Assignment-based Subjective Question

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 1

Optimal Value of alpha in Ridge model is 0.05 .

Optimal Value of alpha in Lasso model is 0.0001

R2 value with alpha 0.05 in Ridge is 90.87% and 90.21% for train & test set of data respectively .

If we double the alpha value ie Alpha = 0.1 the R2 value changes as below :

R2 dipped to 89.71% for train set & 90.1% for test set .

	Metric	Linear Regression	Ridge Regression
0	R2 Score(Train)	0.955212	0.897152
1	R2 Score(Test)	0.959077	0.901210

The most important predictor variable still is : 'LotArea' after the alpha value was doubled.

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?

Answer2

With optimum lambda value of 0.05 in Ridge Regression we get better values that Lasso regression with 0.0001 . The R2 scores for train & test sets for Ridge regression is very close both at 90% level for train & test data set . Whereas the R2 score for Lasso regression has dipped to 84% for Training set & test set score is 89% which is much higher than the train set .

Also evaluating the Mean squared error metrics – we see its very very low for Ridge regression . Hence I have chosen the RIDGE Regression model .

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score(Train)	0.955212	0.908791	0.843299
1	R2 Score(Test)	0.959077	0.902188	0.893752
2	RSS(Train)	0.551220	1.122535	1.928557
3	RSS(Test)	0.351268	0.839590	0.912002
4	MSE(Train)	0.000540	0.001099	0.001889
5	MSE(Test)	0.000800	0.001913	0.002077
6	RMSE(Train)	0.023235	0.033158	0.043461
7	RMSE(Test)	0.028287	0.043732	0.045579

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 3

If the first 5 most important predictor values are not considered ,below 5 are the next most important predictor values :

```
'2ndFlrSF' ,'GrLivArea' ,'BedroomAbvGr', 'KitchenAbvGr', 'GarageArea'
```

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 4

To make the model generalizable and robust : we split the data in Train & test set at 70:30 ratio . We build the model on train data set & keep aside the test data set for final test & evaluation of the model . The metrices like R2 score , Residual sum of squares , mean squared error are evaluated of the model build . We look for a good R2 score . The higher the R2 value the better the model . We then evaluate model against the test set & ensure the R2 value is very close . We look for RSS Values & MSE values very close to 0 . This ensures the model is robust .