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

The optimal values of alpha for the Ridge and Lasso regression turned out to be 20.0 for Ridge and 0.001 for Lasso.

In case if we choose to double the value of the alpha. Which is 40 for Ridge and 0.002 for Lasso.

Ridge at alpha=20.0

Feature	Coef
Exterior1st_AsphShn	0.197435
Neighborhood_NridgHt	0.190890
Neighborhood_OldTown	0.184317
OverallCond	0.181424
Neighborhood_Edwards	0.166024
BsmtFullBath	0.162064
LowQualFinSF	0.149367
Condition2_PosA	0.143683
BsmtExposure_Mn	0.139312
Neighborhood_Timber	0.135271

Ridge at alpha=40.0

Feature	Coef
OverallCond	0.190547
BsmtFullBath	0.153348
Neighborhood_OldTown	0.143147
LowQualFinSF	0.132093
Neighborhood_NridgHt	0.131426
Neighborhood_Edwards	0.121695
BsmtExposure_Mn	0.112331
Exterior1st_AsphShn	0.110910
1stFlrSF	0.109086
Condition1_PosA	0.105090

Lasso at alpha=0.001

Feature	Coef
Exterior1st_AsphShn	1.034098
Neighborhood_NridgHt	0.317291
Neighborhood_Edwards	0.298146
Neighborhood_Timber	0.297077
BsmtFullBath	0.296306
Neighborhood_OldTown	0.252536
SaleType_Oth	0.214646
BsmtExposure_Mn	0.199231
Exterior1st_CBlock	0.173833
LotConfig_FR2	0.173664

Lasso at alpha=0.002

Feature	Coef
Exterior1st_AsphShn	0.762878
Neighborhood_NridgHt	0.292755
BsmtFullBath	0.286954
Neighborhood_Edwards	0.271456
Neighborhood_OldTown	0.246467
Neighborhood_Timber	0.223855
SaleType_Oth	0.214186
BsmtExposure_Mn	0.177480
OverallCond	0.165621
Exterior1st_CBlock	0.153800

As evident from the above pictures doubling of the alpha values for both Ridge and Lasso have made the most important variables coefficients change. In both the above scenarios majority of the important variables remained but their coefficients are impacted.

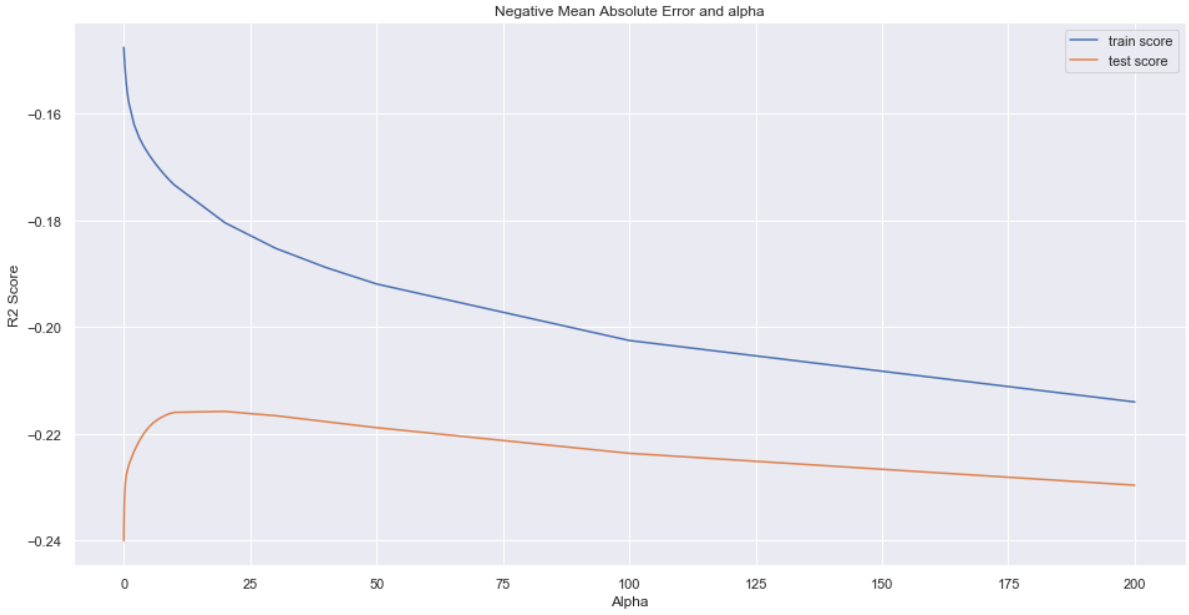
Question 2

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?

In the above case both Ridge and Lasso did a pretty good job in identifying the driving predictors. But I would go with Lasso regression since lasso provides me with feature selection while building the models by making the non-significant feature coefficients to 0. Ensuring only the significant coefficients remain in the model.

