Name – S. Anil

Assignment II – Advance Regression

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?

Ans – The optimal value of alpha for ridge regression is 0.9 and the optimal value of alpha for lasso regression is 0.001.

Changes in the model if we double the value of alpha are as follows –

S.NO	Particulars	Alpha Value	R2_Score – Train data	R2_Score – Test data	Top 5 predictors & Coeff Values
1	Lasso	0.001	0.9022	0.8575	1. MiscVal - 1.520128 2. BsmtHalfBath - 0.541487 3. LowQualFinSF - 0.378263 4. BsmtFullBath - 0.373994 5. HalfBath - 0.331081
2	Lasso	0.002	0.8951	0.8618	1. MiscVal - 1.520128 2. BsmtHalfBath - 0.541487 3. LowQualFinSF - 0.378263 4. BsmtFullBath - 0.373994 5. HalfBath - 0.331081
3	Ridge	0.9	0.9051	0.8517	1. MiscVal 1.482910 2. BsmtHalfBath 0.574493 3. HalfBath 0.421184 4. LowQualFinSF 0.419674 5. BsmtFullBath 0.399218
4	Ridge	1.8	0.9022	0.8566	1. MiscVal 1.224349 2. BsmtHalfBath 0.551467 3. LowQualFinSF 0.399191 4. BsmtFullBath 0.391911 5. HalfBath 0.382626

In the lasso regression by doubling the alpha value there is no change in the most important predictor variables but in case of ridge regression by doubling the alpha value there is a change in the most important predictor variables

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 – I will choose Lasso regression as it will help in feature elimination and as a result the model becomes less complex, robust, and easy to interpret.

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 five important predictor variables after changing the model are as follows –

1. BsmtFinType1

2. Neighborhood_Gilbert

3. LotShape

4. HeatingQC

5. Neighborhood_BrkSide

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 -We can ensure that the model is robust and generalisable by the following points.

1. It must not affect by the outliers in the training data set.

2. Test accuracy is not lesser than the training accuracy.

3. The model should be accurate for the unseen data.

Implications of the same for the accuracy of the model are as follows.

- Outlier analysis to be done to remove the unwanted outliers in the training data set which in turn increases the accuracy of the predictions by the model.