NEURAL NETWORK TRAINING WITH PARTICLE SWARM OPTIMIZATION

Poonam Rajan Pawar | pp1549@rit.edu | Advisor: Dr. Weijie Zhao

RIT Rochester Institute of Technology

INTRODUCTION

The performance of neural networks relies heavily on optimizing trainable parameters such as weights.

Popular gradient-based methods like **Stochastic Gradient Descent (SGD)** and **Adam** often struggle with local minima and noisy gradients.

This project explores **Particle Swarm Optimization (PSO)** as a gradient-free, biologically inspired alternative method for training neural networks.

CORE CONCEPTS

Neural network training enables models to learn data patterns by adjusting weights across layers.

🎯 **Evaluation Metrics**

↓ Lower Loss → ↑Performance

↑ Accuracy → ↑ Correct Predictions

The goal is to minimize loss and optimize weight configuration.

⚙️ **Particle Swarm Optimization (PSO):** It uses a swarm of particles that update their positions based on cognitive and social components.

🧬 **Genetic Algorithms (GA):** It has a population of particles and updates their positions through **crossover and mutation.**

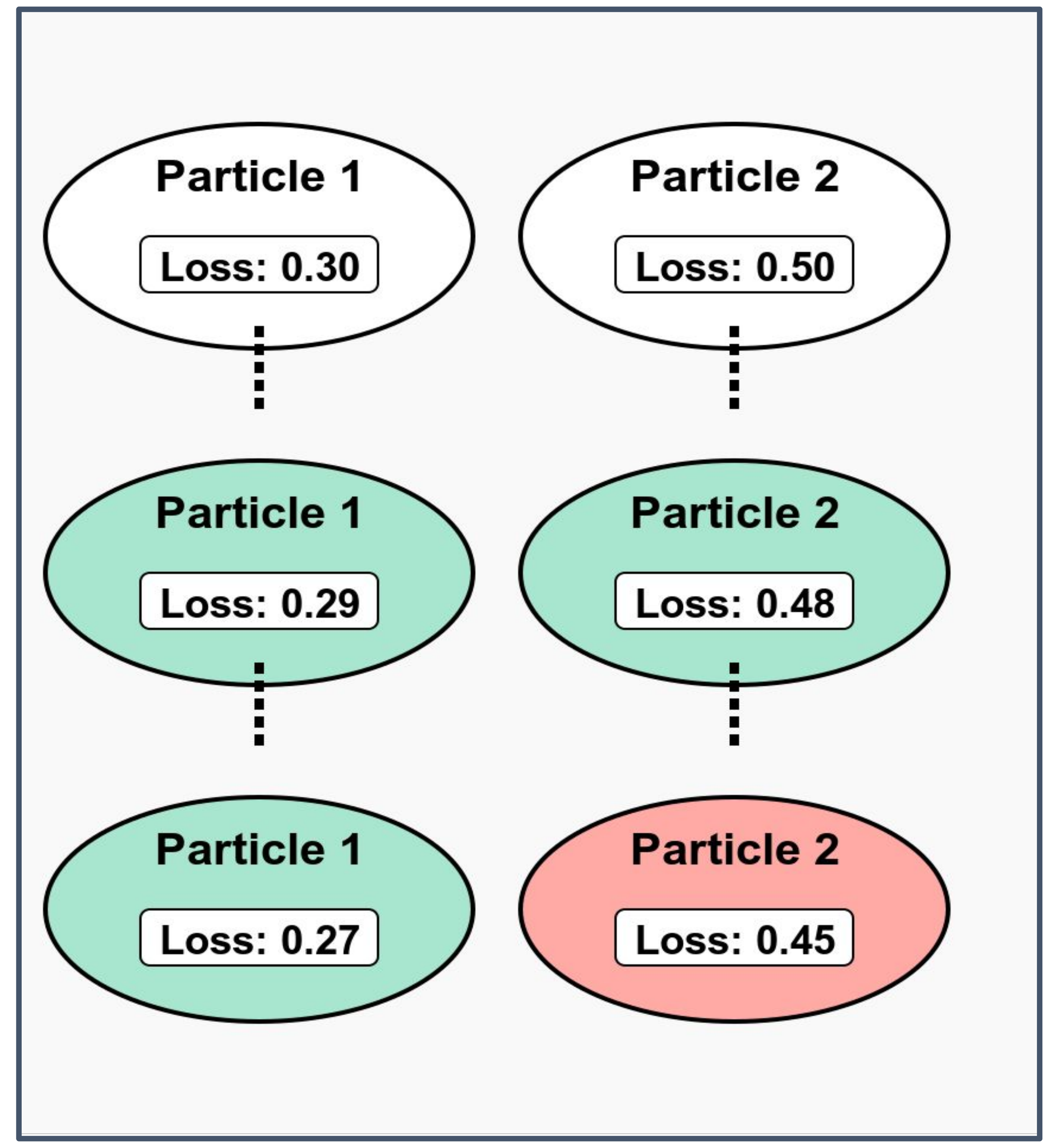
EXPERIMENTAL SETUP

Weight initialization uses pretrained model weights for initial setup. Further these weights are flattened into a particle vector.

