Assignment Part II – Subjective Questions

By – Ishkhan Marzook

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 alpha value for Ridge and Lasso regression is 10.0 and making alphas of both Ridge and Lasso regression doesn't make a considerable change for the mode. will be the most important predictor variables after the change is implemented are,

- **TotRmsAbvGrd**
- **OverallQual**
- **Fireplaces**
- **PoolQC**
- **GarageArea**
- **BsmtExposure**
- **BsmtQual**
- **LotArea**
- **MasVnrArea**
- **OverallCond**

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

Since the current Housing Data Model contains a considerable number of significant features, the best option will be the **Ridge Regression**, Lasso Regression do well when there are a smaller number of significant features.

Hence decided to choose **Ridge Regression** for the Housing Price Prediction Model.

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

I'll choose the following variables,

- **TotRmsAbvGrd**
- **OverallQual**
- **Fireplaces**
- **PoolQC**
- **GarageArea**

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 robust model won't have considerable performance hinder when testing with any kind of unseen data, and to make the mode generalizable we need to build the model in a way to perform well on an unseen dataset.

When we try to make the model more robust and generalizable, it may tend the model to become complex which may cause the model to become biased, hence we need to optimize the model to stand mid in Variance and Biasness.