Advanced Regression Assignment

**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 to 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 Value of alpha for ridge and lasso regression are given as follows -

• Ridge: 10

• Lasso: 0.001

If we choose to double the value of alpha In case of ridge it will lower the coefficients and in case of Lasso the least important features coefficients turning 0.

The most important predictor variable after the change is implemented are those which are significant.

**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 :

Optimal Value of alpha for ridge and lasso regression are:

• ridge: 10

• Lasso: 0.001

As in our case Ridge has better score of 86.6 on test data, its better to go with Ridge model.

**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**:**

On running the same notebook and removing the top 5 significant variables :

We found below variables as next 5 significant.

Lasso:

****

Ridge**:**

****

**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:

A model needs to be made robust and generalizable so that they are not impacted

by outliers in the training data. The model should also be generalisable so that the test

accuracy is not lesser than the training score. The model should be accurate for datasets

other than the ones which were used during training. Too much weightage should not given

to the outliers so that the accuracy predicted by the model is high.

A model needs to be made robust and generalizable so that they are not impacted

by outliers in the training data. The model should also be generalisable so that the test

accuracy is not lesser than the training score. The model should be accurate for datasets

other than the ones which were used during training. Too much weightage should not given

to the outliers so that the accuracy predicted by the model is high.

A model needs to be made robust and generalizable so that they are not impacted

by outliers in the training data. The model should also be generalisable so that the test

accuracy is not lesser than the training score. The model should be accurate for datasets

other than the ones which were used during training.

A model needs to be made robust and generalizable so that they are not impacted

by outliers in the training data. The model should also be generalisable so that the test

accuracy is not lesser than the training score. The model should be accurate for datasets

other than the ones which were used during training.

A model needs to be made more robust and generalisable so that they are not impacted by the outliers in the training data. The model should be accurate for datasets other than the training data