Advanced Linear Regression Subjective Questions

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?

The optimal value of alpha for ridge and lasso regression should not be too low or too high.

If it is too low, model won't handle over fitting

If it is too high, model will under fit

In the Problem Statement, optimal value of alpha for

- Ridge : 3
- Lasso : 0.0001

When we double the alpha values, the difference in r2 for train and test reduced (small reduction in my case)

After the change is implemented, important variables –

- 'GrLivArea'
- 'OverallQual'
- 'TotalBsmtSF'
- 'OverallCond'
- 'GarageArea'
- 'YearBuilt'
- 'Somerst'
- 'NridgHt'
- 'NoRidge'
- 'Crawfor'

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignme nt. Now, which one will you choose to apply and why?

I will choose to apply Lasso regression because it has feature elimination and the model will be Robust

Question 3

After building the model, you realised that the five most important predictor variables in the lasso m odel 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?

We will consider the next five most important predictor variables after sorting –

- 'YearBuilt'
- 'GarageArea'

- 'NoRidge'
- 'Crawfor'
- 'NridgHt'

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?

- Removing outliers while training the data
- Checking the r2 values of train and test values and making sure that the difference is not too high
- Using Lasso Regression if the model overfits, so feature elimination can be applied and model becomes robust