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 double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Ans:

Ridge Regression: The model built on Ridge regression at an optimal value of alpha has:

- Optimal value of alpha = 20
- R2 value of Train Data = 0.90
- R2 value of Test Data = 0.88

Lasso Regression: The model built on Lasso regression at an optimal value of alpha has:

- Optimal value of alpha = 0.001
- R2 value of Train Data = 0.91
- R2 value of Test Data = 0.89

In both lasso and ridge regression models, the effect of doubling alpha has pushed the coefficients closer to 0 (for both positive and negative coefficients). A few of the predictor variables have changed the order of importance. In case of the ridge regression, the first three predictor variables are in the same position where as in lasso most of the important predictor variables retain the order.

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 Regression:

Best value of alpha =20

	Feature	Coef
7	TotalBsmtSF	0.254
2	OverallQual	0.206
8	GrLivArea	0.180
53	Neighborhood_Crawfor	0.171
45	MSZoning_FV	0.163
32	MSSubClass_70	0.138
3	OverallCond	0.129
63	Neighborhood_NridgHt	0.124
102	Exterior1st_BrkFace	0.116
62	Neighborhood_NoRidge	0.108

Double the value of alpha =40

	Feature	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 Regression:

Best value of alpha = 0.001

	Feature	Coef
7	TotalBsmtSF	0.304
45	MSZoning_FV	0.292
53	Neighborhood_Crawfor	0.242
63	Neighborhood_NridgHt	0.209
68	Neighborhood_StoneBr	0.205
2	OverallQual	0.197
8	GrLivArea	0.190
32	MSSubClass_70	0.189
62	Neighborhood_NoRidge	0.189
102	Exterior1st_BrkFace	0.182

Best value of alpha = 0.002

	Feature	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	0.138
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?

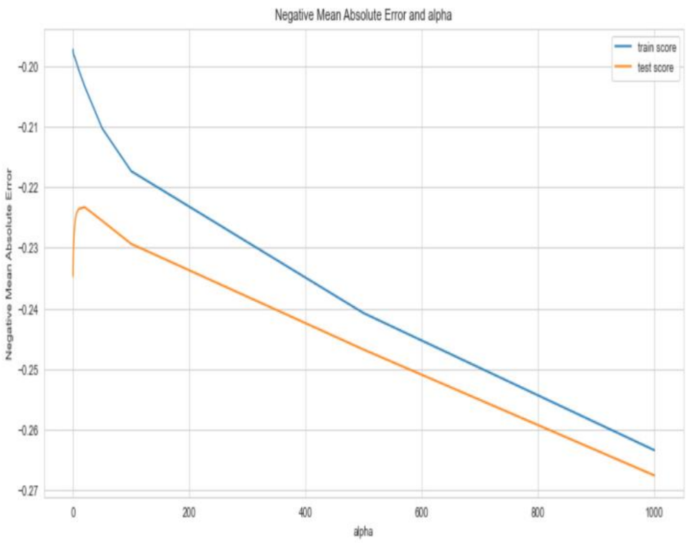
Ans:

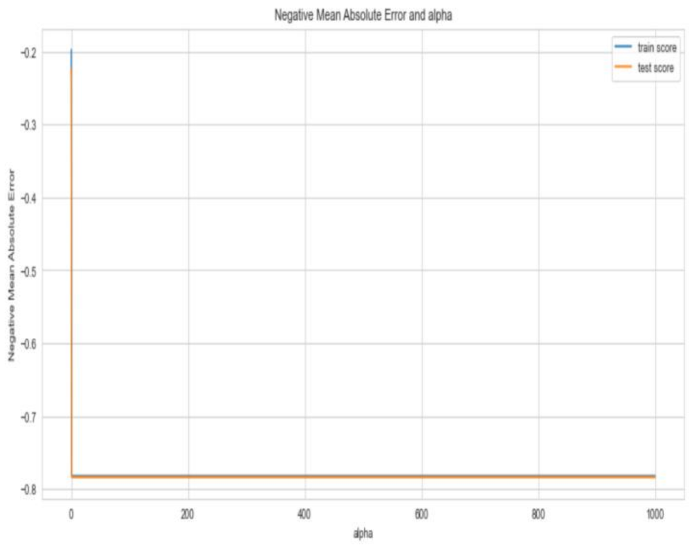
Ridge Regression: Ridge regression adds “squared magnitude” of coefficient as penalty term to the loss function. This regression regularization parameter is a scalar and it enforces the β coefficients to be lower but not zero i.e. it will not eliminate the features that are irrelevant but forces their coefficients close to zero making their impact on the model less. This technique works very well for overfitting issues.
Lasso Regression: Lasso Regression adds “absolute value of magnitude” of coefficient as penalty term to the loss function. This regression regularization parameter is an absolute value unlike ridge but this difference has a huge impact on the trade-off. Lasso is better by reducing the β coefficients and also makes them zero if the feature is irrelevant to the model.

