

# Advanced Regression Assignment

Submitted By

Shamseena VM

## 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?

**Answer:**

Optimum value obtained for alpha used for Ridge regression is 2

Optimum value obtained for alpha used for Lasso regression is 0.001

Ridge: by doubling the value of alpha we can see that, r2 score for train and test set is reduced

Lasso: In lasso, we can see that doubling the alpha value doesn't impact much since r2score of train and test looks similar to original value

```
['Neighborhood_NoRidge', 'OverallQual_9', 'OverallQual_10',  
 'RoofMatl_WdShngl', 'KitchenQual_Fa']
```

Are the important variables.

### 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:**

Will choose to apply lasso since it doesn't impact much on the original value obtained after doubling alpha value.

### 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 :**

```
['MSZoning_RL', 'RoofMatl_WdShake', 'BsmtQual_No Basement',  
 'KitchenQual_TA', 'SaleType_New']
```

Are the new top 5 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:**

1. perform train-test split: Model is trained on the train set and its model is evaluated on the test set.

A good model performs pretty well in test also.

2. Analyze p value

3. Analyse VIF value for correlation

4. Use RFE selection of variables

5. use k fold cross validation

6. use regularization

7. Use Hyper parameter tuning