

## 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?

Based on our final analysis we got to know the Optimal value of Alpha is:

- Lasso Regression = 0.008
- Ridge Regression = 10

When we double value of alpha R2 Score decreases

### Ridge

1. Neighborhood\_NridgHt
2. Neighborhood\_NoRidge
3. BldgType\_Duplex
4. GrLivArea
5. RoofMatl\_WdShngl
6. TotalBsmtSF
7. SaleCondition\_Partial
8. SaleType\_New
9. Condition2\_PosN
10. 1stFlrSF

### Lasso

1. Neighborhood\_NridgHt
2. GrLivArea
3. Neighborhood\_NoRidge
4. TotalBsmtSF
5. SaleType\_New
6. BldgType\_Duplex
7. LowQualFinSF
8. 1stFlrSF
9. SaleCondition\_Partial
10. Functional\_Sev

## 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?

We have made the model using both Ridge and Lasso and the optimal lambda values are:

Ridge = 10

Lasso = 0.008

I would choose Lasso as it helps in features selection and also penalizes for adding more features.

## 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?

After building Lasso Regression model with value of alpha 0.002 now in new model, following below are the new Top 5 variables:

1. 1stFlrSF
2. 2ndFlrSF
3. LowQualFinSF
4. Condition2\_PosN
5. Condition2\_RRAe

#### **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?

A model needs to be:

1. robust and generalizable so that they are not impacted by outliers in the training data.
2. The model should also be generalisable so that the test accuracy is not lesser than the training score.
3. The model should be accurate for datasets used in training.

The outlier analysis needs to be done and only those which are relevant to the dataset need to be retained.

Those outliers which it does not make sense to keep must be removed from the dataset. This would help increase the accuracy of the predictions made by the model. This would help standardize the predictions made by the model. If the model is not robust, it cannot be trusted for predictive analysis.