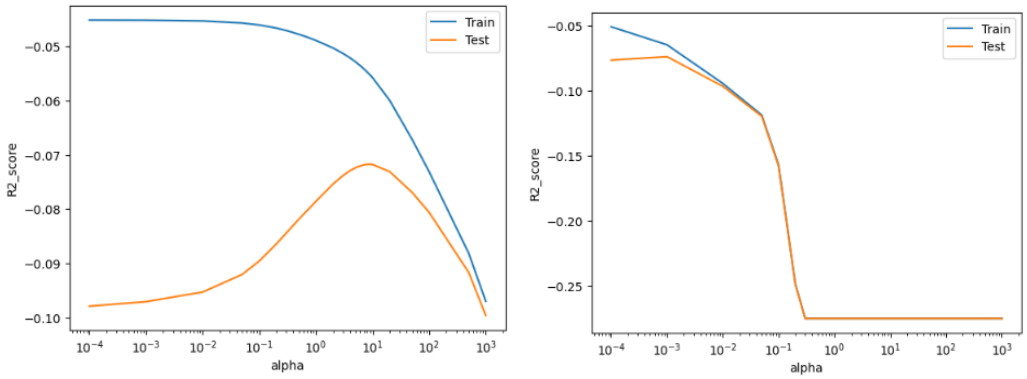
Question 1

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 –



Top 10 variables are -

- OverallQual_9
- Neighborhood_Crawfor
- OverallQual_8
- GrLivArea
- Functional_Typ
- SaleType_New
- SaleCondition_Normal
-
- TotalBsmtSF
- MSSubClass_70
- BsmtExposure_Gd

Optimal value of lambda for Ridge Regression = 9

Optimal value of lambda for Lasso = 0.001

Ridge at alpha = 9

- R2 Score (Train) = 0.95
- R2 Score (Test) = 0.92

Lasso at alpha = 0.001

- R2 Score (Train) = 0.93
- R2 Score (Test) = 0.92

- **Double the value of alpha for both ridge and lasso.**

Observations -

Ridge R2 score at alpha = 9

- R2 Score (Train) = 0.95
- R2 Score (Test) = 0.92

Ridge R2 score at alpha = 18

- R2 Score (Train) = 0.94
- R2 Score (Test) = 0.92

Lasso R2 score at alpha = 0.001

- R2 Score (Train) = 0.93
- R2 Score (Test) = 0.92

Lasso R2 score at alpha = 0.002¶

- R2 Score (Train) = 0.91
- R2 Score (Test) = 0.91

Observation

Top 10 variables after double alpha values for Ridge and lasso regression

- GrLivArea
- Neighborhood_Crawfor
- OverallQual_8
- OverallQual_9
- Functional_Typ
- SaleCondition_Normal
- OverallCond_7
- TotalBsmtSF
- BsmtExposure_Gd
- 2ndFlrSF

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-

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

We got alpha value as 0.001.

New Top 5 Predictors

- 2ndFlrSF
- 1stFlrSF
- MSSubClass_70
- Functional_Typ
- SaleType_New

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 -

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

- 1) 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 modeling.

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

- 1) Outlier treatment should be done in traning. 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.