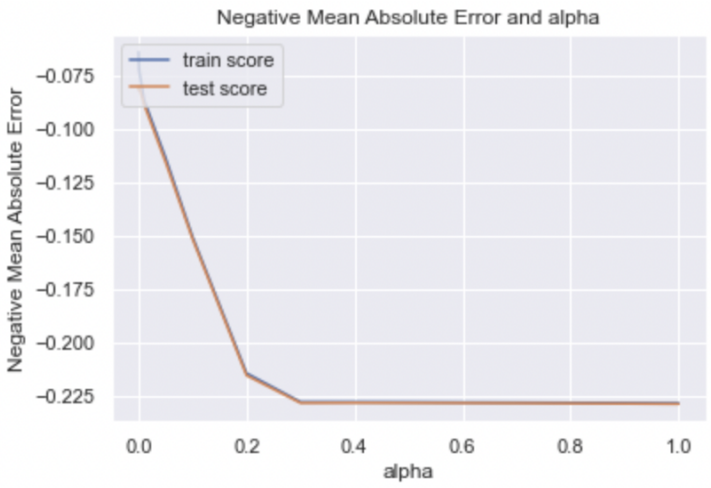
Q1. What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Let's observe the following graphs (*extracted from my workbook*)
Lasso:



Ridge:

Based on that, the optimal value for Ridge and Lasso are as follows:

Ridge	2
Lasso	0.3

Original Predictors for Ridge & Lasso :

Ridge:

- MSZoning_RH

- MSZoning_FV
- MSZoning_RL
- MSZoning_RM
- Neighborhood_Crawfor

Lasso:

- GrLivArea
- OverallQual
- TotalBsmtSF
- OverallCond
- GarageArea

After Doubling :

It is observed that the predictors remain the same but the coefficients have changed.

Q2. 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?

- Ridge has performed better in terms of R2 values of train and test than lasso.
- Lasso is enabling us to choose the predictive variables by assigning a zero value to insignificant features.
- Either can be chosen but based on the principle to always stick to a simple and robust model, I would choose Lasso.

Q3. 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?

If we remove the five most important variables predicted by Lasso then the next set of 5 most important predictors are:

- BsmtFinSF1: Type 1 finished square feet
- LotArea: Lot size in square feet
- LotFrontage: Linear feet of street connected to property
- Fireplaces: Number of fireplaces
- BsmtFullBath: Basement full of bathrooms.

Q4. 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?

- A model must always be robust otherwise it can never be relied upon for predictive analysis as the test data will always be new.

- To ensure that a model is robust and generalizable we must always ensure that the model is accurate for new test data sets. It must be tested thoroughly with multiple test data sets.
- We should also ensure that the test accuracy of the generated model is not less than the training score.
- We should handle outliers in the datasets but at the same time we should not give too much importance and start manipulating outliers such that we decarese the prediction accuracy of the model.