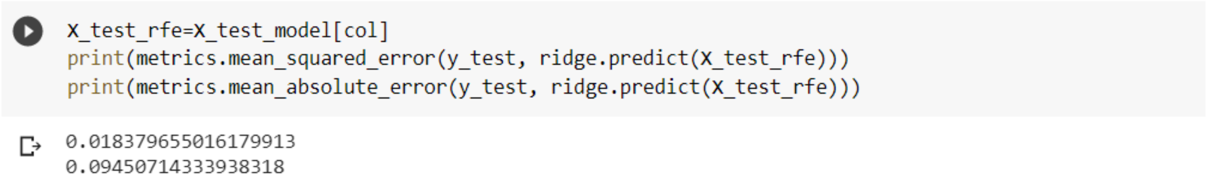
Also depending on the decrease in the coefficient values, the bias might increase slightly but that might result in slight increase in variance to a certain extent. In our case, as the margin of increase is very less, the train and test results improved marginally.



```
[100] # Checking the mean squared error and MAE
      X_test_rfe=X_test_model[col]
      print(metrics.mean_squared_error(y_test, ridge.predict(X_test_rfe)))
      print(metrics.mean_absolute_error(y_test, ridge.predict(X_test_rfe)))

      0.018368440325706177
      0.09447911927302838
```

Figure 3/MSE and MAE for ridge model with optimum lambda=0.01



```
X_test_rfe=X_test_model[col]
print(metrics.mean_squared_error(y_test, ridge.predict(X_test_rfe)))
print(metrics.mean_absolute_error(y_test, ridge.predict(X_test_rfe)))

0.018379655016179913
0.09450714333938318
```

Figure 4/MSE and MAE for ridge model with double of optimum lambda=0.02

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:

I will apply Lasso in this case. Reasons:

Even though, MSE of test sets for both the regressions is almost marginal- MSE of Lasso is still better than that of Ridge (as shown in figure 3 and 4). Other than that, applying Lasso reduces the intermediate step of RFE that's applied in case of Ridge reducing the speed of training and prediction.

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:

But why to build another model excluding the 5 most important predictor variables? That will hamper the prediction and analysis of factors affecting house price causing loss to business. We can process the incoming data in a way we have processed the data before training and then send the processed data to the model for prediction. I have carried out a similar processing in thie end to end project. Please click here to see the project and click here to check the processing part.

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:

According to Occam's Razor, a regression model must be as simple as possible but no simpler.

Meaning of simple but no simpler:

1. If we use almost all the variables for prediction, the model tends to memorize the data and hence creates a model with low bias but high variance(overfitting). Hence to simplify the model, we need to drop irrelevant variables (Multi-collinear, not linearly related to the target variable at all, having > 40% missing values etc). As a result, we might see decrease in the training accuracy but the test accuracy increases leading to slightly higher bias than the early model (but still low bias) and low variance. Now, this philosophy when applied more than needed might lead to drop almost all columns except few and that scenario might lead into increases training and test error-high bias and high variance called underfitting.
   Hence a general and robust model is one which has minimum number of variables leading to optimum low bias low variance condition till it doesn't start going in the lane of underfitting.

2. The lesser the degree of polynomial for the model, the better it is. This concept comes into picture, when the data we have is completely non-linear. In such scenarios, equations with degree of polynomial=1 might not be a good fit. Increasing the degree of polynomial of the

model might lead us to better training accuracy but if it’s overdone, it again mugs up the dataset and overfits and gives inefficient test accuracy. Hence, it’s observed that models with higher degree polynomial react highly even if the change in the data is small. Such models are neither generalised nor robust. Hence the degree of polynomial needs to be kept as minimum as **necessarily** possible to build a robust and a generalised model.