EEE 443/543 - Spring 2025

Project #3

Introduction

In this project, we explore the multicategory perceptron training algorithm (PTA) for digit classification using the MNIST dataset. The objective is to classify digit images into one of 10 categories by training a neural network with 784 input nodes and 10 output nodes, learning a total of 7840 weights. Unlike the single-category PTA in Project #2, this experiment involves data that is not linearly separable, making convergence to zero misclassifications unlikely.

Dataset and Preprocessing

The MNIST dataset includes 60,000 training images and 10,000 test images, each of size 28×28 pixels. These images were vectorized into 784-dimensional input vectors. The dataset was divided as follows:

- Training set: First *n* images and their corresponding labels (varied across experiments).
- Test set: All 10,000 test images and their labels.

The neural network's output was configured to produce one-hot encoded vectors, such as [1, 0, 0, 0 ...] for the digit 0.

Results and Observations

1. **Training with η = 1, ϵ=0, n=50**

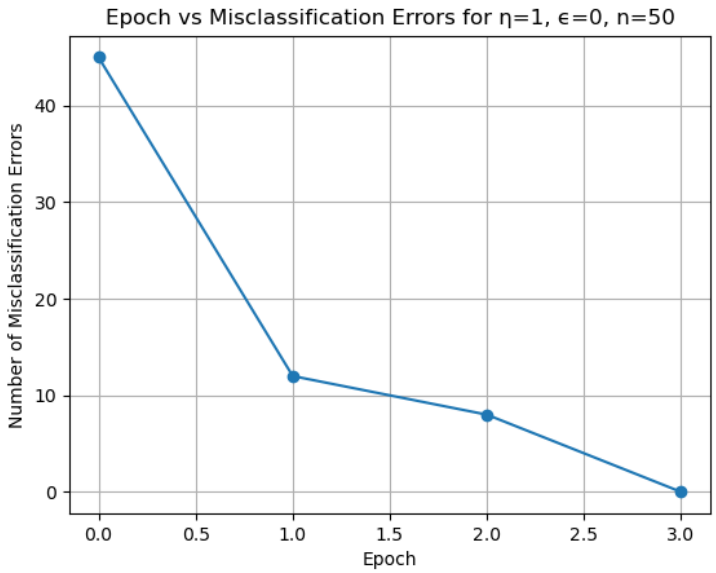

Figure 1

The training process achieved zero errors by the final epoch as outlined in the manual's expectations. When using the weights produced by this amount of data, the error percentage on test set is 42.97% which is not good.

2. **Training with η = 1, ϵ=0, n=1000**

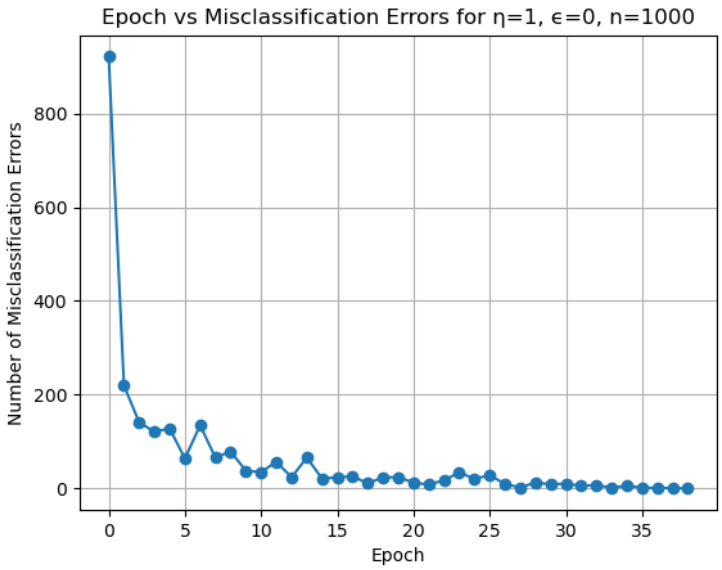

Figure 2

Training resulted in zero errors here as well aligning with the expectations. When using the weights produced by this amount of data, the error percentage on test set is 17.45% which is better than n=50 case but still high.

3. **Training with η = 1, ϵ=0, n=60000**

In this case, it is impossible to reach zero errors, which prevents the algorithm from terminating. However, with a modification allowing the number of epochs to be limited, it was observed that the error count for the given training data does not drop below approximately 8,000. This indicates that ϵ should be at least around 0.13.

4. **Training with η = 1, ϵ=0.15, n=60000**

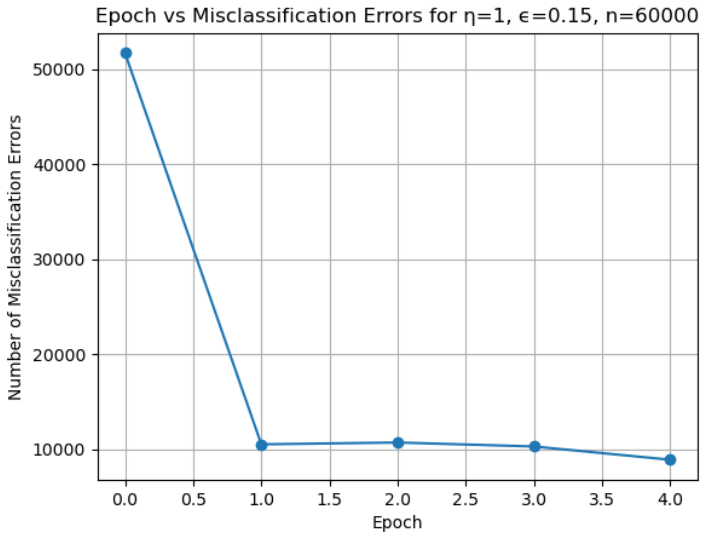

Figure 3

When using the weights produced by whole data, the error percentage on test set is 15.51% which is better than all cases but still high due to limitations of the algorithm.

Comments on i

1) **Effect of η on Convergence**

- Changing η affects both the speed and stability of convergence. A larger η can speed up convergence but may cause oscillations, while a smaller η leads to slower but more stable updates. In this project, η = 1 provided a good balance.

2) **Effect of Initial Conditions**

- Different initial weights led to variations in the number of epochs required for convergence. While convergence was eventually achieved in all cases, the starting point influenced the trajectory significantly.

3) **Effect of Sample Size**

- Increasing *n* resulted in longer training times but improved generalization. Smaller datasets converged quickly but performed worse on the test set. For *n = 60000*, the algorithm required careful tuning of ε to achieve meaningful results without excessive epochs.

Conclusion

The multicategory PTA successfully classified digit images under controlled conditions. However, its performance was influenced by the choice of hyperparameters (η and ε), initial weights, and dataset size. Larger datasets improved generalization at the cost of longer training times and more complex convergence behavior.