

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 regression: 4

The optimal value of alpha for lasso regression: 50

When the model is chosen with the double the optimal value of alpha for both ridge and lasso, there is a slight drop in the accuracy score of training and test sets and some of the coefficient values has been changed from zero to greater numbers.

Question-2:

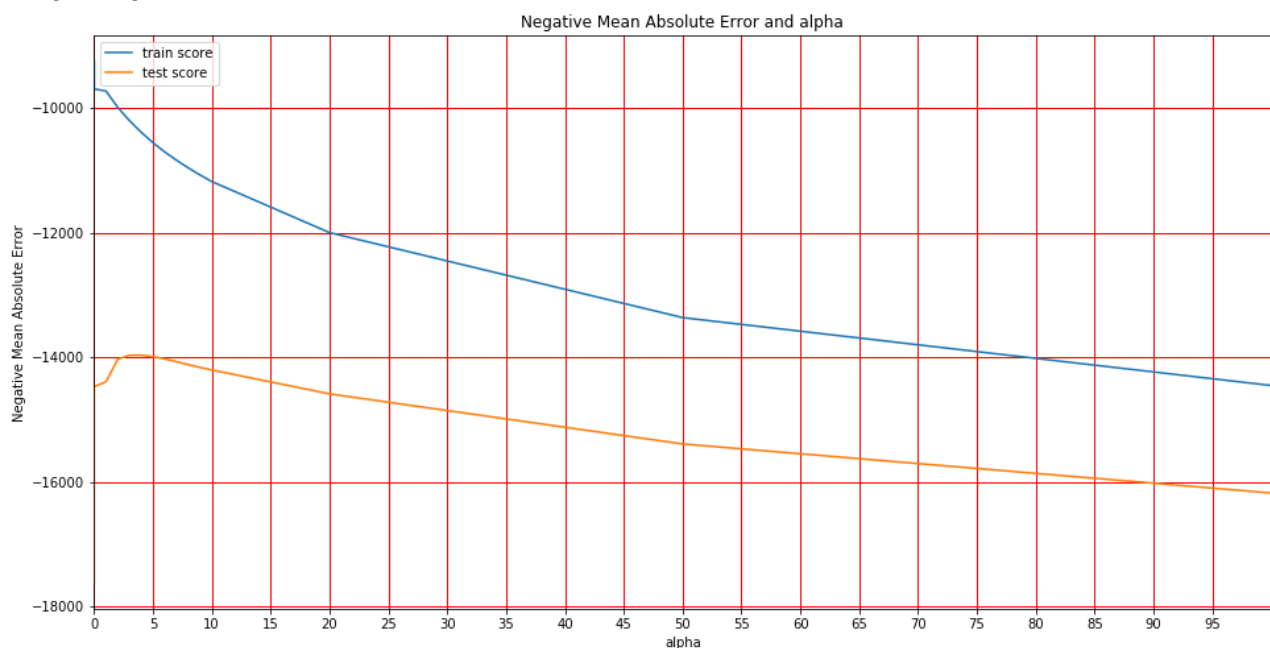
You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer:

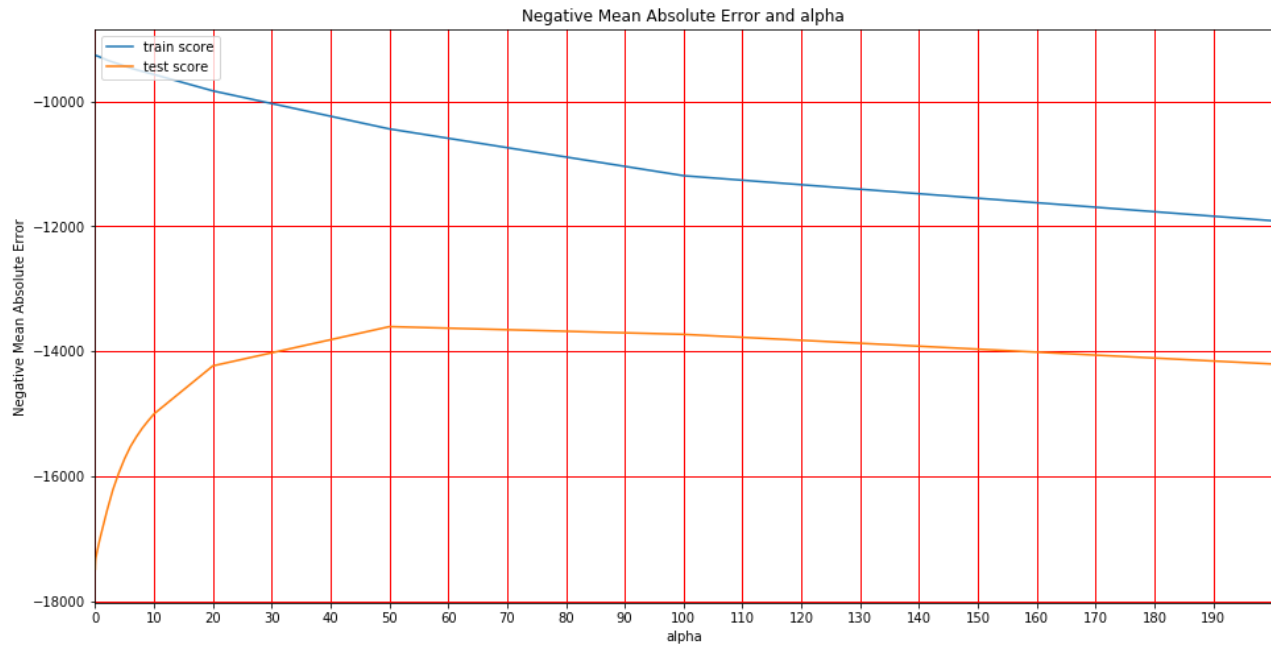
From the below negative mean absolute error and alpha plot of ridge and lasso regression, we got the optimal value of alpha as 4 and 50 respectively.

