## Question 1

**What is the optimal value of alpha for ridge and lasso regression?**

Ans. Optimal alpha value is varied from dataset to dataset given it take its values from 0 to infinity

In the assignment given house price prediction the optimal values are

Lasso Regression – 0.8

Ridge Regression – 2

**What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?**

Ans. When we double the alpha values for Lasso and Ridge regression model

There is no significant change in R2 Score, MSE but there is changes can be observed in train and test datasets.

**What will be the most important predictor variables after the change is implemented?**

Ans. The most important predictor variables are

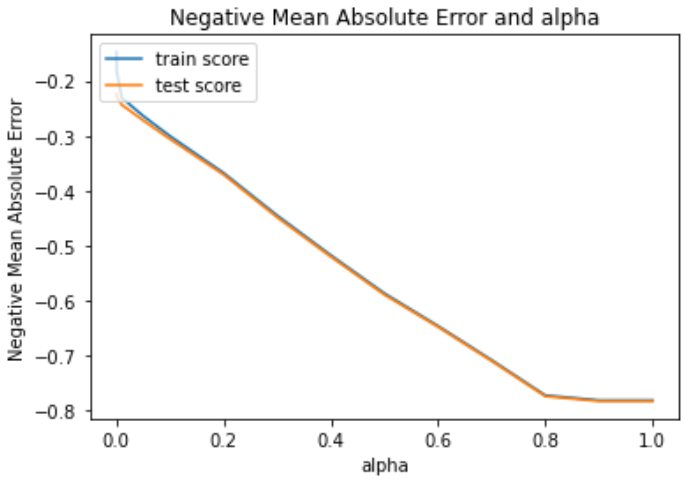
| | Variable | Coeff |
|---|---|---|
| 4 | OverallQual | 0.293 |
| 16 | GrLivArea | 0.248 |
| 6 | YearBuilt | 0.150 |
| 26 | GarageCars | 0.123 |
| 5 | OverallCond | 0.107 |
| 17 | BsmtFullBath | 0.086 |
| 80 | Condition1_Norm | 0.085 |
| 24 | Fireplaces | 0.074 |
| 7 | YearRemodAdd | 0.071 |
| 12 | TotalBsmtSF | 0.062 |
| 209 | GarageType_Attchd | 0.046 |
| 23 | TotRmsAbvGrd | 0.039 |
| 19 | FullBath | 0.037 |
| 3 | LotArea | 0.036 |
| 157 | Foundation_PConc | 0.032 |

## 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?**

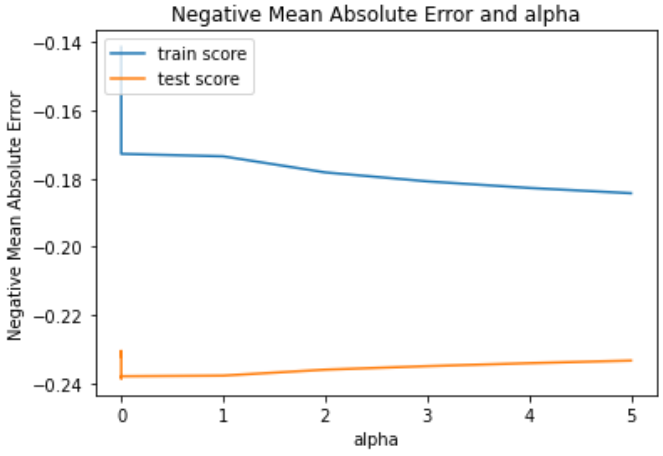
Ans. Observing the both the models lasso and ridge regression with their optimal values of alpha – 0.8 and 2.

From the plots Lasso Regression




```
0.8762843000046499
0.867837466901328
RMSE : 0.37194250305925103
```

And Ridge Regression




```
0.9307475290237582
0.8889966110785108
RMSE : 0.3408707619544575
```

We can say that Lasso Regression does better than the Ridge regression also from the results of R-Square values of lasso and ridge, Lasso Regression is doing well.

So I have choose the **Lasso Regression**.

## 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. The five most important predictor variables are 'OverallQual', 'GrLivArea', 'YearBuilt', 'GarageCars', 'OverallCond', if these variables are not available in incoming data.

By creating a new model the first five important variables will be

BsmtFullBath',

'Condition1_Norm',

'Fireplaces',

'YearRemodAdd',

'TotalBsmtSF'.

## Question 4

**How can you make sure that a model is robust and generalizable?**

Ans. To make sure that a model is robust and generalizable the model has to trained in such a way that

- When a model can learn and understand the underlying hidden pattern of the data in train dataset so that it can predict on the test and as well as the data when given in real-time.
- Scores of Training and Testing data needs to similar (not exactly equal).
- The independent variable/ features should have a good coefficient values

If the model can explain the three points we can say that Model is Robust and Generalizable.

**What are the implications of the same for the accuracy of the model and why?**

Ans.

1. Get more Data – Gather as much of the data possible so our model can learn from the data
2. Treating Null values – Gather the data for null values from the client/ drop if and only the data has greater than 40% of null values.
3. Imputation of data - Proper Techniques has to be used mean/media/mode/contact values while imputation of missing values better have perspective from business end as well.
4. Outlier Treatment - Proper Techniques has to be used mean/media/mode/contact values has to be used.

5. Transforming the variables – Use either of the Standard/Minmax Scaler as per the data and model you are using.
6. Feature Selection – choose the higher values coefficient variables and also try to drevie new columns if possible from the given data.
7. Implementing correct Model – Use the different algorithms linear, polynomial Regression model understand the underlying patterns in the data accordingly build the right model.
8. Cross Validation – Sometime we might get higher accuracy then we have to perform cross validation.