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 and lasso regression are as below:

Alpha for ridge – 0.8

Alpha for Lasso – 50

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:

R2 score for redge regression is slightly higher than the Lasso for test dataset, As per the assignment evaluation, I will choose Redge regression. Usually Lasso model would provide the more R2 score, but in my evaluation it's Ridge regression has slightly higher values, because the optimal value of Alpha for Lasso is slightly higher.

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:

Following will be the most important predictor variables

1. TotRmsAbvGrd
2. BsmtFinSF1
3. Street_Pave
4. Neighborhood_NridgHt
5. GarageArea

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:

Model can be built considering whether the pattern follows Linear or Non linear regression. If the pattern is Linear, we can apply Linear Regression Model and Multi Linear Regression. If not Model has to be extended to check if Linear can be extended by applying Polynomial Regression or data transformation, If not Redge and Lasso regression model can be applied by evaluating coefficients and optimal value of Lambda. The smaller the optimal value of lambda, it would be better to avoid underfitting, which will have less variance and Errors.