Part-II

Assignment-Advanced Regression

Q-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?

A-1 The optimal value of Ridge and Lasso Regression as obtained from the model is:

Ridge: 0.01

Lasso: 0.0001

b.) If we double the values of alpha for both ridge and lasso:

1. In Ridge, if we double or increase the value of alpha there will be a problem of underfitting in the model. The variance of the model will be less and the bias of the model will be high.

2. In the case of Lasso regression, if we double the value of alpha there will be a decrease in the variance and the bias of the model will increase.

c.) It will be decided based on the coefficients that I will get when the alpha values have been doubled. The ones that will be more significant will be our predictor variables.

Q-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?

A-2 The optimal value of Ridge and Lasso Regression as obtained from the model is:

Ridge: 0.01

Lasso: 0.0001

In **Ridge regression**, R-square on the Train set and Test set came out to be: **0.938 and 0.891**

Whereas, In **Lasso regression** R-square on the Train and Test set came out to be **0.928 and 0.891**. There is a slight dip in the R-square value of the Train Set as compared to the Ridge Regression Train set value.

I'll go ahead with Ridge Regression in this model as it can explain around 94% of the data on the Train set and 89% of the data on the Test set.

Also, the **MSE (Mean Squared Error) on Train and Test** set in Ridge Regression is slightly lower than Lasso Regression.

Ridge Regression:
MSE (Train):0.09
MSE (Test):0.133

Lasso Regression:
MSE (Train):0.10
MSE (Test):0.133

Q-3 After building the model, you realized 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?

A-3 If the 5 most important predictor variables are not present in the lasso model then the next 5 most important predictor variables will be:

Dropping the 5 most important predictor variables and running the notebook again:

1. BedroomAbvGr
2. Exterior2nd_BrkFace
3. Exterior1st_MetalSd
4. Exterior2nd_AsphShn
5. Exterior1st_Plywood

Q-4 How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

A-4 To make a model robust and generalizable we need to carry out the outlier treatment of our dataset with utmost care. This will make sure that there are no irrelevant values present in the data set and we get all the outputs in the Range. Also, we need to Regularize our model so that it has low bias and low variance.

The values that we get for Mean Squared Error should be as low as possible both on the Train and the Test sets. In our models of ridge and Lasso regression the values obtained are:

Ridge Regression:
MSE (Train):0.09
MSE (Test):0.133

Lasso Regression:
MSE (Train):0.10
MSE (Test):0.133

In the above case, Ridge Regression has a slightly better MSE on Train set as and when compared with Lasso Regression.

The accuracy of the model will surely get better when the data is filtered out of all the Null values or say the outliers present in the dataset.