REPORT – PROGRAMMING ASSIGMENT 1

IMPLEMENTATION OF LOGISTIC REGRESSION ALGORITHM

BY

SAI ABHISHEK THOTA - SXT210056

RAHUL GAURI – RXG200002

Part 1

For the learning rates {0.01, 0.1, 0.333} and for the iterations {10,100,1000,10000} we see that the least testing error is obtained when learning rate is 0.33 and the number of iteration have been set to 10000.

Between testing and training error, we choose testing error as a metric to select our best parameters for the model as the ultimate goal of building a machine learning model is to make accurate predictions on new, unseen data.

Part 3 – Comparision of scikit-learn Logistic Regression Algorithm

As observed in the above outputs, we have obtained the least Training Error and Test Errors at the Learning Rate = 0.33 and Best Iteration = 1000 using our own implementation of SimpleLogisiticRegression(). The train error we have gotten was 0.1803 and the Test Error we have found was 0.1828. When we use scikit-learn Logistic Regression Algorithm, the Training Error we have found was 0.1639 and Test Error was 0.1782. There is a small difference in the values obtained through SimpleLogisiticRegression and scikit-learn LogisticRegression.

Some of the reasons for the difference are mentioned as below

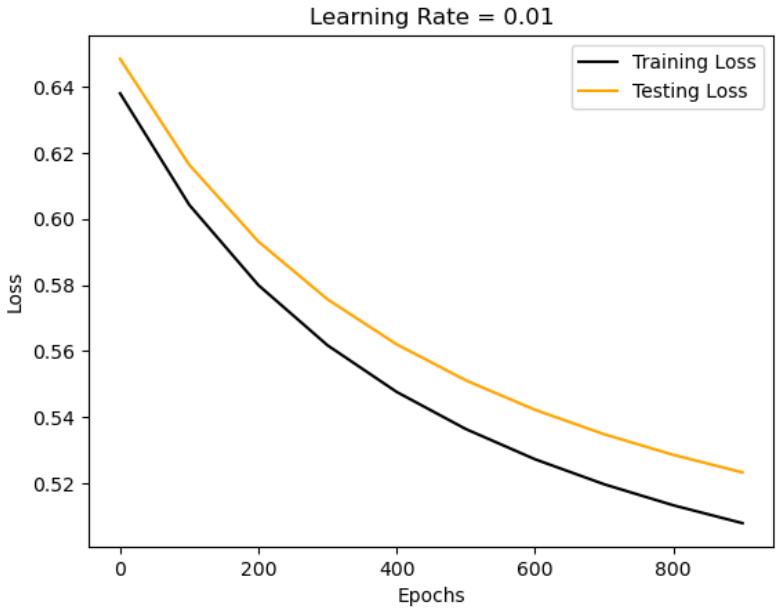
1. Optimization Algorithm – scikit-learn uses an algorithm called L-BFGS (Limited-Memory Broyden–Fletcher–Goldfarb–Shanno algorithm) while our implementation uses gradient_ascent to converge our results.

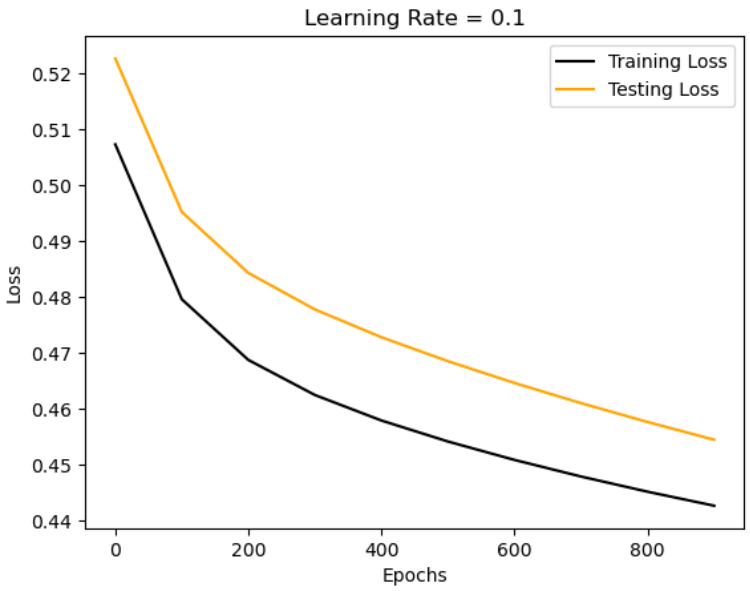
2. Regularization – scikit-learn has a parameter called “C” which prevents overfitting. Based on the overfitting that is happening in our model, we can modify the value of C accordingly to apply more regularization.

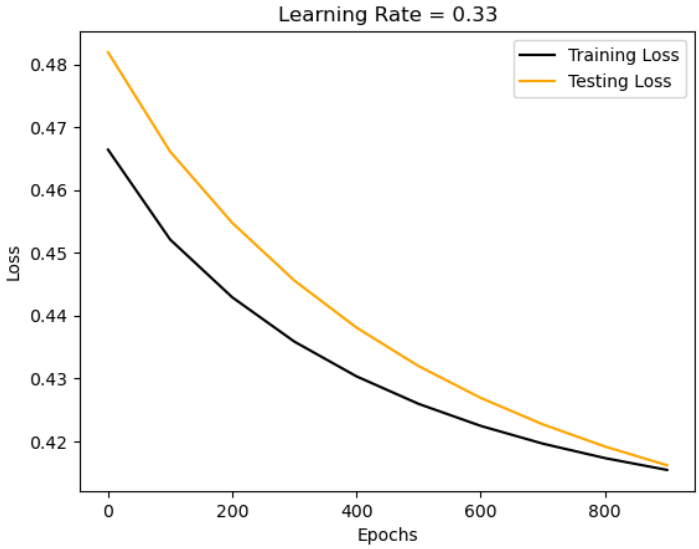
3. Multiparameter tuning – scikit-learn has the capability to tune parameters such as ‘solver’, ‘penalty’ and ‘C’. The performance of our algorithm can be altered by altering the values of these parameters.

Part 4 – Plotting of graphs

We have plotted 3 graphs for 1000 iterations for different learning rates as depicted below






Learning Rate = 0.33
Training Loss
Testing Loss
0.48
0.47
0.46
0.45
0.44
0.43
0.42
Loss
0
200
400
600
800
Epochs