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?
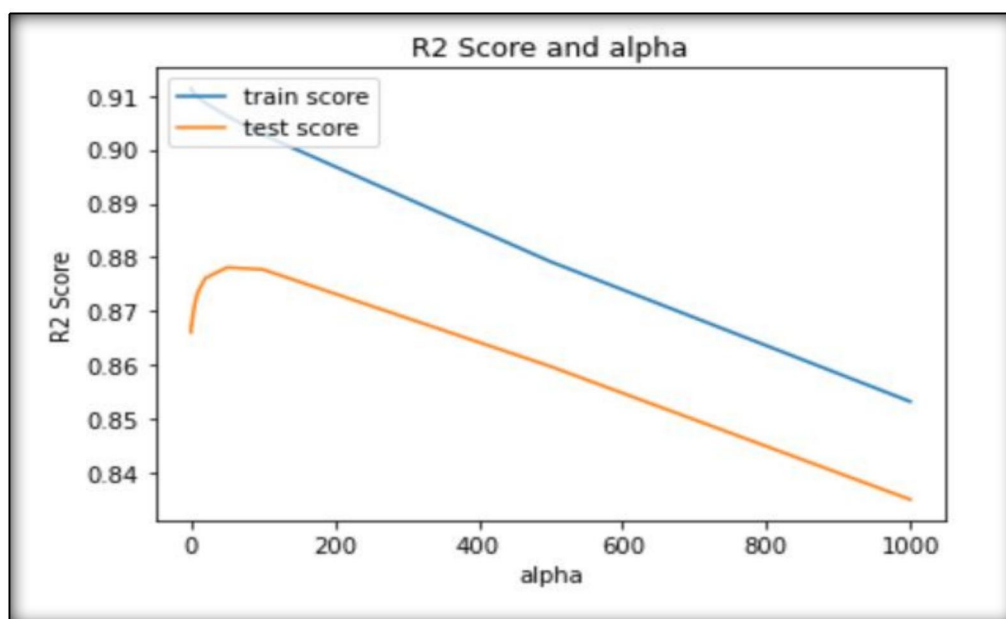
Optimal Value
Lasso:0.001
Ridge:9

If we choose the double the value of alpha for ridge and lasso.
The prediction accuracy remain same there is small change in the coefficient values.

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?

For Ridge:

	mean_train_score	mean_test_score	param_alpha
24	0.906228	0.878142	50
25	0.902744	0.877761	100
23	0.908643	0.876173	20
22	0.909639	0.873707	10.0
21	0.909752	0.873305	9.0
20	0.909870	0.872856	8.0
19	0.909994	0.872352	7.0
18	0.910125	0.871784	6.0
17	0.910265	0.871139	5.0
16	0.910420	0.870403	4.0

Optimal value for ridge is 9 because after alpha increase there is no big change in r2score so we choose 9 .

For Lasso:
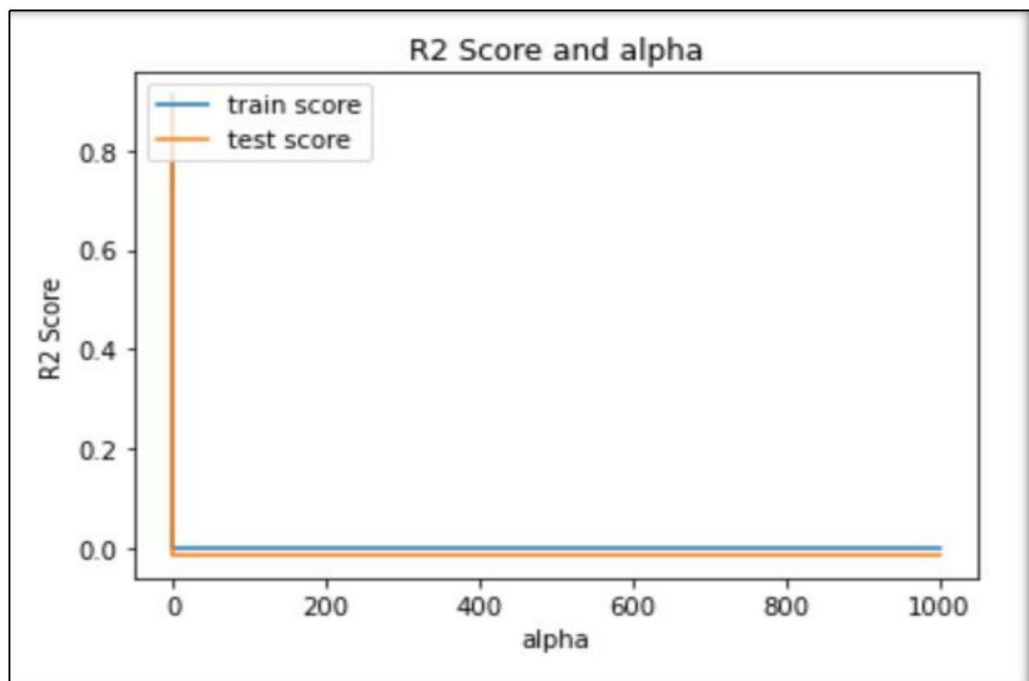
|	mean_train_score	mean_test_score	param_alpha
1	0.907071	0.881937	0.001
2	0.891066	0.877371	0.01
0	0.910182	0.875308	0.0001
3	0.805029	0.793059	0.05
4	0.686689	0.674830	0.1
5	0.375769	0.360684	0.2
6	0.007784	-0.009933	0.3
7	0.000630	-0.013396	0.4
22	0.000000	-0.013624	10.0
20	0.000000	-0.013624	8.0
21	0.000000	-0.013624	9.0
26	0.000000	-0.013624	500
23	0.000000	-0.013624	20

