Assignment 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:

Optimal value of alpha for Ridge is 100 and for Lasso Regression is 0.01

When the value of alpha is doubled for Ridge and Lasso, the mean test score is changed for Ridge regression from 89.95% to 89.71% and for Lasso Regression, the mean test score changed from 89.88% to 88.78%.

The most important predictor variables after the change with Lasso regression is:

1. LowQualFinSF
2. LotArea
3. BsmtUnfSF
4. LotFrontage
5. MasVnrArea
6. Neighborhood_NoRidge
7. Exterior1st_Wd Sdng
8. Neighborhood_Somerst
9. GarageYrBlt
10. OverallQual

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:

As per the determined optimal value of lambda for Ridge and Lasso regression, both seems to be good fit. But I will proceed with Lasso regression, as there is less variation between the mean train score and mean test score.

In Lasso, some of these coefficients become 0, thus resulting in model selection and easily interpretation is possible, particularly when the number of coefficients is very large.

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?

The five most important predictor variables are:

1. LowQualFinSF
2. LotArea
3. BsmtUnfSF
4. LotFrontage
5. MasVnrArea

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:

A model is robust if the value does not change drastically when the training data is changed. The model is generalised if it performs good on training set and also performs well on the test data or any new data.

If the model is robust and generalisable, then the model will perform will have good accuracy on both training and test data.