

### 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 value of alpha for Ridge is 0.5.

The optimal value of alpha for Lasso is 0.001.

Change in R2 Score after doubling the alpha for Ridge Regression.

R2 Score	Ridge Train	Ridge Test
Before	0.93713397	0.936403986
After	0.936629889	0.936419451

Change in R2 Score after doubling the alpha for Lasso Regression.

R2 Score	Lasso Train	Lasso Test
Before	0.937588123	0.935913354
After	0.937559754	0.936194396

There is a very small decrease in the r2 Score after doubling the alpha and there is no change in features that are most important for the model.

### 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:

Lasso regression performs better when there is data set with fewer features and has a low correlation. But in this assignment, there are more than 60 features with high correlation with each other. But in this condition, Ridge regression performs better than lasso Regression as this data set has a high number of correlated 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?

Most important features of Ridge:

Before dropping: Neighborhood\_StoneBr ,MSZoning\_FV ,MSZoning\_RH ,MasVnrType\_Stone ,Neighborhood\_Crawfor

After dropping: RoofStyle\_Shed ,Exterior1st\_BrkFace ,BsmtExposure\_Gd ,Condition1\_PosN ,LandContour\_HLS

Most important features of Lasso:

Before dropping: MSZoning\_FV ,MSZoning\_RH ,MSZoning\_RL ,MSZoning\_RM ,Neighborhood\_StoneBr

After dropping: MasVnrType\_Stone ,MasVnrType\_None ,MasVnrType\_BrkFace ,RoofStyle\_Shed ,Exterior2nd\_CmentBd

### 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 :

List of methods to make sure that model is robust and generalizable.

- A. Outlier Detection: Outliers should be detected properly and removed as outliers have a large impact on the magnitude of the linear regression model's coefficient.
- B. feature scaling: In the assignment, SalePrice has a higher value compared to the other variables which will create more impact compared to other important

smaller numerical features. Hence it is important to scale features before feeding data to the model.

- C. Parameter Tuning: This can be achieved by grid search and cross-validation. It is used to find the optimal value that balances the trade-off between the magnitude of the coefficient and fit.
- D. Cross validations: It is used to check whether the model is overfitting to the test data or not. In this process training data is spilt into several subsets and the training model on each subset and an average is taken.