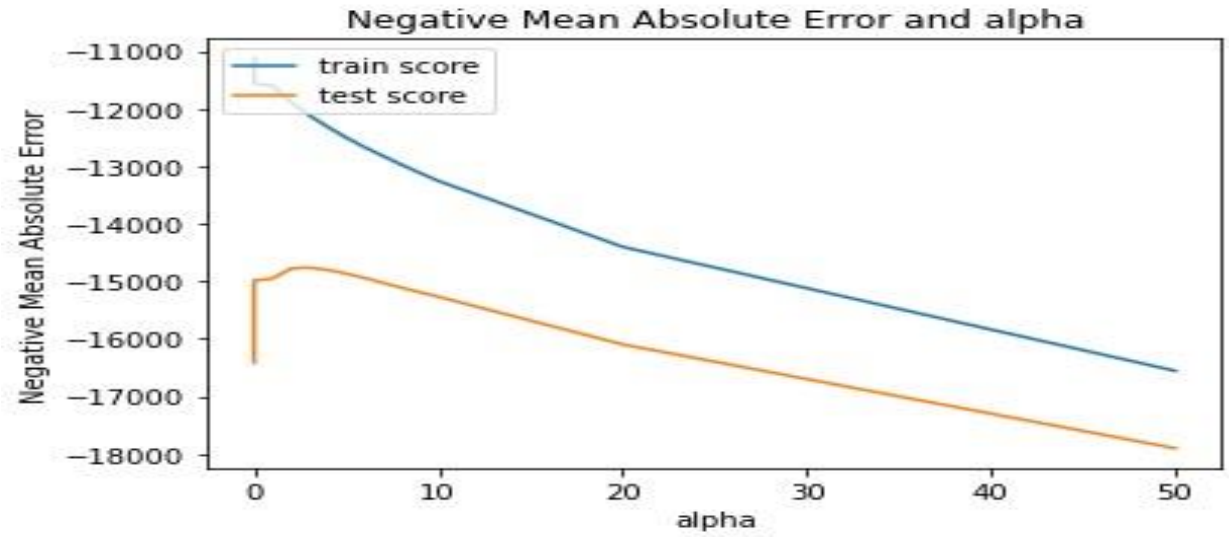
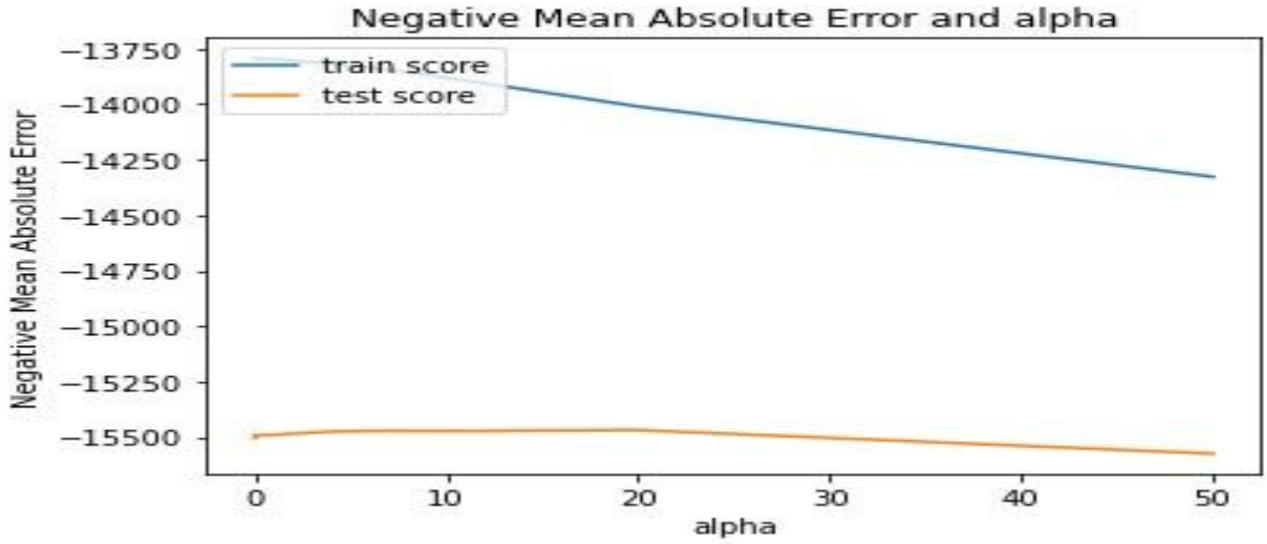
Question 1: What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer 1: Based on the experiments & modelling, the alpha for Ridge & Lasso regression was selected where the test score is maximum when we plot the Negative Mean Absolute Error against Alpha.

Ridge Regression (RFE Features): The best Alpha as seen from the graph was found to be **0.9**. If we double the alpha, we are increasing the regularization. As seen from the graph the Error would increase leading to a poor fit.



Lasso Regression (RFE Features): The best Alpha as seen from the graph was found to be **20**. If we double the alpha, we are increasing the regularization. As seen from the graph the Error would increase leading to a poor fit.



With increase in alpha or double it, only the coefficients of the model are reduced further. While changing the same, the most important predictors remained the same with the most optimal alpha with the co-efficient decreasing further in absolute magnitude.

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 2: Four Experiments/Models were tried with the below results:

1. For Ridge we observe :
    a. All Features Model (alpha = 3.0): R2 is 0.927 with ~180 features
    b. RFE Features Model (alpha = 0.9) : R2 is 0.916 with 50 features
2. For Lasso we observe :
    a. All Features Model (alpha = 100) : R2 is 0.926 with ~180 features
    b. RFE Features Model (alpha = 20) : R2 is 0.915 with 50 features

For the final model Lasso with RFE features (top 50) was selected. The reason are as follows:

a. The models made over 50 features have very similar R-square. So choosing it over 180 features as it's a simpler model.
b. Between Lasso & Ridge, we can see that Lasso has an alpha of 20, which further reduces dependence on low meaningful features pushing them towards 0, without effecting the performance majorly.
c. Hence Lasso in this case would give us a more robust & generalized model.

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?

Answer 3: The 5 most important features are:

a. BedroomAbvGr - Positively related with price
b. OverallCond - Positively related with price
c. 2ndFlrSF – Negatively related with price
d. LotArea - Positively related with price
e. SaleType_Con - Positively related with price

On removing these 5 features, and remodeling, the next top 5 features are:

a. WoodDeckSF
b. Neighborhood_NridgHt
c. BldgType_2fmCon
d. BsmtFinType2_No Basement
e. Exterior1st_Wd Sdng

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 4: Models are trained on training data and evaluated only data unseen during training also referred as Test Data. While we build a model and optimize on training data, the goal of the model is to perform good on test data. Also the difference in performance between training and test performance should be minimal. This showcases that a model is robust and generalizable.

Hence a robust & generalizable model is preferred to be as simple as possible without affecting the performance. If a model's performance changes a lot with small change in training data, the model has learnt the data and not the patterns from the data. This is a case of poor generalization referred as overfitting. Also an oversimplified model which fails to relate to the data has poor generalization and is referred as under-fitting.

The implications of these are poor performance of the model in real life over unseen data, as the model has not learn the underlying representation or pattern. To tackle this we use Regularization which strikes a balance between overfitting & under fitting, helping to train a model which is robust.