Advanced Regression Assignment

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 to 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 lambda for Ridge regression is 10 and optimal value of lambda for Lasso Regression is 0.001. if we double the value of alpha, in case of ridge it will lower the coefficients and in case of Lasso regression there will be more of less important features becoming zero.

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 :

Got a good score for Both Ridge and Lasso Regression. As we had a lot of features in this dataset, we will be going with Lasso Regression as the coefficients of lesser important features becomes zero thus it reduces the difficulty in feature selection process.

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 :

On running the same notebook by removing the top 5 features, the new top 5 features are as follows :

- GarageType_BuiltIn

- GarageType_Detchd

- GarageType_No Garage

- GarageType_Others

- GarageFinish_No Garage

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 following Changes can be made to the model :

- Build a model that's resistant to outliers : Tree based models are generally not affected by the outliers present in the data.

- Try to use a more robust error metric

- Transform your data : Try a log based or a exponential transformation if your dataset is skewed.

- Remove the outliers.