## 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 :

For Lasso : alpha = 0.01

For Ridge : alpha = 2

It can also be seen that for Ridge and lasso, on doubling the alpha value, R-Squared goes Down very slightly.
Root mean squared error has gone up very slightly for both ridge and lasso.
Top 5 most important predictor variables using:

ridge:

| | | |
|---|---|---|
| 50 | Neighborhood_Crawfor | 0.108 |
| 29 | MSZoning_FV | 0.099 |
| 31 | MSZoning_RL | 0.087 |
| 66 | Neighborhood_StoneBr | 0.081 |
| 54 | Neighborhood_MeadowV | -0.087 |

lasso:

| | | |
|---|---|---|
| 4 | OverallQual | 0.120 |
| 13 | GrLivArea | 0.116 |
| 9 | TotalBsmtSF | 0.042 |
| 5 | OverallCond | 0.035 |
| 28 | prAge | -0.083 |

## 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:

Though the model performance by Ridge Regression was better in terms of R2 values of Train and Test,  it is better to use Lasso, since it brings and assigns a zero value to insignificant features, enabling us to choose the predictive variables.

**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:

After excluding top_features = ['prAge', 'GrLivArea', 'OverallQual', 'OverallCond', 'TotalBsmtSF'], then below 5 are the most important now :

| | | |
|---|---|---|
| **7** | BsmtFinSF1 | 0.130 |
| **8** | BsmtFinSF2 | 0.085 |
| **128** | ExterCond_Fa | 0.081 |
| **17** | HalfBath | 0.063 |
| **5** | OverallCond | 0.054 |

**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:

In summary:

- a model is robust when any variation in the data does not affect its performance much.
- a generalizable model is able to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.

The way to ensure model is robust and generalized is :

- to check the test score and ensure training and test set are almost similar.
- This eventually means that training set can reduce a bit to avoid overfitting.
- So the model should not be too complex in order to be robust and generalizable.
- Also outlier analysis and correlation analysis need to be done and only relevant attributes should be retained in the final dataset.

If model is not robust, it cannot be trusted for predictive analysis.

If we look at the from the perspective of accuracy, a too complex model will have a very

high accuracy.

So, to make our model more robust and generalizable:

- we will have to decrease variance which will lead to some bias and this will a cause some decrease in accuracy.
- In general, we have to find balance between model accuracy and complexity.

This could be achieved by Regularization techniques like Ridge Regression and Lasso

Regression