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 is 50 for Ridge – 100 for Lesso,
- If double the value of alpha for both ridge and lasso, the values with remain the same .
- In Ridge predictor variables are   - Neighborhood_Stone Br,GarageArea,Neighborhood_NridgHt,TotalBsmtSF,GrLivArea , KitchenQual,Neighborhood_Names,Neighborhood_Edwards,BldgType_TwnhsE,GarageFinish
- In Lasso, predictor variables are - TotalBsmtSF ,SaleType_New ,MSZoning_RM,GarageType_Attchd ,GrLivArea,Neighborhood_NAmes ,Neighborhood_OldTown,KitchenQual ,SaleCondition_Partial ,RoofStyle_Gable.

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:

Both Ridge and Lasso give very similar results in terms of performance same RMSE value .

So, Lesso will be selected.

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 five most important predictor variables

'GrLivArea','GarageType_Attchd','MSZoning_RM','SaleType_New','TotalBsmtSF'

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer:

If the accuracy of both training and testing data are nearly same and acceptable, we can say the model is robust.