Advanced Linear 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 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 for ridge and lasso are:

- Optimal value of alpha for Ridge Regression : 7
- Optimal value of alpha for Lasso Regression : 0.001

If we choose double the value of alpha for both ridge and lasso:

After doubling the alpha value for both Ridge and Lasso Regression the R2 score for train and set will not change much but there may be some waviness present.

In case of Ridge Regression, the coefficients will be lowered. And in the case of Lasso regression, more less important feature coefficients will turned into zero.

The most important predictor variables after the change is implemented are those which are significant which are shown below:

For Ridge :

	Featuere	Coef
7	TotalBsmtSF	0.312639
3	OverallCond	0.233065
6	BsmtFinSF2	0.155921
4	MasVnrArea	0.127365
50	Neighborhood_IDOTRR	0.094328
45	LotConfig_Inside	0.094298
5	BsmtFinSF1	0.085949
2	OverallQual	0.075793
10	GrLivArea	0.069792
40	LotShape_IR3	0.066194

For Lasso:

	Featuere	Coef
7	TotalBsmtSF	0.324764
3	OverallCond	0.245932
6	BsmtFinSF2	0.151318
4	MasVnrArea	0.129671
50	Neighborhood_IDOTRR	0.095692
5	BsmtFinSF1	0.087591
2	OverallQual	0.069784
40	LotShape_IR3	0.067558
10	GrLivArea	0.067451
23	ScreenPorch	0.057586

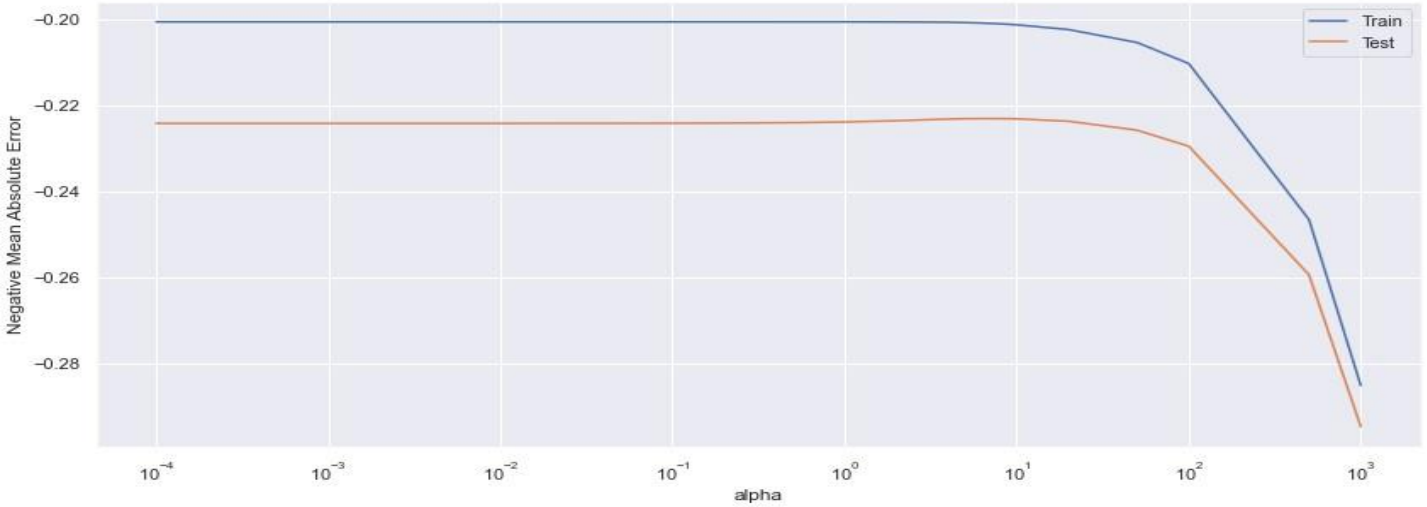
Question:- 2

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:

