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?

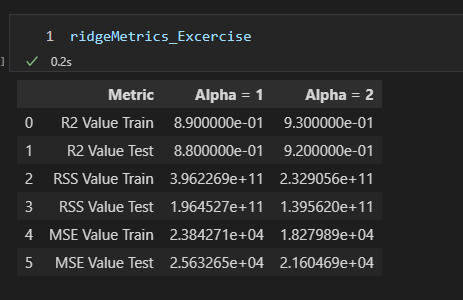
Ans:

Optimal value of alpha for Ridge regression is 1.0

Optimal value of alpha for Lasso regression is 100

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

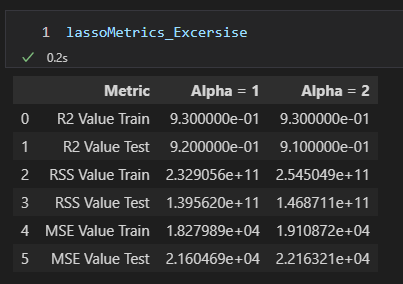
The change in metrics is as shown below.



**The r2 value improves for doubled alpha in ridge.**

1. Upon doubling the alpha value for lasso, i.e 200

The change in metrics is as shown below.



**The r2 value decreases for test of doubled alpha in lasso.**

**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 9.3 and for test it is 9.2.

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?

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?