Problem Statement - Part II

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:
The Optimal value of alpha for ridge is 2 and for lasso it is 0.0001.

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 will choose Lasso as its giving feature selection option also. It has removed unwanted features from model without affecting the model accuracy, which makes model generalized and simple and accurate.

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:

The most important predictor variables are as follows:

	Lasso Co-Efficient
LotFrontage	0.146535
Total_porch_sf	0.072445
HouseStyle_2.5Unf	0.062900
HouseStyle_2.5Fin	0.050487
Neighborhood_Veenker	0.042532

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:

To make model robust and generalisable 3 features are required:

1. Model accuracy should be > 70-75%: I our case its coming 80%(Train) and 81%(Test)

which is correct.

2. P-value of all the features is < 0.05

3. VIF of all the features are < 5

Thus we are sure that model is robust and generalisable.