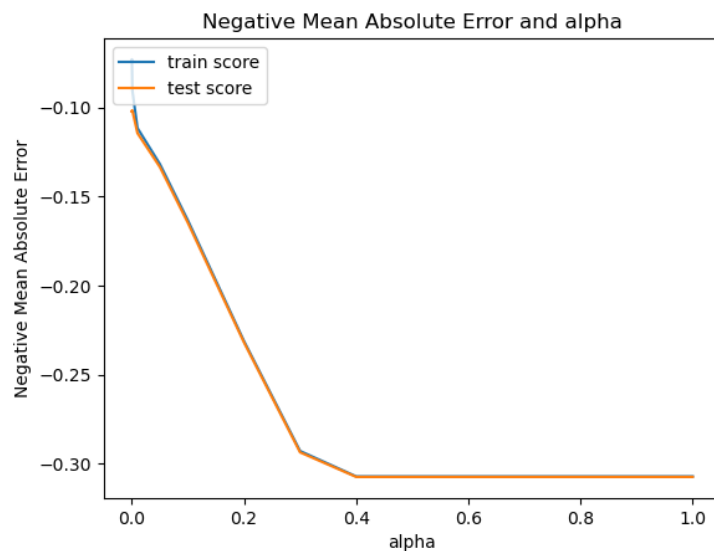


1. What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

In my final model, The optimal values that I got alpha for Ridge is 0.3 and for Lasso is 0.01.

- Even though if I double the values it is not affecting the performance as we can see in the plot it becomes flat after these alpha values



- After Tuning, In Ridge Model the important predictor variables are - RoofMatl\_WdShngl 0.592

RoofMatl\_CompShg 0.447

MSZoning\_RL 0.407

MSZoning\_FV 0.391

MSZoning\_RH 0.374

MSZoning\_RM 0.367

RoofMatl\_WdShake 0.367

RoofMatl\_Tar&Grv 0.316

Condition2\_PosA 0.313

- In Lasso, the important predictor variables are - OverallQual 0.134

GrLivArea 0.094

GarageCars 0.060

MSZoning\_RL 0.047

YearRemodAdd 0.042

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?

The model performance by Ridge Regression was better in terms of  $R^2$  values of Train and Test, However we will use Lasso Model because Model Interpretability is better in Lasso and since there are more features. If you suspect that only a subset of features is truly important, Lasso may be more suitable. We will consider multiple factors before choosing Models such as

Model Interpretability: We want better interpretability hence we choose Lasso

**Collinearity: It is effective to use Ridge but however we have removed correlated features hence we can use Lasso**

**Feature Importance:**

**If you have reason to believe that only a specific subset of features holds true significance, Lasso regression could be a more appropriate choice**

Hence we choose Lasso with

Test dataset  $R^2$  : 0.8560261902024202

Train dataset  $R^2$  : 0.8255210305312168

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?

The five most important predictor variables in our Lasso model are - OverallQual 0.134

GrLivArea 0.094

GarageCars 0.060

MSZoning\_RL 0.047

YearRemodAdd 0.042

If we remove those five and the next five most important variables would be

YearBuilt 0.038

GarageType\_Attchd 0.020

TotalBsmtSF 0.018

1stFlrSF 0.014

1GarageArea 0.012

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?

Ensuring that a model is robust and generalizable we use multiple strategies it involves:

Cross-Validation: Validating on multiple datasets

Hyper parameter tuning: optimizing the parameter so that grid searches randomly

Check if there is lot of difference between Test data, Train data performances which might indicate overfitting or underfitting