

**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?**

Ridge - Alpha 10

Lasso – Alpha 0.001

After doubling the value of alpha for ridge, we get same R2 values, but the 5 important features have changed from –

Neighborhood\_Crawfor 1.093888

OverallCond\_9 1.081702

OverallQual\_9 1.081224

GrLivArea 1.078804

OverallQual\_8 1.077316

To

Neighborhood\_Crawfor 1.075603

GrLivArea 1.074821

OverallQual\_8 1.069886

OverallQual\_9 1.064332

Functional\_Typ 1.055627

In case of lasso no major change was observed

## 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 –

The R2 score on test data is 0.91 for Ridge Regression and for Lasso model it is 0.90 so we prefer Ridge regression

## 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?

Functional_Typ	1.057482
OverallQual_7	1.043467
YearRemodAdd	1.038396
Condition1_Norm	1.035013
CentralAir_Y	1.030925

## 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?

Model should be robust for any variation in data and it should not affect its performance much. And it should be generalizable i.e. being able to adapt to new and previously unseen data. We

have to make sure that it doesn't overfit. Such a model will memorize all the patterns in the training data, but fail to generalize to the patterns in test data.