The key difference being Lasso shrinks the less important features coefficients to zero thus, removing irrelevant features. So, this works well for feature selection in case we have large features.

In the model we built using Ridge and Lasso the overall accuracy for both Ridge and Lasso are almost same but since the model we built has large number of features including dummies we can prefer Lasso.

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.





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.

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?

Ans:

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

	Feature	Coef
7	TotalBsmtSF	0.304
45	MSZoning_FV	0.292
53	Neighborhood_Crawfor	0.242
63	Neighborhood_NridgHt	0.209
68	Neighborhood_StoneBr	0.205

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

	Feature	Coef
97	Exterior1st_BrkFace	0.230
7	GrLivArea	0.221
2	OverallQual	0.220
81	HouseStyle_1Story	0.135
31	MSSubClass_70	0.132

Question 4

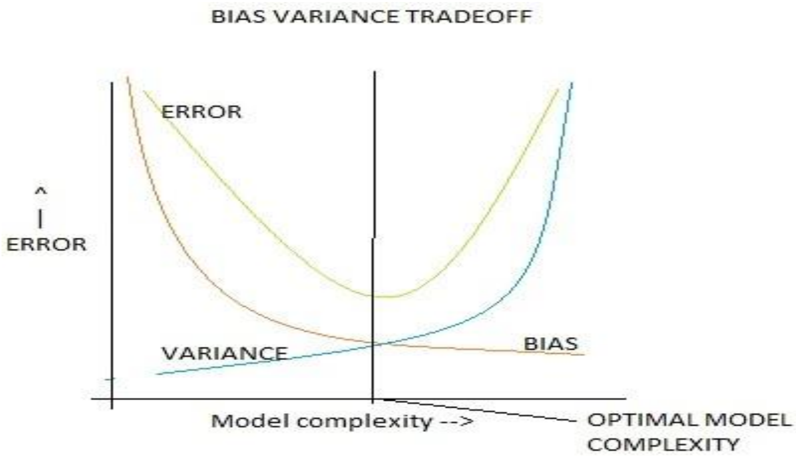
How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

Ans:

A model is said to be generalisable if it has not been overfitting the training data. In other words it does not memorize the training data. When the model comes across data that is different from what it has been trained on, it should still give a reasonable response with acceptable error. The model is considered robust if its results are consistently accurate even if some of the input variables changed.

A model can be made generalizable and robust by striking a balance between overfitting, underfitting and accuracy. Regularization is one such technique that reduces the overfitting by penalizing the coefficients that are large. The model has to be at that complexity level that can recognise the underlying patterns but generalised enough not to memorize the training data. Simpler models are more robust and generalisable.
The implications of keeping a model robust and generalisable does bring down the accuracy score on the training data but will be more consistent on the test set. This is because we compromise on the complexity of the model to make it more generalisable.

The below figure shows the bias and variance trade off with respect to the model complexity. The optimal complexity is when the model is having enough bias to be generalised and enough variance where the model gives the least error.



The below figure shows the bias and variance trade off with respect to the regularization parameter. The optimal value is when the model is having enough bias to be generalised and enough variance where the model gives the least error. The cross validation set is used to minimize the error for the model while choosing the best parameter