

Subjective Questions

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?

Answer:

The optimal value of alpha for ridge and lasso regression are:

Best alpha value for Lasso: {'alpha': 0.001}

Best alpha value for Ridge: {'alpha': 20.0}

After doubling the alpha for both Ridge and Lasso regression the R2 score of train, test data set for both ridge and lasso will not change much but some waviness will be present, and the coefficients of both regressions will undergo some change.

The most important predictor variables after the change implemented are:

For Ridge:

	Feaure	Coef
7	TotalBsmtSF	0.222
2	OverallQual	0.210
8	GrLivArea	0.170
53	Neighborhood_Crawfor	0.130
3	OverallCond	0.124
45	MSZoning_FV	0.111
32	MSSubClass_70	0.103
1	LotArea	0.096
63	Neighborhood_NridgHt	0.090
102	Exterior1st_BrkFace	0.085

For Lasso:

	Featuere	Coef
7	TotalBsmtSF	0.292
45	MSZoning_FV	0.256
53	Neighborhood_Crawfor	0.246
2	OverallQual	0.215
8	GrLivArea	0.195
63	Neighborhood_NridgHt	0.186
32	MSSubClass_70	0.151
62	Neighborhood_NoRidge	0.145
102	Exterior1st_BrkFace	0.137
3	OverallCond	0.134

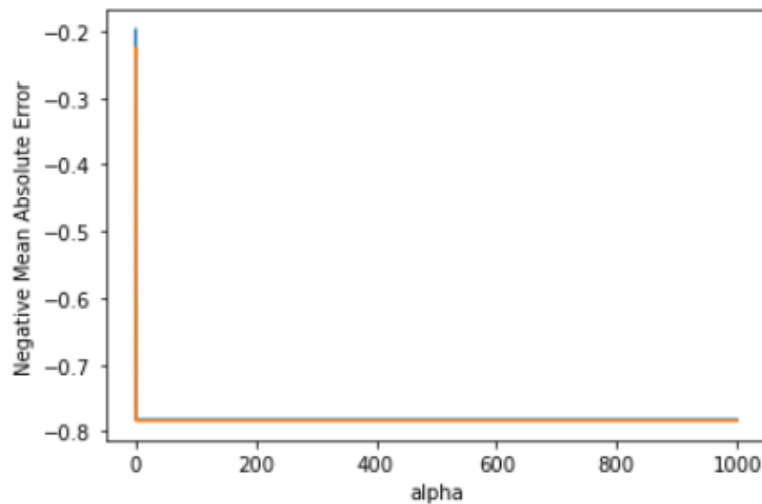
Question 2

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:

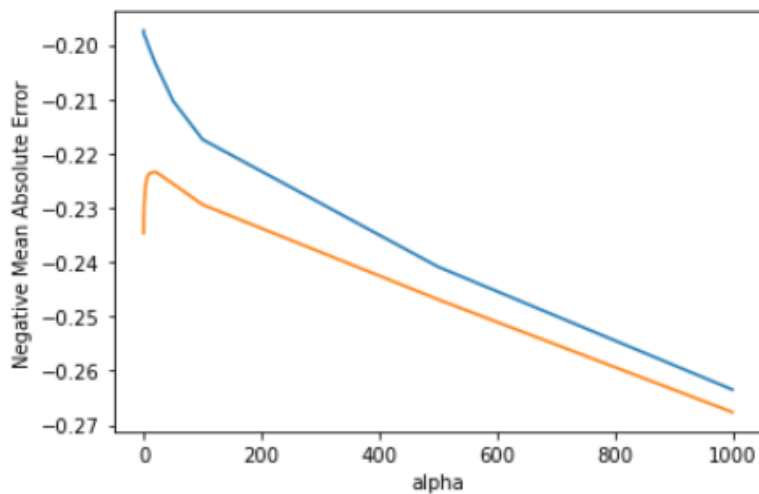
We consider the optimal value for Ridge and Lasso regression based on plots and best hyperparameter. And chose a value of alpha where we have good training as well as the test score.

Lasso Regression Plot:



Therefore, based on the plot and hyperparameter we choose the value of 0.001 lambda/alpha for Lasso regression, because it has the best train and test score.

Ridge Regression Plot:



Therefore, based on the plot and hyperparameter we choose the value of 100 lambda/alpha for Lasso regression, because it has the best train and test score.

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?

Answer:

After Building the model the five most important predictor variables in the lasso model are:

	Featuere	Coef
7	TotalBsmtSF	0.222
2	OverallQual	0.210
8	GrLivArea	0.170
53	Neighborhood_Crawfor	0.130
3	OverallCond	0.124

Now by recreating another model excluding the five most important predictor variables then the most important variables are:

	Featuere	Coef
2	OverallQual	0.197
62	Neighborhood_NoRidge	0.190
8	GrLivArea	0.190
32	MSSubClass_70	0.189
102	Exterior1st_BrkFace	0.182

Question 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?

Answer:

A model is considered to be robust if the model is stable, i.e., does not change drastically upon changing the training set. The model is considered generalizable if it does not overfits the training data and works well with new data.

Its implication in terms of accuracy is that a robust and generalizable model will perform equally well on both training and test data i.e., the accuracy does not change much for training and test data.

