Lab 9 Summary

In this lab, we were able to practice linear regression which is the process of modeling the relationship between a scalar response and explanatory variables. We also had the opportunity to analyze logistic regression which is a type of classification algorithm based on supervised learning.

We used Mean squared error (MSE), which gives us the squared value of the difference between the original values and predicted values. This value shows us the efficiency of our dataset when a Machine Learning Algorithm is tested on it.

Logistic Regression

We have split our dataset into 60% training and 40% test data. We have then fitted Logistic Regression to our training dataset. Upon predicting the target value using logistic regression on the training dataset, we received a Mean squared error of 0.115. When we calculated the Mean Squared Error for the test data set, we received a value of 0.203.

Classification Report :

precision recall f1-score support

0 0.78 0.76 0.77 82

1 0.81 0.83 0.82 100

accuracy 0.80 182

macro avg 0.80 0.79 0.79 182

weighted avg 0.80 0.80 0.80 182

[[46 10]

[ 8 57]]

Our True Negative and True Positives are high enough to say that the model we used is efficient on the dataset.

Linear Regression

Like the previous method, our dataset has been split by a ratio of 3:2. An SGD classifier has been for optimization. We fitted the SGD classifier to our training dataset. Upon running the mean squared error algorithm on our dataset, we received the following values :

MSE on training set: 0.4462809917355372

MSE on test set: 0.3956043956043956

Conclusion

Upon careful introspection, the mean squared error of our dataset is higher for Linear Regression as compared to Logistic Regression. Due to our dataset having binary target values, Linear Regression would not be an efficient model to run on it because it relies on values spread out over a line.