CS F464 | Machine Learning

Aditya Agarwal (2017B1A71075H)

Manisha Katariya (2017B3A70354H)

Jalaj Bansal (2017B3A71610H)

Assignment 2

2A. Logistic Regression

1. A very brief description of your model and its implementation.

Our model implements **logistic regression** with both **Gradient Descent (GD)** and **Stochastic Gradient Descent (SGD)** for binary classification. The dataset used has 4 attributes and two classes. The probabilities for each sample are calculated using a **sigmoid function** with a threshold value of 0.5. We tested the model for various learning rates and iterations and calculated accuracy, loss, fscores, precision and recall. We created **10 independent 70:30 splits** and trained the model.

2. The most important feature in the dataset.

Upon training the model with both GD and SGD, we find **attr1** to be the most important feature since its weight has the highest absolute value across all learning rates.

3. Final Results and Metrics

Gradient Descent (Learning Rate = 0.001)

Average Training Accuracy (5000 epochs): 99.76%

Testing Accuracy: 92.50%

Loss: 0.24

Precision: 1.0

Recall: 0.92

F-Score: 0.96

Stochastic Gradient Descent (Learning Rate = 0.1)

Average Training Accuracy (50000 epochs): 96.02%

Testing Accuracy: 92.18%

Loss: 0.24

Precision: 1.0

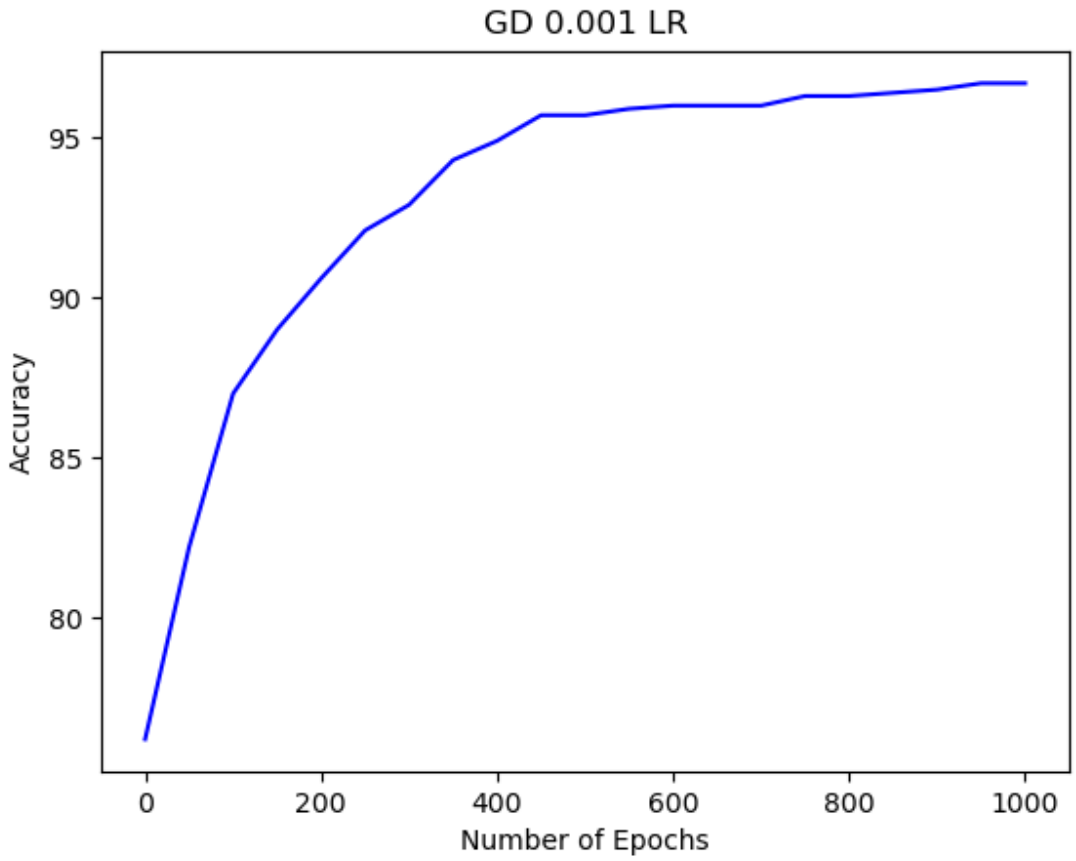
Recall: 0.92

F-Score: 0.95

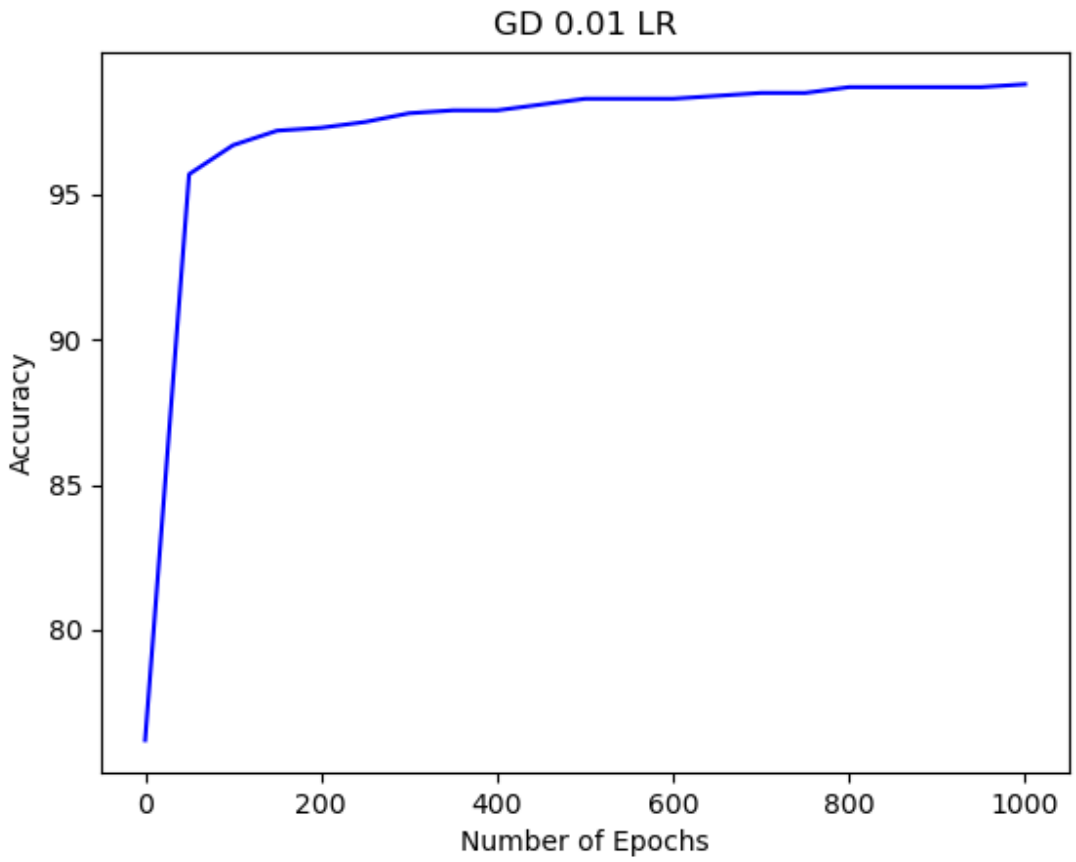
4. Accuracy Plots for varying learning rates

