

### 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 :

From the model we have built we could see that the optimal value of alpha for Ridge and Lasso are

Ridge – 5

Lasso – 0.0002

After doubling the alpha value for both Ridge and Lasso, we do not see any significant change, the mean squared error is slightly higher than the old value. That mean choose higher value will result in more error.

New MSE Ridge - 0.0175067

Old MSE Ridge – 0.0174621

New MSE Lasso - 0.0178796

Old Mse Lasso – 0.0177438

Below are the Most important predictor variable based on Ridge regression

OverallQual

MSZoning\_FV

Neighborhood\_StoneBr

GrLiveArea

SaleCondition

Below are the Most important predictor variable based on Lasso regression

MSZoning\_FV

MSZoning\_RH

GarageType\_Builtin

OverallQual

Neighborhood\_StoneBr

### 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 :

I will choose Ridge Regression as we have observed less SME compared to Lasso. Also it includes all the variable in the final model unlike Lasso which ignores the variable whose coeff becom zero

### 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 :

Five most important predictor variable now are:

ToatalBsmtSF ,SaleCondion ,KitchenQual ,GrLiveArea ,GarageCars

#### 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:

We should make the model as simple as possible although it will result in low accuracy but it will be more robust and generalized. As the model is simple the Bias will be very high and Variance will be very low. So the model will perform poorly unless we improve the complexity of the model and balance the variance and bias