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?
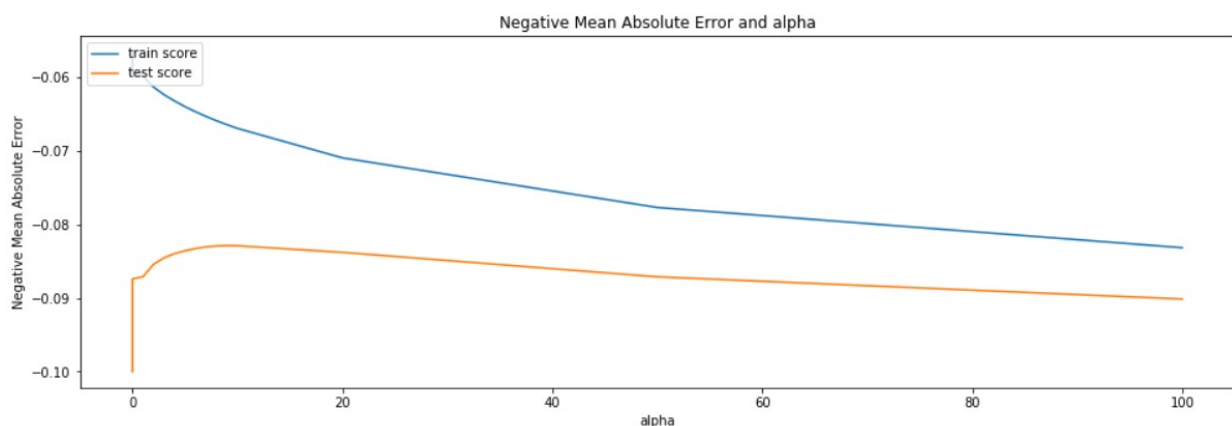
Answer:

Below are the optimal value of alpha for Ridge and Lasso regression:
- Ridge – 9
- Lasso – 0.0004

Below are the changes in the model if we choose double the value of alpha:
- Low value of alpha make less error
- Increasing the value of alpha reduce the accuracy of the model and increase the error value.

Ridge

Negative Mean Absolute Error and alpha

As per the above figure for the Ridge:
- Doubleing the value of alpha increased the error and reduce the accuracy in the train score.
- And, In the test score it will increase for the certain point of the alpha but after it will start reducing so we can say dobuling the value of alpha will reduce the accuracy of the model and increase the error.
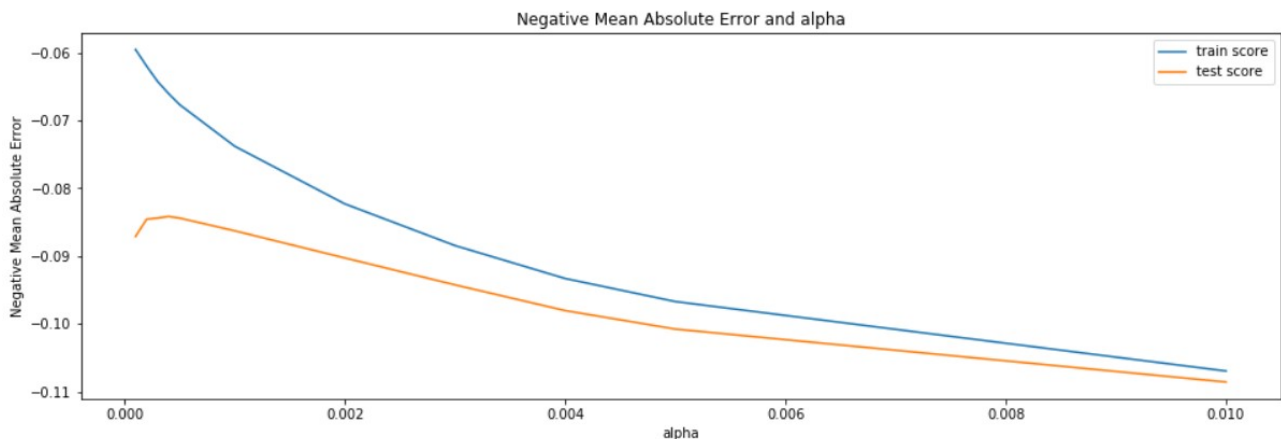
R2 Score:- R2 socre reduced

Before (aplha)	After (alpha*2)
0.9362080474015231	0.9294426413243642

RMSE:- RMSE value increased

Before (alpha)	After (alpha*2)
0.012831685330116334	0.013062555152318448

Lasso

Negative Mean Absolute Error and alpha

As can see in the above figure for Lasso
- Doubling the value of alpha reduce the accuracy and increased the error

R2 Score:- R2 socre reduced

Before (aplha)	After (alpha*2)
0.9355367045198499	0.9245292633683727

RMSE:- RMSE value increased

Before (alpha)	After (alpha*2)
0.012842851960487225	0.013127131500000416

Below are the most important predictor variables after the change is implemented
- **OverallQual_9**

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

Below are the R squred value of aplha for Ridge and Lasso regression:
- Ridge – 0.9362080474015231
- Lasso – 0.9355367045198499

Below are the RMSE value of alpha for Ridge and Lasso regression:
- Ridge – 0.012831685330116334
- Lasso – 0.012842851960487225

As can see in the above details R-squred and error value is same but slightly low for the lasso. Lasso help for the feature selection and coefficient value of one of the feature is zero so we will choose lasso over ridge.

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

Lasso regression indirectly performs feature selection and in the lasso model lesser important features cofficients become zero.

Increasing the value of alpha in lasso increasing the feature selection as many features coefficients values become zero.

Below are the list of the five most import predictor is after excluding the five most important predictor variables.

- TotRmsAbvGrd_8
- TotRmsAbvGrd_7
- TotRmsAbvGrd_6
- RoofMatl_Metal
- ExterQual_Fa

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?

Answer

Below are the point that we have to consider for the robust and generalisable model:

- As we have read in the course when we have two models that show almost similar results for the training or test data then we should select the simpler models because simpler models are robust. Simpler models requires fewer training data to train the model & easy to train.
- Simpler model will have low vraiance and high bias but the complex models have high variance and high bias (Bias Variance Trade-off).
- Simpler models prevents overfitting.
- It's easy to train simpler model and execution of the algorithm will take very less time.
- To make simple model we should use regularization techniques.
- Ridge & Lasso regression used to make the regression model simpler while balancing the bias-variance tradeoff.

Bias and Variance helps to maintain accuracy of the models, Bias Variance helps to minimize the total error.

Bias-variance tradeoff: Error vs Model Complexity, showing underfitting zone, overfitting zone, generalized error, variance, and bias.