ASSIGNMENT :ADVANCE REGRESSION

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 : Optimal value for alpha for ridge is 5.0 and lasso is 0.001.

In the Model when the alpha is change and re iterated I donot see much difference in the train and test scores.i.e. before doubling and after doubling the scores of train and test seems pretty much near .

However Lasso tends to do well if there are few significant parameters and Ridge works well with more parameter.

#LotArea is the most important predictor variable.

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 :After creating model in both Ridge and Lasso we can see that the r2_scores are almost same for both of them but as lasso will penalize more on the dataset and can also help in feature elimination .Hence considering Lasso as the final mode

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 :

BsmtFullBath

MasVnrArea

BsmtExposure

BsmtFinType2

CentralAir

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 : Based on outliers , test accuracy the model can obtain the robust and generalizable

The outliers which are not required to be removed/capped and the outliers which adds weightage to the analysis should be retained.In this way the accuracy of a model can increase .

If the model is neither robust nor generalizable it cannot be trusted.