

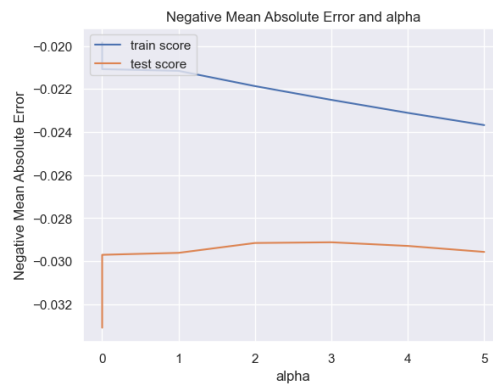
## Assignment-based Subjective Questions

### Question 1

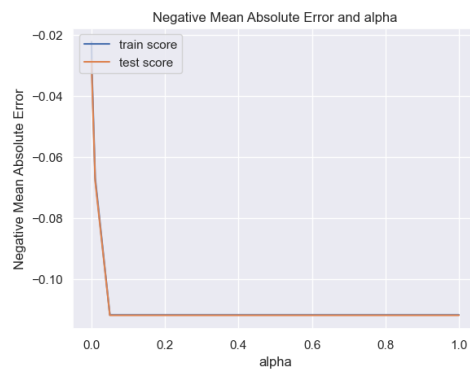
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

**Answer:**

Optimal value for Ridge: 3



Optimal value for Lasso: 0.0001



When we make the above alpha values to double, the algorithm will penalize more and try to generalize the model. Similarly, in Lasso it will try to reduce more number of coefficients to zero and R2 value also decreases.

The coefficients are more penalized and below is updated score for Ridge Regression and Lasso regression.

**Ridge Regression:**

	Ridge Regression
R2 Score(Train)	94.48
R2 Score(Test)	89.04

**Lasso Regression:**

	Lasso Regression
R2 Score(Train)	94.4
R2 Score(Test)	90.5

After changing alpha to double the most predictor variables for Ridge:

GrLivArea
OverallQual
BsmtFinSF1
OverallCond
1stFlrSF

After changing alpha to double the most predictor variables for Lasso:

GrLivArea
OverallQual
BsmtFinSF1
OverallCond
GarageArea

## 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:** While model prediction and evaluation for Ridge and Lasso regression, as it is noticed that in Lasso regression R2 score for Test data is closer to R2 score of train data as compared to Ridge regression.

The RMSE is also better in Lasso regression, and it has simpler equations because of less variables.

So, its better to choose Lasso regression and use further for analysis.

### 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?

**Answer:**

After excluding the 5 most important predictor variables and on re-training the Lasso regression. The new generated model has below 5 variables which are most important.

1stFlrSF
2ndFlrSF
GarageArea
Neighborhood_Crawfor
Neighborhood_StoneBr

### 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?

Model is said to be more robust if it's able to identify and generalize the patterns in the data than memorizing them. Regularization methods would help avoid overfitting then by making the model generalizable.

As we can see the R2 for Test and train dataset both are good. The accuracy of the model is quite good.

Generalization refers to your model's ability to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.

The model should be as simple as possible. Let's see about Bias-Variance trade off. The simpler the model the more the Bias, but less variance and more generalizable. It's implication in terms of accuracy is that a robust and generalizable model will perform equally on both training and test data i.e. the accuracy doesn't change much for training and test data.