As Ridge Regression model had most of the coefficients values as non-zero and model was fairly complex where as Lasso Regression model was simple with most of the coefficients values as zero hence we would consider that Lasso Regression model with the alpha value as 50.

Ridge Regression Model:



Lasso Regression Model:



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?

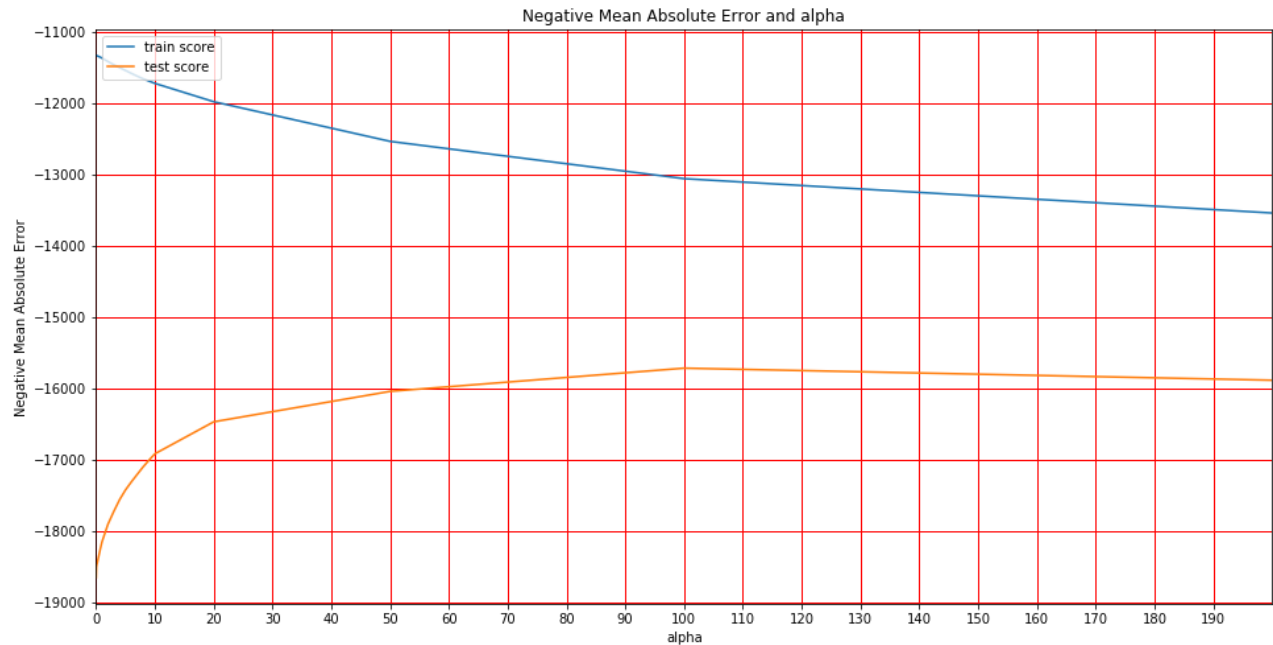
Answer:

From the lasso regression model the five most important predictor are

1. OverallCond
2. SaleCondition
3. Neighborhood
4. Functional
5. KitchenQual

After excluding these above 5 predictor variables and created the lasso model.

- Training set accuracy score: 0.92
- Test set accuracy score: 0.89
- Optimal alpha value: 100
- Adj. R-Squared: 0.80



Top 5 important predictor variables from the new model:

1. OverallQual
2. Exterior1st
3. SaleType
4. MSSubClass
5. Exterior2nd

Question-4:

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?

Evaluation metrics of the model (such as accuracy, sensitivity, specificity etc.) has not much impacted on changing the training dataset and doesn't exhibits overfitting with training data and predicts the new data (test data) as expected then the model is considered as robust and generalizable

In general model should behave/perform equally with training and test dataset without overfitting and underfitting the data.