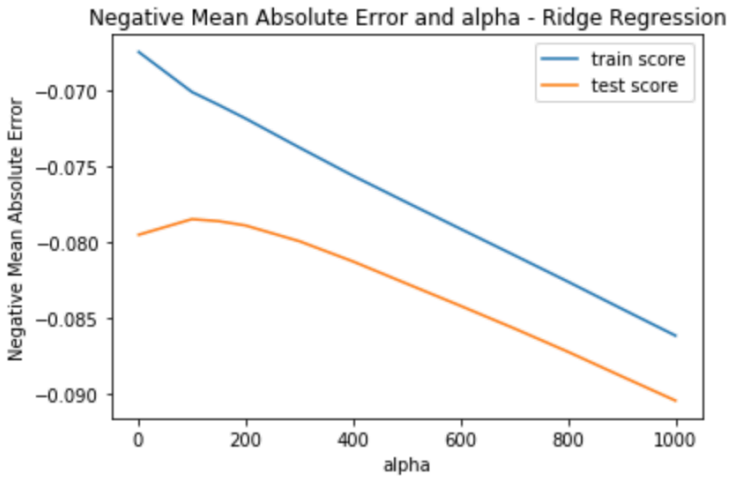
Advanced Regression Assignment - Part II (Subjective)

By Saiyana Ramisetty

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: The optimal value of alpha for ridge and lasso regression is where the mean squared/absolute error remains constant or decreases with changes in alpha values. Alpha values must be chosen wisely because, higher values alpha reduce overfitting and significantly high values can cause undercutting too. Also with the increase in alpha values, the model complexity reduces.



If we consider the above graph, we can clearly see that the test score error decreases nearly from alpha = 100. Thus the optimal value of alpha can be chosen as 100 in the above case.

If the alpha value is doubled for ridge and lasso, there will be a decrease in the R squared value and increase in the Mean square error. When we increase the value of alpha to a higher value, the coefficients tend to approach zero. Thus by doubling the alpha value the value of coefficients might decrease and approach towards zero. However there might not be change in the most important predictor variable after this change (but the coefficient of the variable decreases)

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: After we obtain the optimal value of lambda for ridge and lasso regression. We need to consider the performance metrics such as R-squared values for train and test, as well as mean squared error to evaluate the model performance. The performance metrics that I obtained after finding the optimal lambda value for both the regression models are as below.

Regression	alpha	Train R-square	Test R-square	Mean Square Error
Ridge	100	0.924	0.901	0.014
Lasso	0.001	0.928	0.909	0.013

The performance metrics are similar for both the regression techniques, however the Lasso Regression model is performing better with slightly better evaluations than Ridge. Lasso can also be chosen over Ridge because it performs Feature Selection and Regularisation, where are Ridge only performs Regularisation.
Lasso gives us few number of predictor variables (making the not significant variables coefficients zero), thus we obtain a better set of predictor variables. Where are Ridge Regression assigns coefficients to the attributes without performing feature reduction. Lasso regression optimises the predictors and there by making the model less complex. Unless the performance measures of Ridge are way better than Lasso, the Lasso regression model is a better choice for selection. That is why I am choosing the Lasso model over Ridge model.

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: Lasso (Least Absolute Shrinkage and selection operator) is a regression analysis method that performs both variable selection and regularisation in order to enhance the prediction of the model.

The Lasso method puts a constraint on the sum of the absolute values of the parameters. In order to do so the lasso method applies a shrinking process where is penalises the coefficients of the regression variables shrinking some of them to zero (which are not significant, thus performing the feature selection)

After we build the Lasso model, we get the coefficients of the features used in the dataset, the coefficients of the insignificant features will be returned as 0 (which specify that the feature is insignificant or not useful). Thus for obtaining the best features, we have to get the features which have high magnitude of their coefficients (this implies that the change in unit of the dependent feature will reflect the change in the dependent variable, thus higher the value of coefficient, higher is the impact on the dependent variable).

	Feature	Coefficient
13	BsmtExposure	0.119
2	LotFrontage	0.073
98	RRAn - Condition2	-0.070
3	LotArea	0.048
9	ExterQual	0.039
...	...	...
66	Crawfor - Neighborhood	-0.000
67	Edwards - Neighborhood	-0.000
68	Gilbert - Neighborhood	-0.000
70	MeadowV - Neighborhood	-0.000
100	1Fam - BldgType	-0.000

Consider the above case, lasso gives the coefficient for the variable (Crawfor - Neighbourhood) as zero specifying that the feature is not useful in predicting the output variable. And the coefficient for (BsmtExposure) is greater implying that it is a better predictor variable.

If the top 5 important predictor variables are removed, we can considered the next top 5 predictor variables or can choose the parent variable (if the removed variable is a derived attribute). We can repeat the process of feature selection using the magnitude of the coefficients method again after rebuilding the model to obtain the five most important predictor variables.

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 considered robust and generalisable if the output variable is predicted correctly even if one or more assumptions made on the predictor variable changes due to unseen circumstances. This can also be explained using Bias-Variance trade off.
Thus the model should not be very complex or very simple. And the robustness of the model highly depends on outliers in the data. Also more weightage must not be given to outlier removal because, the model built this way, can have very drastic changes if it encounters an outlier in test data.

Few of the changes we can make to ensure that the model is robust are:
1. Using a model which is resistant to outliers
2. Using a more robust error metric
3. Transforming the data
4. Removing outliers
5. Considering robust metrics for representation (Ex: Median instead of Mean)

We need to look at the robust estimators, which minimise the sum of the absolute values of the errors instead of the sum of squares making it more resistant to the outliers.

Similarly considering the accuracy of the model, the model must be robust and not complex enough to make predictions such that the test accuracy is nearly equal or greater than the train accuracy. The model should be generalisable so that the test accuracy is greater than the training accuracy. That means, it must be accurate for the datasets other than the ones which are not used during training.