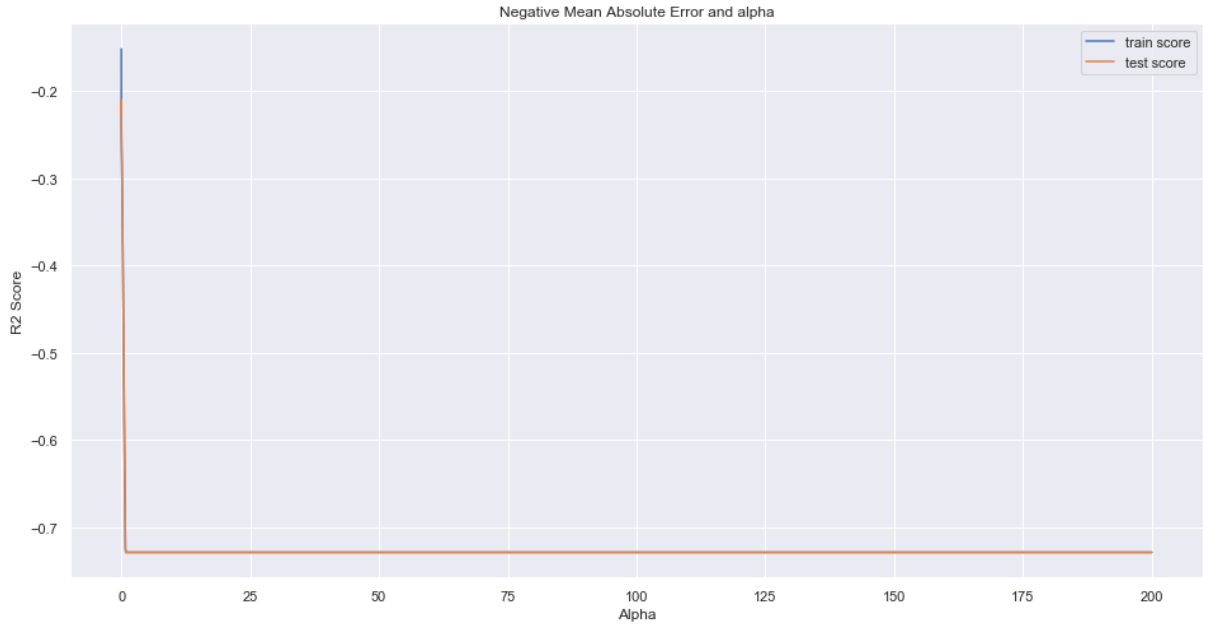
The below ridge regression plot shows the test R2 value increases at first and then slowly tends to a bit lower value and n the training R2 keeps decreasing. So, in accordance with bias and variance tradeoff I choose an alpha value which defines the model at a point which is not too complex nor too simple.

Alpha=20.0



Lasso regression has both test and train R2 scores dropping. So, in order to find the best model the alpha has to be chosen at very low value.

Alpha=0.001



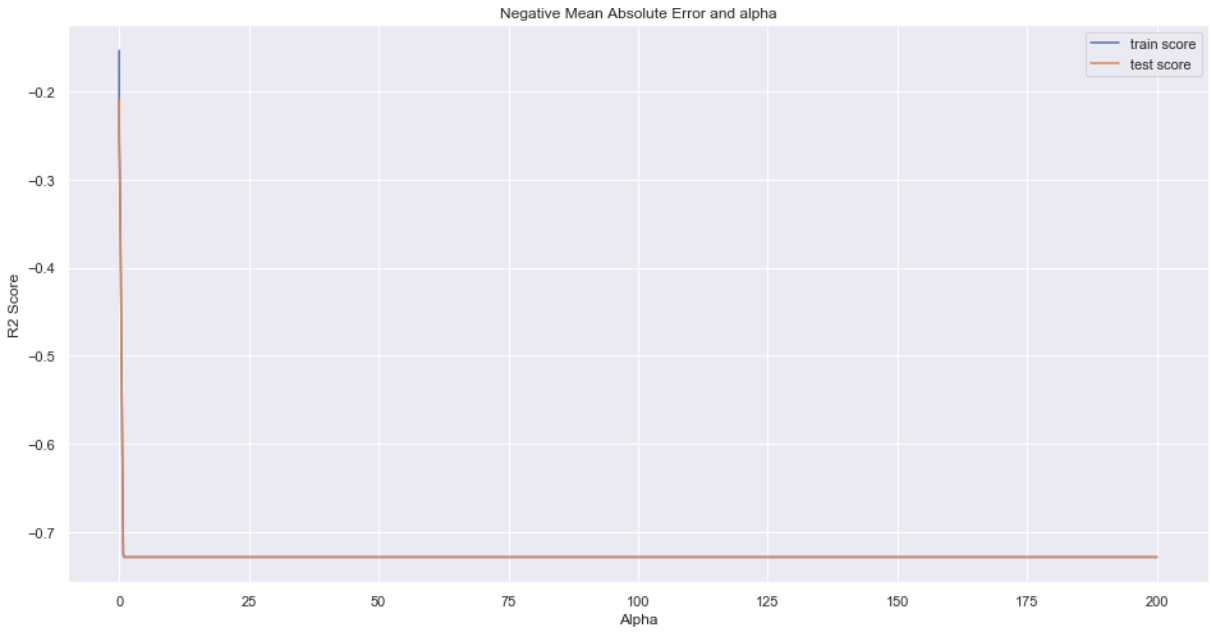
Question 3

After building the model, you realized 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?

Excluding the top five predictors and performing the lasso regression.

The best alpha optimal value came to be 0.001 and below are the top five features.

Feature	Coef
RoofMatl_Roll	0.969639
Neighborhood_Crawfor	0.312363
BsmtFullBath	0.279358
Neighborhood_NWAmes	0.259050
SaleType_Con	0.240119



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?

A Model must be simple to be robust and generalizable. A simple model is not sensitive to the specifics of the training data like complex models. Complex models tend to change wildly with changes in the training data set. The problem with complex models is they overfit the data which will give very bad outputs when tested on unseen data.

So, in other words complex model gives the accuracy of the fit but it fails badly while performing on the unseen data. While simple model irrespective of the training errors made does a good job on unseen data.

Then again simple models have low variance, high bias and complex models have low bias, high variance. There should be a tradeoff between this usually called as bias variance tradeoff. Where we thrive to find the delicate point to ensure model is not too naïve and complex.