TOPIC – ADVANCED LINEAR REGRESSION (THEORY BIT)

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 lasso is 1 and ridge is 10.

If we choose double the value of alpha for both regressions, it effects the rsquare value. For Ridge regression, rsquare decreases for train set and remains same for test set. For Lasso regression, rsquare decreases for train and test set.

The most important predictor variables after change would be –

1. OverallQual
2. TotalBsmtSF
3. Neighborhood
4. Exterior1st
5. LotArea
6. CentralAir

===

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

I have taken Lasso regression over Ridge as there are so many features, and removing the features with less coefficients seems to be fit.

==

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 include : OverallQual, LotArea, Neighborhood, Exterior1st, BsmtQual

==

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

The model holds high r square value which is good fit. Though my model couldn't give an exact linear line but still model built close to the actual values making sure it doesnot overfit and is accurate.

==