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:

- Optimal value of Alpha for Ridge: **5.0**
- Optimal value of Alpha for Lasso: **100**
- Models choose to double the Value of Alpha for Ridge & Lasso are: **10** & **200**

Most Important Predictor Variables:

- OverallCond_3
- OverallQual_10
- Neighborhood_Timber
- GarageType_Attchd
- BsmtExposure_Mn
- Neighborhood_NridgHt
- RoofStyle_Gable
- Exterior1st_BrkComm
- Neighborhood_Edwards
- ExterQual_Fa

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:

There are two main regularization techniques, namely Ridge Regression and Lasso Regression. They both differ in the way they assign a penalty to the coefficients.

Lasso method is usually used in machine learning for the selection of the subset of variables. It provides greater prediction accuracy as compared to other regression models. Lasso Regularization helps to increase model interpretation.

The less important features of a dataset are penalized by the lasso regression. The coefficients of this dataset are made zero leading to their elimination. The dataset with high dimensions and correlation is well suited for lasso regression.

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:

Top features are **OverallCond_3, OverallQual_10, Neighborhood_Timber, GarageType_Attchd & BsmtExposure_Mn** as this have high coefficient value. After dropping the top features there is no major change in r2 score. From 94.42% to 94.59% for Train and from 35.20% to 37.50% for Test.

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 will be considered too robust if it does not change extremely while changing the training data.

Model will be considered as generalisable, if it tries to overfit the training data and properly work with unknown or new data.

The implication of accuracy is that a robust and generalisable model will perform well on both training and test data.