Gradient Descent

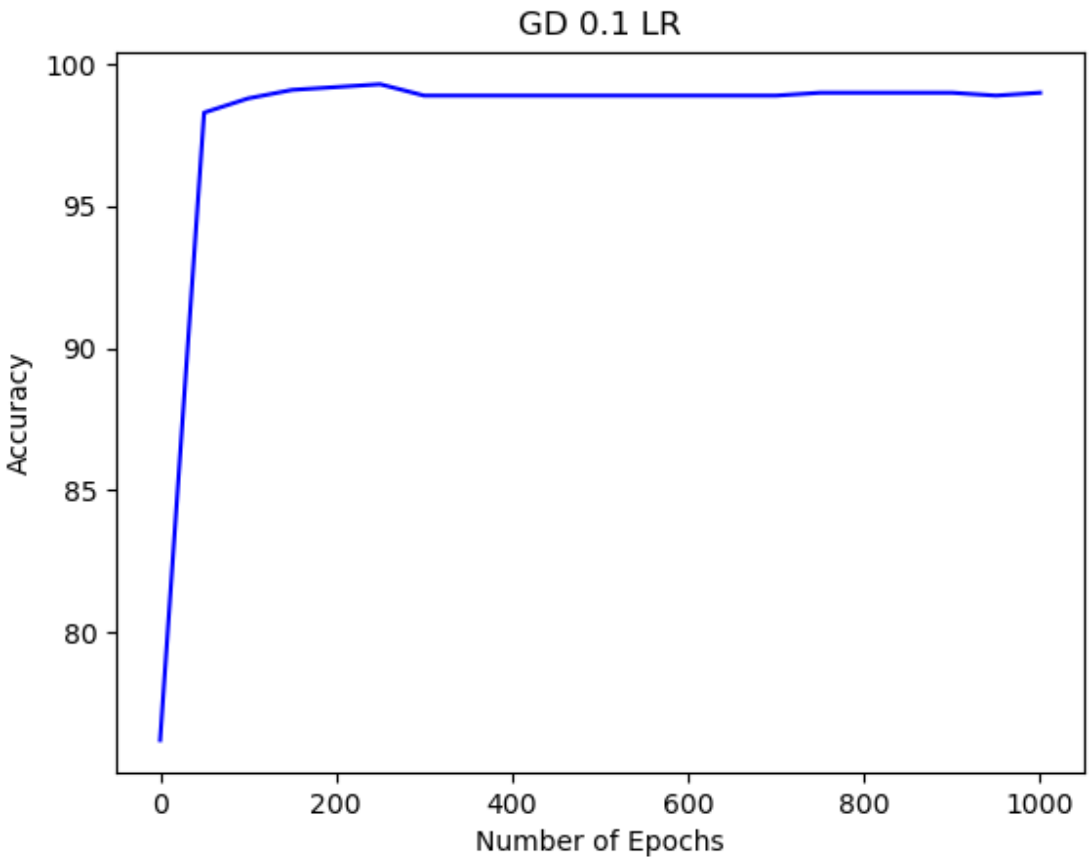
Learning Rate: 0.001



Learning Rate: 0.01

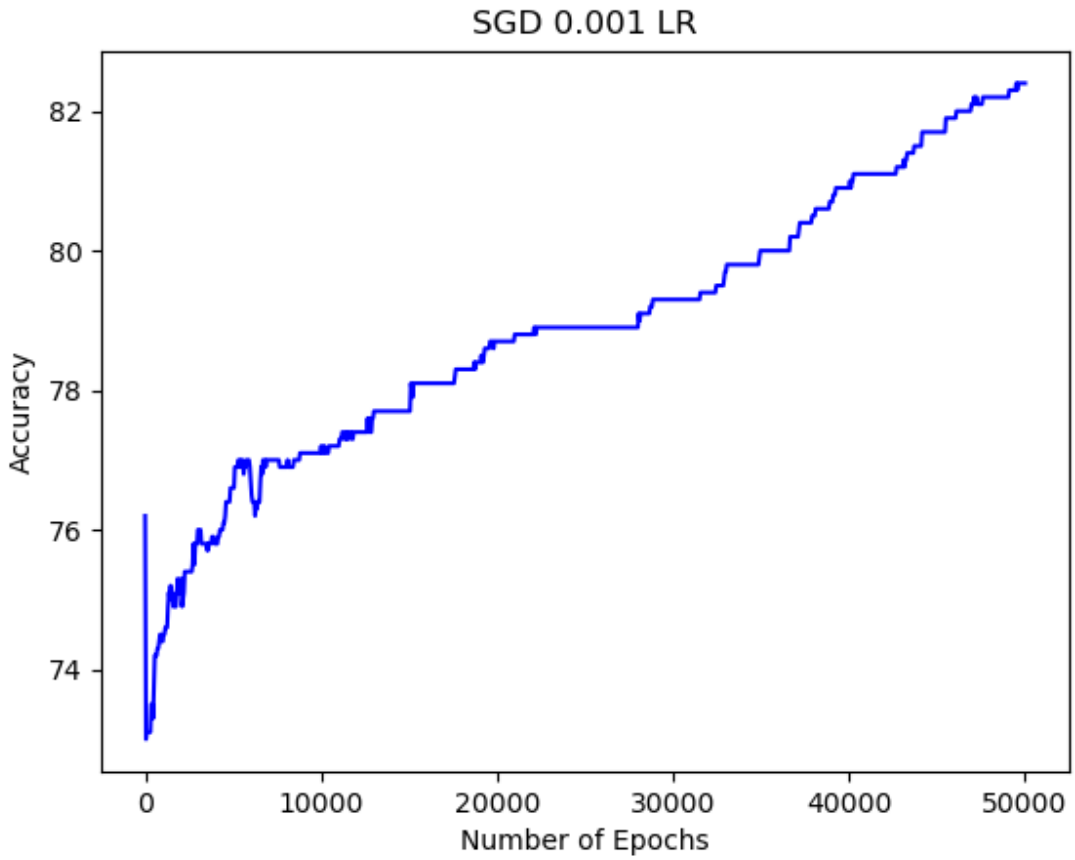


