## 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 Values:**

1. Optimal Value for Ridge regression – 0.0001
2. Optimal Value for Lasso regression – 0.0001

**Changes Observed on Doubling:**

1. There is no much changes in the metric obtained on doubling the alpha values in ridge and Lasso.
2. We have the Ridge model which performs slightly better than Lasso , however the Lasso model's training accuracy dropped by just 3% compared to the previous model before doubling.

**Important Top-5 Predictors after the change from Lasso model:**

1. TotalBsmtSF
2. OverallQual
3. 2ndFlrSF
4. SaleCondition_Alloca
5. GarageArea

## 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:

1. We obtained the same alpha value for Ridge and Lasso model however we could not comment much on the accuracy as both the models seem to perform better in both training and test data.

2. We prefer Lasso to Ridge as in the case of Lasso it pushes certain coefficients to zero, which makes very easy for the feature selection.

## 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. In the previous run with optimal alpha value we obtained the following top five predictor variables in the lasso model and they are as follows.

    - PoolArea
    - PoolQC_None
    - TotalBsmtSF
    - OverallQual
    - 2ndFlrSF
2. Now if we dropped these five variables and re-run the lasso model for the dataset then we had the following top-five predictor variables for the optimal value alpha=4:
    - BedroomAbvGr
    - GarageArea
    - FullBath
    - OverallCond
    - Neighborhood_StoneBr

## 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:

| | Metric | Linear Regression | Ridge Regression | Lasso Regression |
|---|---|---|---|---|
| 0 | R2 Score (Train) | 0.90 | 0.94 | 0.93 |
| 1 | R2 Score (Test) | 0.81 | 0.85 | 0.87 |
| 2 | RSS (Train) | 16.49 | 9.28 | 10.99 |
| 3 | RSS (Test) | 13.95 | 10.86 | 9.44 |
| 4 | RMSE (Train) | 0.13 | 0.10 | 0.10 |
| 5 | RMSE (Test) | 0.18 | 0.16 | 0.15 |
| 6 | MAE(Train) | 0.09 | 0.07 | 0.08 |
| 7 | MAE(Test) | 0.11 | 0.10 | 0.10 |

1. From the above table, we can confirm that the model has performed well on training data as well as in the test data which means there is no potential overfitting issue.
2. We could infer that the baseline Linear Regression model has 9% difference between test and train accuracy and the regularized Ridge model (9%) and the lasso model (6%) .
3. Since the model able to generalize well on the unseen dataset the model is generalizable for future unseen datapoints and thus the model is robust.