Optimal value for Lasso is 0.001 because after this alpha value increase then there will big change in r2_score of train set and test set.

--

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 Five Variable:
1 GrlivArea
2 OverallQual
3 Yearbuilt
4 OverallCond
5 TotalbsmtSF

Coefficients in the Lasso Model
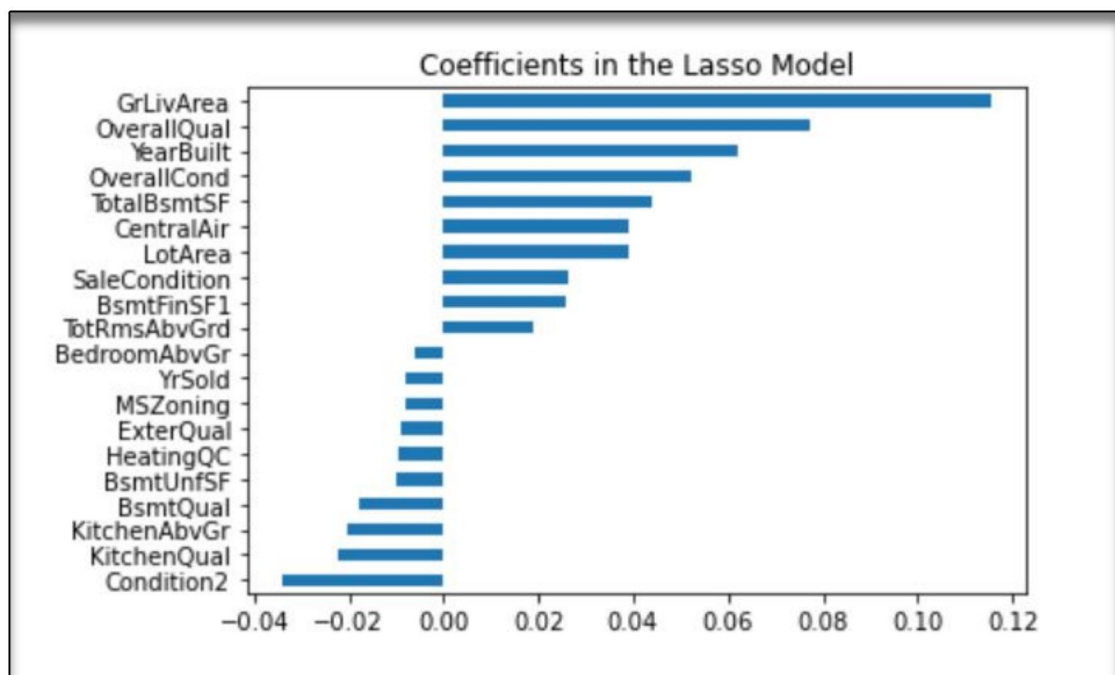
After excluding the top 5 predictor variable

