

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:

The optimal value for Ridge and lasso regressions are as follows:

1. Ridge: 0.001
2. Lasso: 0

If the value of alpha is doubles, the ridge regression coefficients will reduce and in lasso regression, there will be more coefficient values 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:

Ridge results: Train: 93.15, Test:96.053

Lasso results: Train:93.15, Test:96.057

Though Ridge and Lasso, both are giving good results, since Lasso has the advantage of feature selection, by coefficients of lesser important features becoming zero, we can choose to apply lasso.

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: 'GarageType_BuiltIn', 'GarageType_CarPort', 'GarageType_Detchd', 'GarageType_none' are the important variables.

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:

To have a robust and generalisation model, onus is on the underlying data. The underlying data must be of good quality and should cover a lot of scenarios. Also the Exploratory Data Analysis steps like missing value treatment, outlier treatment etc must be premed rigorously. The better the quality and diversity of data, the better the accuracy of the model created using it.