

### 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:**

The optimal value of alpha for ridge and lasso regression

Ridge Alpha 1

lasso Alpha 10

But in the given dataset calculated values:

Ridge : 4

Lasso : 0.0001

After doubling the lambda :

~ R2 square value of new ridge and new lasso are more than old ridge and old lasso

~RSS value of new ridge and lasso are less than the old one

~MSE value of new ridge and lasso are less than the old

After implementing the changes variables are:

1. MSZoning\_RL
2. MSZoning\_RM
3. GrLivArea
4. OverallQual
5. MSZoning\_FV
6. TotalBsmtSF
7. OverallCond
8. MSZoning\_RH
9. Foundation\_PConc
10. Garagecars

### 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:** The optimal value of lambda in this case

for ridge : 4

For lasso : 0.0001

The  $r^2$  score of ridge is higher than lasso, so will go with Ridge regression in this case.

### Question 3

After building the model, you realized 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:** according to lasso 5 important predictor variables are

MSZoning\_RL, MSZoning\_RM, GrLivArea, MSZoning\_FV, OverallQual .

After processing few steps the 5 new predictors with coefficients will be:

TotalBsmtSF : 0.092219

FullBath: 0.083168

HalfBath: 0.069498

OverallCond: 0.054315

FireplaceQu: 0.04717

### 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?

**Ans:**

- a. The model should be generalized so that the test accuracy is not lesser than the training score.
- b. The model should be accurate for datasets other than the ones which were used during training.
- c. Too much importance should not be given to the outliers so that the accuracy predicted by the model is high.
- d. To ensure that this is not the case, the outliers analysis needs to be done and only those which are relevant to the dataset need to be retained.
- e. Those outliers which it does not make sense to keep must be removed from the dataset.

f. If the model is not robust, It cannot be trusted for predictive analysis