Problem Statement -Part II

Ques: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:

Optimal Val for ridge:20
Optimal Val for lasso:0.001

If we double the value for alpha then

Ridge—Alpha changed from 20 to 40 will result in increase of RMSE val

Lasso-More Features will get removed

Top most predictor variables:

	Feature	Coef	mod
0	MSSubClass	11.819914	11.819914
90	Condition2_RRAe	-0.187161	0.187161
16	BsmtFullBath	0.114787	0.114787
60	Neighborhood_Edwards	0.112939	0.112939
75	Neighborhood_StoneBr	0.101380	0.101380
4	OverallCond	0.091284	0.091284
70	Neighborhood_OldTown	0.088734	0.088734
61	Neighborhood_Gilbert	-0.064410	0.064410
119	Exterior1st_CBlock	0.058937	0.058937
80	Condition1_PosA	0.058771	0.058771

Ques2: 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 lot of features in given data set.It make more sense if we select lasso as it removes unwanted features

Ques3: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?

Answer3:Model accuracy will get impacted after dropping main predictor variables.Next set of variables after dropping above mentioned variables are 1stFlrSF, MSSubClass_90, MSSubClass_120, TotalBsmtSF, HouseStyle_1Story

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: Simplifying the model enhances robustness and generalizability at the cost of some accuracy, exemplified by the Bias-Variance trade-off. A balance between bias and variance is crucial to prevent overfitting (high variance) or underfitting (high bias). Striking this balance ensures the model performs consistently well on both training and test data, maintaining accuracy across different datasets.