Assignment Part-II

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:

The optimal value of alpha for ridge and lasso regression

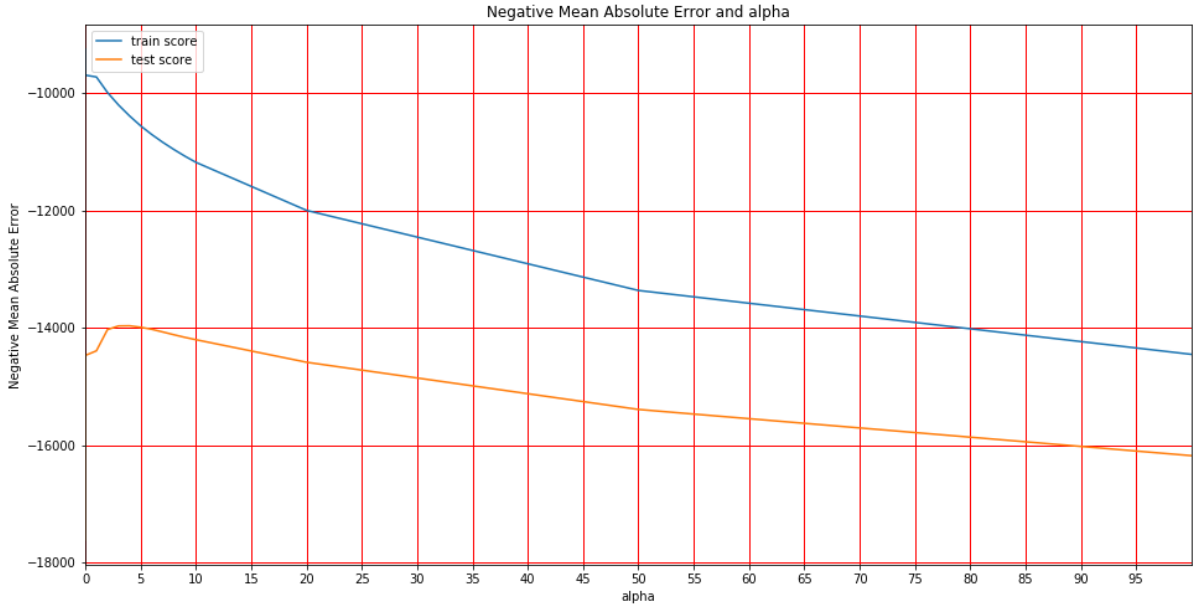
Ridge Alpha 1
lasso Alpha 10

R2score on training data has decreased but it has increased on testing data

---

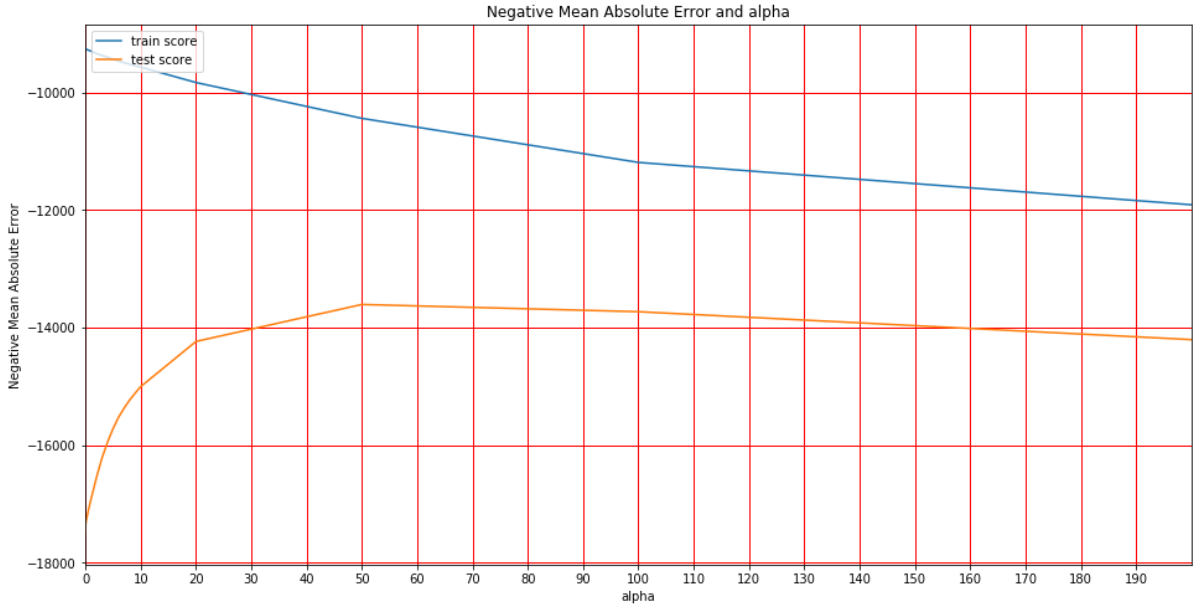
Question-2:
You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer:
We would decide that on the basis of plots and chose a value of alpha where we have good training as well as the test score.
Ridge regression plot:



Based on the plot, we choose 4 as the value for lambda for Ridge Regression, since it has the best train as well as the test score.

Lasso Regression Plot:



Based on the plot, we choose 50 as the value for lambda for Lasso Regression, since it has the best train as well as the test score.

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:
five most important predictor variables

11stFlrSF----------First Floor square feet
GrLivArea----------Above grade (ground) living area square feet
Street_Pave---------Pave road access to property
RoofMatl_Metal------Roof material_Metal
RoofStyle_Shed------Type of roof(Shed)

Question-5:
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?

Answer:

A model is considered to be robust if the model is stable, i.e. does not change drastically upon changing the training set. The model is considered generalisable if it does not overfits the training data, and works well with new data. Its implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data i.e. the **accuracy does not change much for training and test data.**