

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:

I am getting alpha for Ridge as '6' and alpha for Lasso as '0.0005'. If we double the value for alpha or both Ridge and Lasso we get slightly off-set from the accuracy of our model. Neighborhood is the most important predictor we get in Ridge and Lasso regression.

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:

We have determined the optimal value of Lambda by some manual fine tuning of our data. Based on the Final metric table we see Ridge is performing better than Lasso for our data analysis.

Out[42]:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	8.647847e-01	0.844158	0.830027
1	R2 Score (Test)	-2.279971e+23	0.829277	0.826449
2	RSS (Train)	4.093063e+00	4.717463	5.145213
3	RSS (Test)	3.099159e+24	2.320631	2.359076
4	MSE (Train)	6.331569e-02	0.067974	0.070989
5	MSE (Test)	8.411721e+10	0.072789	0.073390

Question-3:

After building the model, you realized 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:

The five most important predictor variables after removing the initial five most important predictors are FullBath, LotConfig_CulDSac, Functional_Maj2, Condition2_RRAe, Foundation_PConc. Based on our Lasso Model.

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:

The model should be generalized so that there is no large gap in test accuracy and the training accuracy.

The model should be accurate for datasets other than the ones which were used during training. Those outliers which it does not make sense to keep must be removed from the dataset. If the model is not robust, It cannot be trusted for predictive analysis.

