Mitali Seth

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?

Ans:
Alpha value for Ridge is 6
Alpha value for Lasso is 0.0001

Doubled the Ridge value:

There is a drop of 1% in the r2 score post doubling the value for ridge

Doubled the Lasso value:

There is a drop of 1% in the r2 score post doubling the value for lasso

```
final_metric2
```

Out[325]:

	Ridge2 Regression	Lasso2 Regression
0	0.947219	0.953977
1	0.911882	0.911431
2	7.215491	6.291611
3	5.118878	5.145046
4	0.090346	0.084364
5	0.116063	0.116360

In [326]:
```
# without doubling result
final_metric
```

Out[326]:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.638171e-01	0.950637	0.958631
1	R2 Score (Test)	-3.031050e+17	0.911930	0.905829
2	RSS (Train)	4.946401e+00	6.748234	5.655415
3	RSS (Test)	1.760770e+19	5.116098	5.470522
4	MSE (Train)	7.480292e-02	0.087371	0.079985
5	MSE (Test)	2.152581e+08	0.116032	0.119984

In [ ]:

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?

Solution

Lasso regression would be best, because there are too many features, Lasso will help in feature elimination as well. Hence I would chose Lasso for robust model.

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?

Solution:

After dropping 1st top 5 feature and built model again then below are the next top 5 features :

```
In [305]: betas['Lasso'] = lasso.coef_

In [306]: betas['Lasso'].sort_values(ascending=False)[:5]
          # after dropping top 5 feature in lasso, these are top 5 feature

Out[306]: OverallQual_9            0.132236
          GrLivArea                0.115567
          SaleCondition_Partial    0.099857
          Functional_Typ           0.096865
          Neighborhood_Crawfor     0.093903
          Name: Lasso, dtype: float64
```

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?

Solution : Model will be **robust** if any variation in the data does not affect its performance much. Model will be generalisable when accuracy of training data and test data are close

To make sure a model is robust and generalizable, we have to **take care it doesn't overfit**.
This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data.