

### 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 lasso = 0.01, ridge = 2.

If we double the value of alpha in ridge, we can see the  $y_{\text{train}}$  values are slightly increasing and  $y_{\text{test}}$  values are slightly decreasing. In case of lasso both the  $y_{\text{train}}$  and  $y_{\text{test}}$  values are slightly decreasing.

Variables : MSZoning\_FV, MSZoning\_RL, Neighborhood\_Crawfor, MSZoning\_RH, MSZoning\_RM, SaleCondition\_Partial, Neighborhood\_StoneBr, GrLivArea, SaleCondition\_Normal, Exterior1st\_BrkFace

### 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 : I'll prefer to go with the lambda for lasso regression as it has many co-efficients near to '0' over ridge.

### 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 : "GrLivArea", "OverallQual", "OverallQual", "TotalBsmtSF", "BsmtFinSF1"

#### Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Ans : To make a model robust and generalizable, first it should be simple. In simple terms more the bias less the variance and more generalizable. The accuracy must not change for test and train models.