**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 of alpha for ridge regression : 10.

Optimal value of alpha for lasso regression : 0.003

On doubling the value of alpha , the model accuracy is decreased but model will be more simpler.

Most important predictor variable is "GrLivArea" (Above grade (ground) living area square feet) for both ridge and lasso.

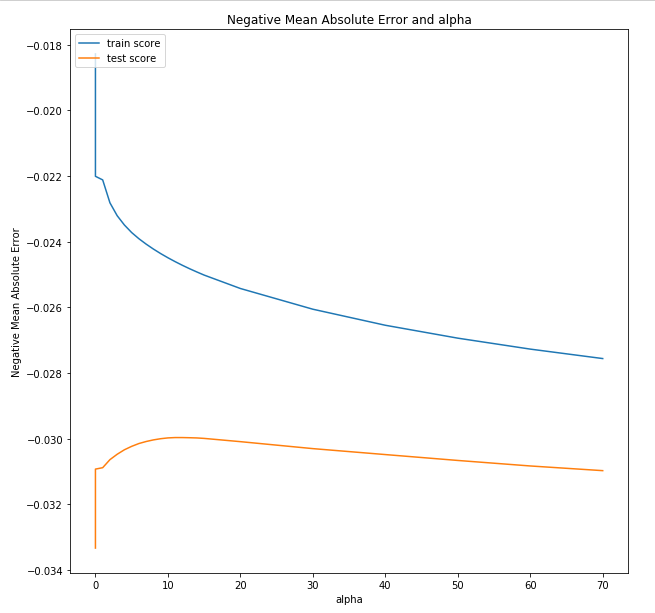
**Question 2**

**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 :**

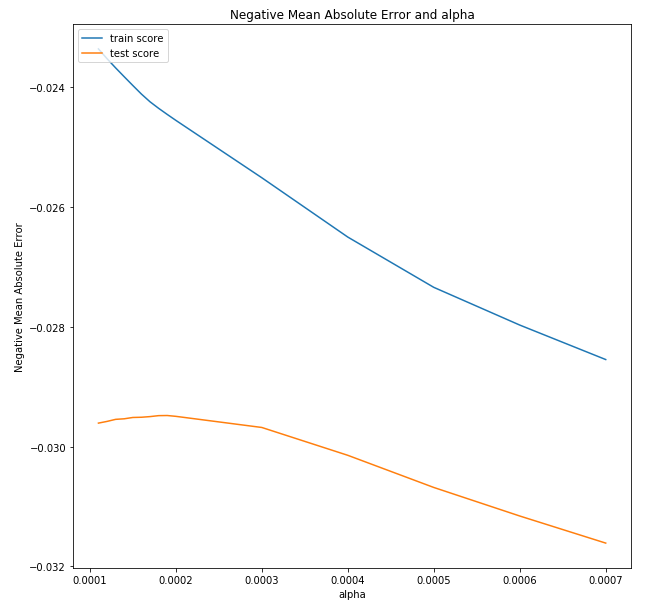
For optimal value of lambda, we can check graph of negative mean absolute error

**Ridge Regression**



The value of error starts dipping at alpha = 10 . Hence, 10 is the optimal alpha to use.

**Lasso Regression**



The value of error starts dipping sharply at alpha = 0.003 . Hence, 0.003 is the optimal alpha to use.

Based on the graphs and alpha values, lasso will be a better option because it helps in feature selection and coefficients are small as compared to ridge 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?**

**Answer :**

Top 5 features after removing are : TotalBsmtSF, GarageCars, 2ndFlrSF, Fireplaces

**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 :**

We can make sore model is robust and generalizable by doing the following:

1. Outlier handling : we have to make sure that outliers in data are handled properly. There are many ways to handle it. One way is to remove them. This is not a good way as it may decrease the predictive power of model or it the feature has some business implications. Another way is to transform the feature so as to decrease the variance between.

2. Regularization : These techniques ensures that the model does not overfit and remain generalizable by simplifying the model.Ridge (L2) and Lasso (L1) are two such techniques for regularization.

Wj = Coefficient vector

Lasso ( L2) has an additional benefit of feature selection. It makes the coefficients of features of low importance to 0.We have to tune the regularization parameter lambda to that the model becomes simple and generalizable without sacrificing too much accuracy.

Cross validation : We can check if the model is generalizable using cross validation. K - fold cross validation is one of the cross validation technique that is widely used. The test data is divides into k parts. The model is created on k-1 parts and tested on the remaining part. This process is repeated for every part (fold) of the data.

Model Selection: Model has to be selected such that it is least susceptible to outliers.