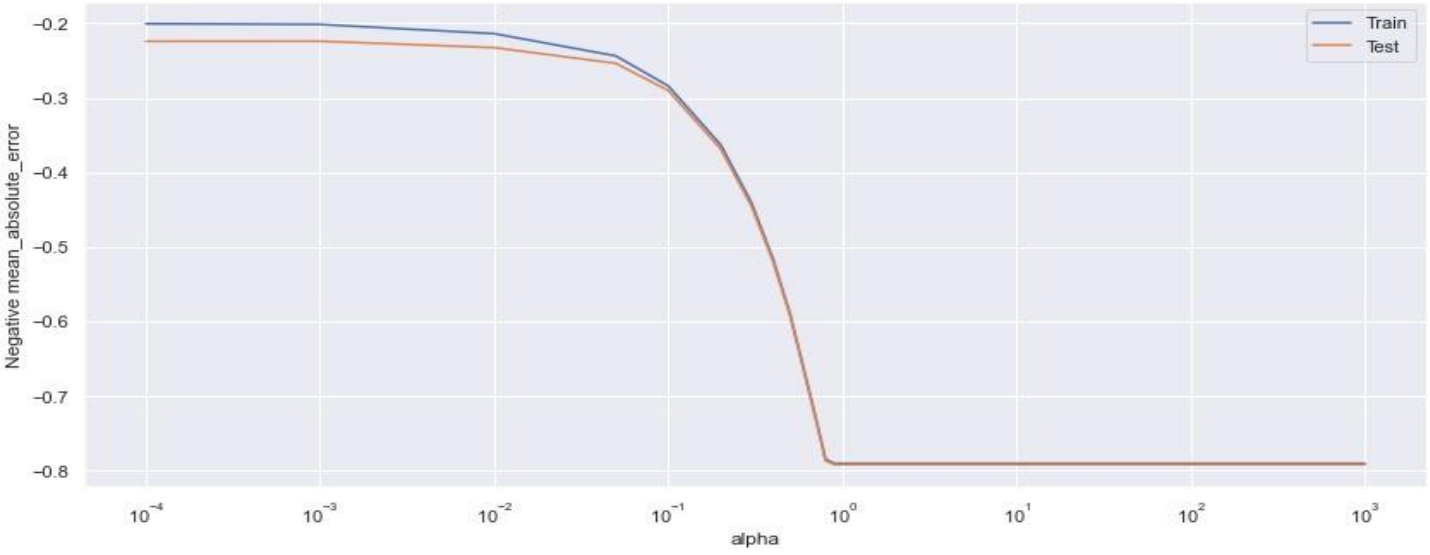
We determined the best optimal value of alpha based on plots and best hyper parameter and chose the value of alpha where we got best train and test score.

For Ridge:



Hence based on the above plot and best hyperparameter we chose the value of alpha as 7 as it gives minimum train and test error.

For Lasso:



Based on above plot and best hyper parameter we chose the value of alpha as 0.001 as it gives minimum train and test error.

Both Ridge and Lasso Regression give the similar results with slight changes in R2 score, MSE. Lasso regression would be a better option it would help in feature elimination and the model will be more robust.

Question:-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?

Ans:

Five most important variables in the lasso model:

	Featuere	Coef
7	TotalBsmtSF	0.324764
3	OverallCond	0.245932
6	BsmtFinSF2	0.151318
4	MasVnrArea	0.129671
50	Neighborhood_IDOTRR	0.095692

After creating another model excluding this attributes, we get following five important variables:

	Features	Coef
2	OverallQual	0.402213
40	LotShape_IR3	0.223138
1	LotArea	0.208306
43	LotConfig_FR2	0.159498
42	LotConfig_CulDSac	0.099702

Question 4:-

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

Ans:

A model is considered as robust and generalizable when the model is not affected by the outliers in the training data and works well for test data, if the model is stable that is it doesn't change drastically when changing the training data.

We can check whether the model is robust and generalizable by testing the model against unseen data and check the various metrics. If there is no drastic changes for metrics of test data we can consider our model to be robust and generalizable.

It's implication in terms of accuracy is that a robust and generalized model will perform equally well on both training and test data that is the accuracy doesn't change much for the training and test data.