Part – II: Subjective Questions

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 is 5.0 and for Lasso regression is 0.001.

From Fig. (1) & (2), we can conclude that if we choose double the value of alpha for both ridge & lasso regressions then the Lasso regression model's R2 score values will have larger variation w.r.t train data score (blue line) & test data score (red line). In the case of Ridge regression model both train data score & test data score will approach to same values.

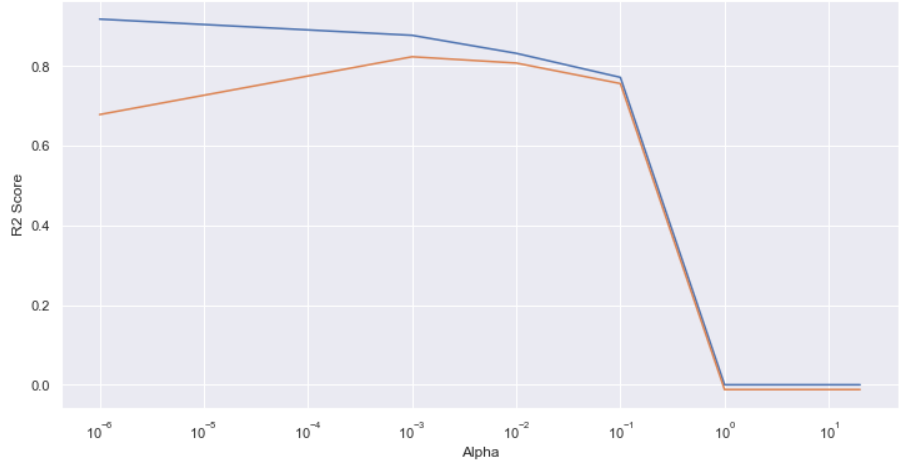

Fig. (1): Plot of Alpha values vs. R2 scores of Lasso regression model.

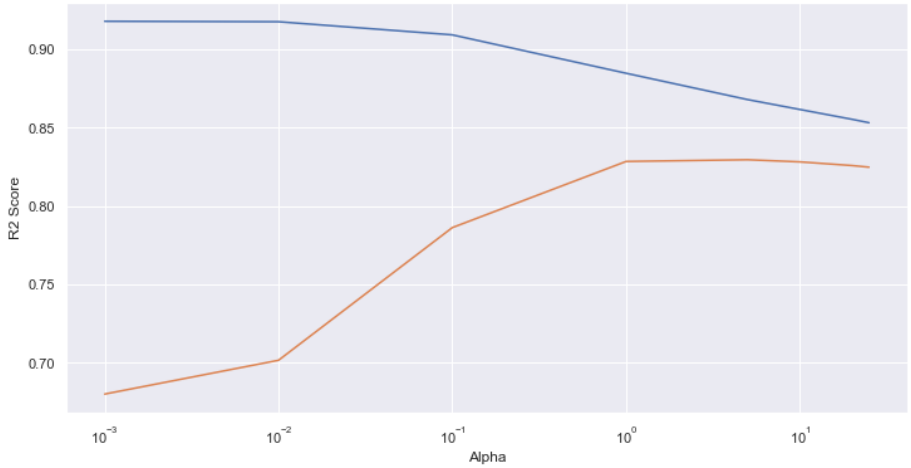

Fig. (2): Plot of Alpha values vs. R2 scores of Ridge regression model.

Following are the most important predictor variables after the change is implemented.

- OverallQual
- MSZoning_FV
- MSZoning_RL
- RoofMatl_WdShngl
- Neighborhood_Crawfor

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:

The optimal value of lambda for lasso regression is 0.001 and lasso regression helps in feature selection. As we increase the value of lambda more coefficients will be zero. Thus, I would like to select the Lasso regression model.

Question 3:

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:

After the removal of five important predictor variables and running the code again, following variables became more significant.

Lasso Model

- GarageCars
- GrLivArea
- Neighborhood_NridgHt
- BsmtFinSF1
- SaleCondition_Normal

Ridge Model

- TotalBsmtSF
- Neighborhood_NridgHt
- SaleCondition_Normal
- SaleCondition_Partial
- Functional_Typ

Question 4:

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer:

Model can be made more robust and generalized by using methods which can include outliers in the analysis without affecting the prediction variables. Also, other performance metrics can also be included in the analysis rather than relying on only one or two metrics like MSE or RMSE. With respect to data, some other transformations can be used to make them normal distributed if they show skewness.

******

Answers submitted by

Sanjay M. Belgaonkar
sanjaymb@ieee.org