

## 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:

Optimal value of alpha for ridge: 1

Optimal value of alpha for ridge: 10

After make the double alpha for ridge and lasso i.e. 2 and 20

For Ridge: Coeff values are decreases as alpha will increase.  $r^2$ \_score of train data will slightly drop

For Lasso: As alpha value increased more features removed from model. But  $r^2$ \_score of train data will slightly drop

Important Predictor Variables are: ExterQual\_TA, OverallQual, Exterior1st\_Stone, TotalBsmtSF, GrLivArea

## 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  $r^2$ \_score of lasso is slightly higher than lasso for the test dataset so we will choose lasso regression to solve this problem

## 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. BedroomAbvGr
3. Street\_Pave
4. Exterior2nd\_CBlock
5. ExterQual\_Gd

## 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 generalized so that the test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones which were used during training