Outstanding Project-2 July ML Batch 1

Name: Sanjay Marreddi

sanjay.marreddi.19041@iitgoa.ac.in

Problem Statement - Part II (subjective Assignment)

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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Sol :

I choose the following values of Alpha for both the Regressions

Alpha_values = [1e-15,1e-13,1e-11,1e-9,1e-7,1e-5,1e-4,1e-3,1e-2,1e-1,0.2,0.3,0.3,0.4,0.5,0.7,0.9,1,2.5,3,5,10,20,30,40,100]

The Optimal Value of alpha for RIDGE Regression obtained is **0.1**

The Optimal Value of alpha for LASSO Regression obtained is **30**

When we double the value of alpha for RIDGE Regression, the score decreased from **0.89056** to **0.88766**

When we double the value of alpha for LASSO Regression, the score decreased from **0.89303** to **0.88546**

The most important predictor variables obtained from RIDGE Regression : - 'Street', 'OverallQual', 'Fireplaces', 'LandSlope', 'GarageCars', 'FullBath', 'HalfBath', 'OverallCond', 'Functional', 'MasVnrType', 'SaleCondition', 'BsmtFullBath', 'TotRmsAbvGrd', 'RoofMatl', 'Foundation', 'GarageCond', 'LandContour', 'PavedDrive', 'RoofStyle', 'BsmtCond', 'ExterCond', 'CentralAir', 'Exterior2nd', 'Neighborhood', 'BsmtFinType2', 'Heating', 'MoSold', 'YearRemodAdd', 'YearBuilt', 'LotConfig', 'GarageYrBlt', '3SsnPorch', 'ScreenPorch', 'MasVnrArea', 'GrLivArea'

The most important predictor variables obtained from LASSO Regression : -

'OverallQual', 'GarageCars', 'Fireplaces', 'Functional', 'OverallCond', 'SaleCondition', 'YearBuilt', 'YearRemodAdd', 'GrLivArea', 'BsmtFinSF1', 'GarageArea', 'WoodDeckSF', 'TotalBsmtSF', 'MasVnrArea', '1stFlrSF', 'LotArea', 'MSZoning', 'LotFrontage', 'Street', 'LotShape', 'LandContour', 'Utilities', 'LotConfig', 'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'HouseStyle', 'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType', 'ExterCond', 'Foundation'

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?

The Optimal Value of Lambda in Ridge Regression is 0.1 and the Corresponding Score is **0.89056.**

The Optimal Value of Lambda in Lasso Regression is 30 and the Corresponding Score is **0.89303.**

I will choose Lasso Regression over Ridge Regression because of the following reasons :

1. It has more R2_SCORE when compared to Ridge Regression which in turn implies high accuracy in making prediction.
2. Moreover, if we go in depth about their working algorithm, we can see that though the Lasso Algorithm is faster than Ridge Algorithm, Lasso has the advantage of completely reducing the un necessary parameters, i.e. bringing coefficients to zero which is very helpful in our case as we have many Independent variables to choose for predicting the House Prices.

Hence, I will choose to apply the Lasso Regression in this case !

Question 3

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?

According to my understanding, we can make sure that a model is robust and generalisable if it satisfies the below conditions :

1) Less Bias in Prediction
2) Less Variation i.e. If its output is accurate even if one or more input variables are changed unfortunately. In short words it should be tolerant to noise.
3) Having more accuracy in Test Data is utmost Important. In this case of Regression, R2_Score more than 0.95. It in turn implies that accuracy is more than 95% .
4) These type of conditions are needed at a high preference when our Model is trying to solve a Real-World problem.
5) Moreover, when working on health-related Problems we even demand 98-99% accuracy to say that our model is robust.