

### 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 are,

Ridge Alpha = 5

Lasso Alpha = 0.0001

When we double the value of alpha for our ridge regression no we will take the value of alpha equal to 10 the model will apply more penalty on the curve and try to make the model more generalized that is making model more simpler and no thinking to fit every data of the data set .from the graph we can see that when alpha is 10 we get more error for both test and train. Similarly when we increase the value of alpha for lasso we try to penalize more our model and more coefficient of the variable will reduced to zero, when we increase the value of our  $r^2$  square also decreases.

Important variables in Ridge Regression are ,

1. MSZoning\_RH
2. MSZoning\_FV
3. GarageArea
4. Fireplaces
5. Neighborhood\_Crawfor

Important variables in Lasso Regression are ,

1. GrLivArea
2. OverallQual
3. LotFrontage
4. FullBath
5. OverallCond

### 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:

It is important to regularize coefficients and improve the prediction accuracy also with the decrease in variance, and making the model interpretably.

The  $r^2$ \_score of ridge is slightly higher than lasso for the model so we will choose lasso regression to solve this problem

Thus, I chose Lasso 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?

**Answer:**

1. LotArea
2. FullBath
3. 1stFlrSF
4. ExterCond
5. MSZoning\_RH

### 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:**

The model should be as simple as possible, though its accuracy will decrease but it will be more robust and generalisable. It can be also understood using the Bias-Variance trade-off. The simpler the model the more the bias but less variance and more generalizable. Its implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data i.e. the accuracy does not change much for training and test data.