Learning Rate: 0.1

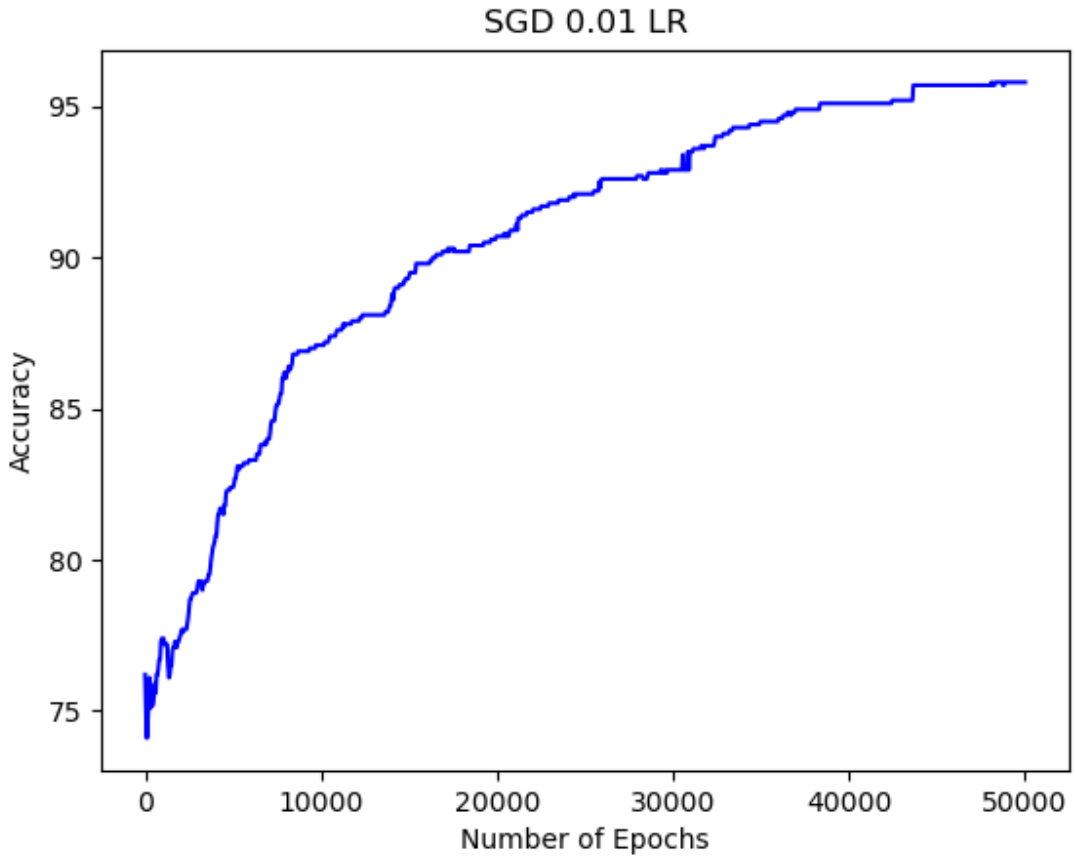


Stochastic Gradient Descent

Learning Rate: 0.001



Learning Rate: 0.01



Learning Rate: 0.1