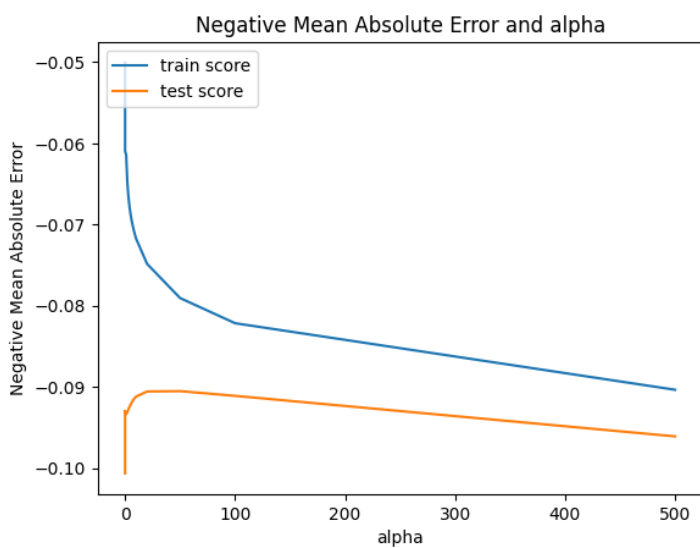


# Advanced Regression Assignment

## Problem Statement - Part II

- 1) **Question:** 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:**



### Ridge

The optimal value of alpha predicted for:

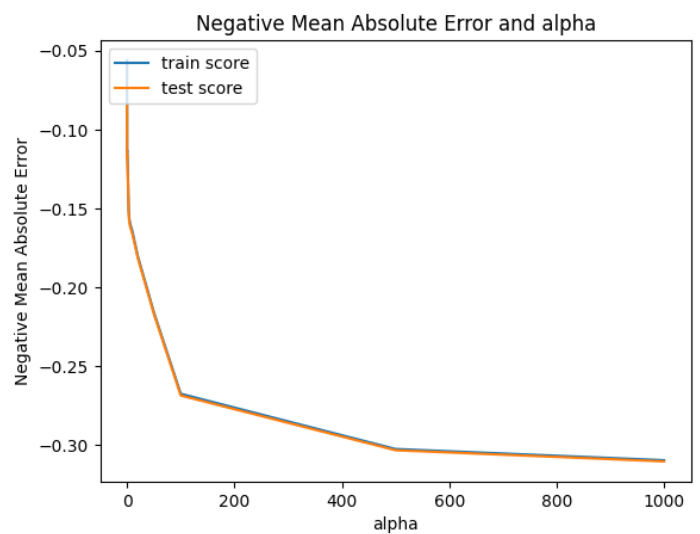
- Ridge Regression: 20
- Lasso Regression: 0

The R-squared value for Test data comes out to be for:

- Ridge Regression: 0.88
- Lasso Regression: 0.74

Top significance in predicting the price of a house:

- RoofMatl\_WdShngl
- SaleCondition\_Partial
- RoofMatl\_CompShg
- RoofMatl\_Roll
- RoofMatl\_Membran
- RoofMatl\_Metal



### Lasso

- RoofMatl\_Tar&Grv
- GarageType\_na
- GarageQual\_na
- GarageCond\_na

- 2) **Question:** 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:**

- Depends on use case –
- case 1
- Lasso regression will perform better - the number of predictor variables is significantly less.
- Lasso regression is able to shrink insignificant variables completely to zero and remove them from the model.
- case 2
- Ridge regression will perform better - the number of predictor variables is significantly more and their coefficients are roughly equal.
- Ridge regression performs better because it keeps all of the predictors in the model.

- 3) **Question:** 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:**

We got alpha value as 0.001.

New Top 5 Predictors:

- 2ndFlrSF
- 1stFlrSF
- MSSubClass\_70
- Functional\_Typ
- SaleType\_New

- 4) **Question:** 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:**

How can you make sure that a model is robust and generalizable?

- 1) The model should not be impacted by outliers in the training data.
- 2) The test accuracy is greater than the training score
- 3) The model should be accurate for new Test datasets other than the ones which were used during the modeling.

The implications of the same for the accuracy of the model and why?

- 1) Outlier treatment should be done in training. This will increase the accuracy of the predictions made by the model.
- 2) Confidence intervals can be used which will standardize the predictions made by the model.