

### 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?

Ans :-

a) Optimal Value for Ridge Regression = 20

Optimal Value for Lasso Regression = 100

b) After doubling the alpha :-

For Ridge the  $r^2$  is has decreased by .1 in test as well as train

For Lasso also the  $r^2$  is has decreased by .1 in test as well as train

c)

These are the explaining variable for Ridge

- OverallQual\_9
- Neighborhood\_NoRidge
- Neighborhood\_NridgHt
- GrLivArea
- Neighborhood\_StoneBr
- BsmtExposure\_Gd
- 2ndFlrSF
- Condition1\_Norm
- GarageCars
- Functional\_Typ

These are the explaining variable for lasso

- OverallQual\_9
- OverallQual\_10
- Neighborhood\_NoRidge
- Neighborhood\_StoneBr
- GrLivArea
- Neighborhood\_NridgHt
- OverallQual\_8
- Neighborhood\_Crawfor
- SaleType\_New
- BsmtExposure\_Gd

### 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?

Ans :-

It will depend on the use case, For Example :-

- If we have too large coefficients and reducing them is the primary focus we should go for Ridge.
- If we have multiple variables and want to go for feature selection then we should use Lasso.

### 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?

Ans :-

Now the top 5 variables are :-

- OverallQual\_9
- OverallQual\_10
- Neighborhood\_NoRidge
- Neighborhood\_StoneBr
- RoofMatl\_WdShngl

### 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 can say that the model is robust when if we vary the data the output of the model doesn't affect much on its performance.
- A generalizable model is a kind of model that can predict good in case on unseen and new data as well.
- In such cases we check for the variance in the model correspondingly we check for over fitting, if the model is over fitted then we can say that the model is not much robust and generalized.
- If we will talk about accuracy, a complex model will have a high accuracy and thus will have high variance but as we know robust and generalized models does not have high variance thus we will need to decrease the variance which could be achieved by lasso and ridge.