Advanced Regression Assignment

Name: Saurabh Dugdhe

**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?**

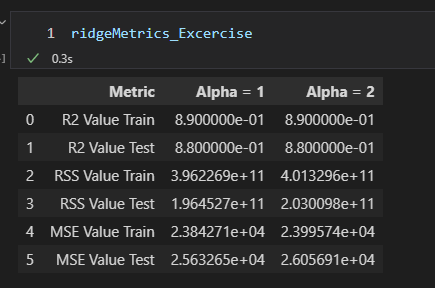
**Ans:**

Optimal value of alpha for Ridge regression is 1.0

Optimal value of alpha for Lasso regression is 7

1. Upon doubling the alpha value for ridge, i.e 2.0

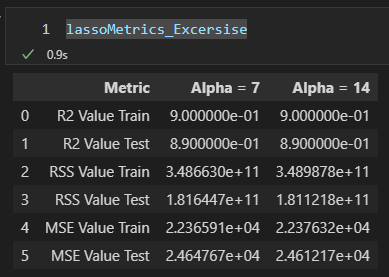
The change in metrics is as shown below.



**The r2 is not impacted much**

1. Upon doubling the alpha value for lasso, i.e 14

The change in metrics is as shown below.

****

**The r2 value is not impacted much**

**Important predictors:**

1. The predictors for Ridge model are:



1. The predictors for Lasso Model:



**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:**

For Ridge regression,

The r2 value for train is 0.89 and for test it is 0.88

For Lasso regression,

The r2 value for train is 9.0 and for test it is 8.9.

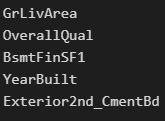
As we can see that the lasso regression performs better. The r2 scores for lasso are better than that of ridge regression. Hence, we will use Lasso regression.

**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?**

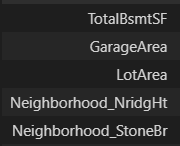
**Ans:**

The top 5 predictors for first Lasso model are:



We removed above 5 predictors and built the lasso model again.

The new top 5 predictors for Lasso model are:



**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 robust and generalisable model does not overfit or underfit. It is able to perform well on test data.

A robust model does not change much if we make small changes in training data. It is able to predict test values with similar accuracy.

Implications: When we make the model generalized and robust, we get similar accuracy on train and test data. The bias-variance trade off needs to be considered for this. We need to have an optimal level of bias and variance so as to make the model more robust.