Coefficients in the Lasso Model
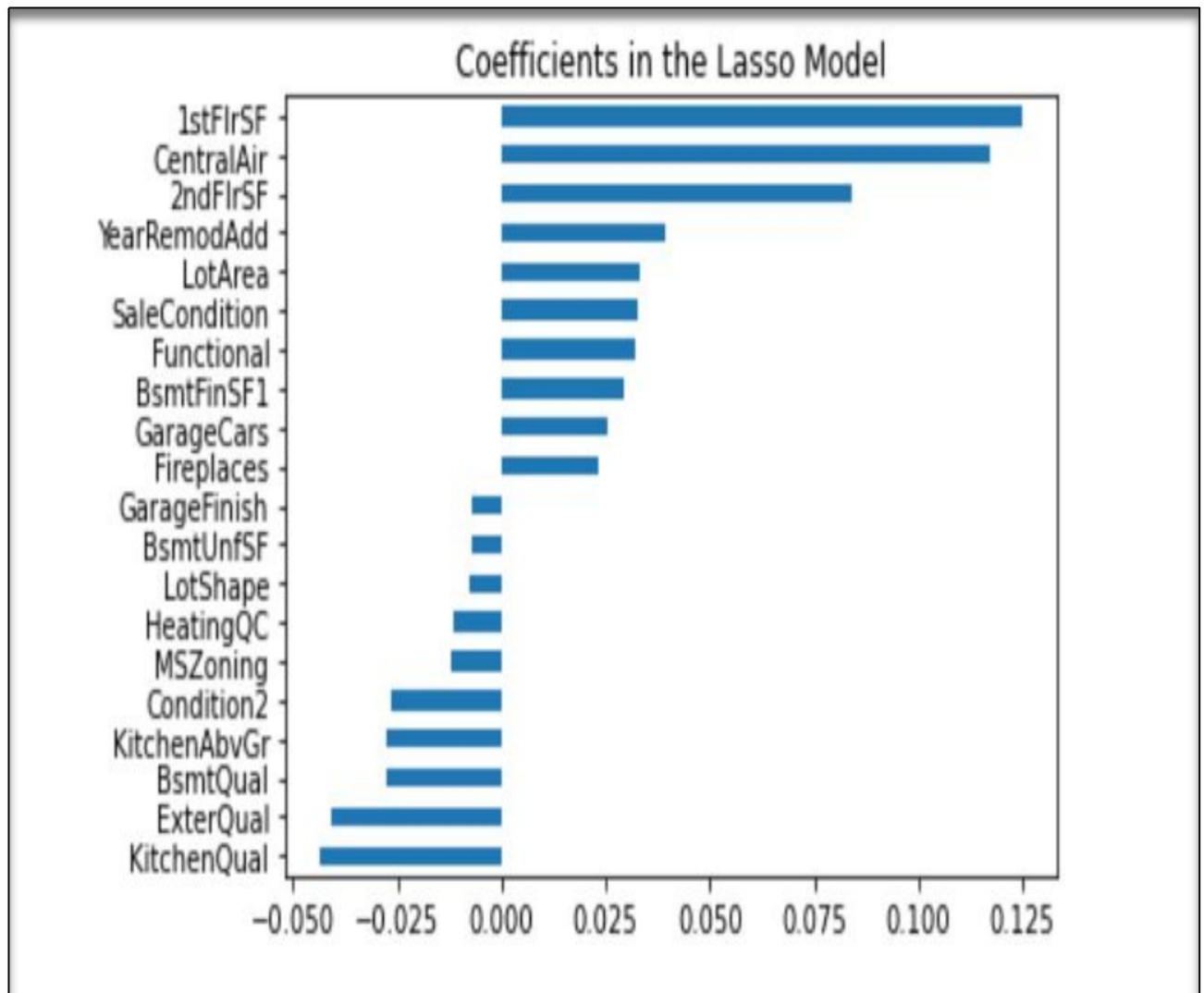
Top Five Variable:
1 Central Air
2 1stFlrSF
3 2ndFlrSF
4 YearRemodAdd
5 LotArea

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?

A Model is considered to be robust if the model is stable I.e does not change drastically upon changing the training set.The model is considered generalisable model if it does not overfits the training data and works well with new data.
Its implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data I.e the accuracy does not change much for training and test data.