

### 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 and lasso regression is as follows:

alpha for ridge: 4.0

alpha for lasso: 0.0

For ridge model with double value of alpha viz. 8.0, it is observed the coefficients reduced in magnitude.

Since alpha for lasso is 0, there is no expected change here.

After implementing the change, most important predictor variable is **OverallQual**

### 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:

Based on doubling of alpha value, here are few observations:

1. Accuracy for training data though reduced, it remain same for test data
2. RSS value is reduced and also difference between train and test has reduced
3. MSE does not have much observable change

Since variation remain similar for train and test data, model with double the alpha may be preferable.

### 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:

Based on analysis, Ridge model is the best performing model.

First top 5 features for Ridge model are: overallQual, Neighborhood, TotalBsmtSF, GrLivArea and BsmtFinSF1

Next top 5 features are: 1stFlrSF, 2ndFlrSF, FullBath, PoolArea and OverallCond

#### **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:**

A robust model shall have

1. good  $r^2$ \_score on training data as well as unseen data.
2. Low RSS and MSE values

Regarding implication towards accuracy of the model, selection of features and apt hyperparameter tuning may be pivotal. Also, handling of outliers may play an important role. In our assignment, since data itself is small size no outliers are dropped.