Report

Name: Soham Mehta
ID: 111496015
Name on Kaggle: sohammehta95
Ranking Obtained: 2533
Score: 0.0652044

Lasso Regression:

The LASSO (Least Absolute Shrinkage and Selection Operator) is a regression method that involves penalizing the absolute size of the regression coefficients. When there are many variables in a dataset and some variables are very highly correlated, the variance keeps increasing. Hence, to fix this problem Ridge regression is used which puts a constraint on how big the Betas can get. It also tries to minimize the sum of squares of the residuals. The same thing is also done in Lasso regression where L1 norm is used instead of an L2 norm. The key difference between them is that the constraint graph of the Lasso regression is a Diamond instead of a circle because when we graph the absolute value, the sum of square residuals obtained in lasso is always less than a specific value say "C" which lies on each side of the axis to form a diamond shape.

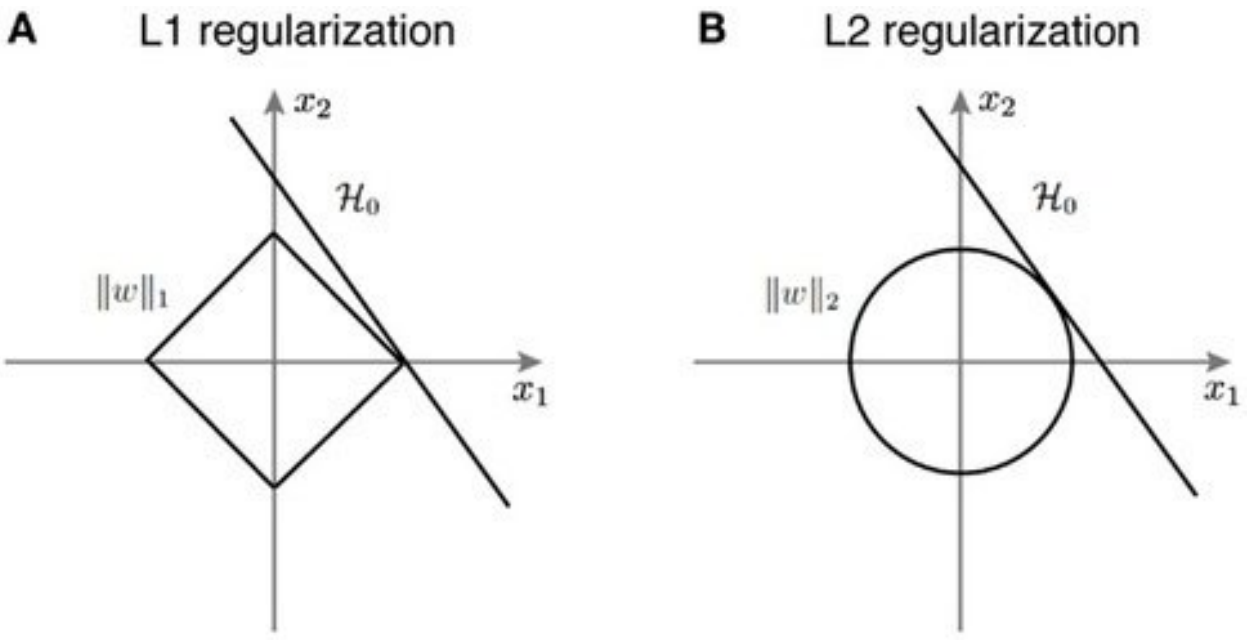

Figure 1 L1 norm of Lasso Regression V/S L2 norm of Ridge regression

The most important property of Lasso regression is that there is a highest chance for a function to hit a given corner than any other point on the diamond which makes sure that one or many of the coefficients are zero.

Implementation

```
#Implementing Lasso Regression

from sklearn import datasets, linear_model
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
from sklearn.model_selection import train_test_split


target = train["logerror"].values
feature_names = trainy.columns.values


X_train_lasso, X_test_lasso, Y_train_lasso, Y_test_lasso = train_test_split(trainy,
target, test_size=0.3)

lasso = linear_model.Lasso(alpha=10.0)
lasso.fit(X_train_lasso, Y_train_lasso)
Y_prediction_lasso = lasso.predict(X_test_lasso)

print("Mean squared error: %f" %mean_squared_error(Y_test_lasso, Y_prediction_lasso))
print("Mean absolute error: %f" %mean_absolute_error(Y_test_lasso,
Y_prediction_lasso))
```

In any equation, when the data is plotted, the curve is generalized for the linear parameters. However, as the power of the variable increases, the regression starts taking more curved shapes to satisfy all the points. This may cause the error to be decreased, however, the future prediction might not work properly as the shape of the graph is not defined well. Hence Lasso regression uses the “Alpha” parameter which suppresses the weightage of the higher powered variable coefficients. It controls the sum of the absolute value of the coefficients and the lasso coefficients become zero in a certain range.

Observation

The kaggle dataset gave a better fitting for the linear regression than both the ridge and lasso regressions. One of the major reasons might be that the data given by kaggle must be more fitting in a straight line than for the higher valued coefficient terms which are controlled by lasso and ridge.

Conclusion

Hence, in conclusion, Lasso regression can shrink large coefficients and avoid overfitting by forcing the sum of the absolute values of regression coefficients to be extremely less or even zero. The disadvantage of lasso can be seen in this particular dataset. However, as the value of alpha would decrease, the performance would increase and the complexity of lasso would decrease as the fitting of the curve would be more generalized.