Setup 1: PSO

Particles in PSO navigate the search space by adapting their trajectories based on the best-performing particle, dynamically refining their future movements.

Throughout the iterations while the overall loss decreases, suboptimal particles (high loss, low contribution) persist across generations due to PSO's lack of an elimination mechanism.

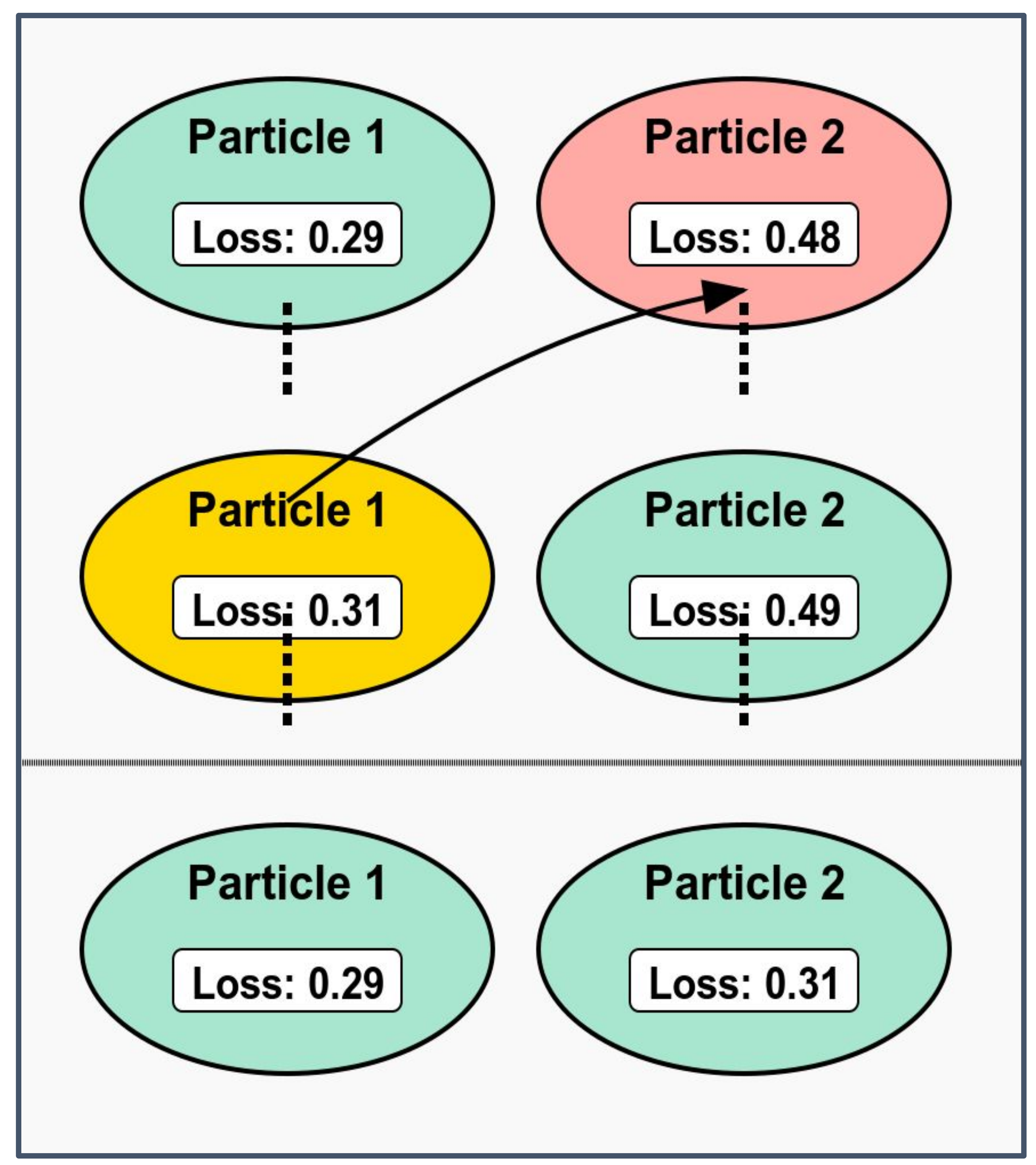


Setup 2: PSO + GA

The hybrid PSO-GA method combines GA's exploratory crossover and mutation with PSO's exploitative search, improving optimization efficiency.

Weak PSO particles are replaced by elite GA solutions, maintaining diversity and preventing premature convergence.

GA often exhibits slow convergence.



RESULTS

DATASETS	METRICS
IRIS - MLP	**Setup 1** Loss: 0.33 → 0.32 (↓ 3%) Accuracy: 93% → 93% (→ 0%) **Setup 2** Loss: 0.61 → 0.43 (↓ 29.5%) Accuracy: 89% → 93% (↑ 4.5%)
Wine - MLP	**Setup 1** Loss: 0.42 → 0.42 (→ 0%) Accuracy: 97% → 100% (↑ 3.1%) **Setup 2** Loss: 0.19 → 0.05 (↓ 73.7%) Accuracy: 97% → 100% (↑ 3.1%)
CIFAR- 10 - CNN	**Setup 1** Loss: 0.20 → 0.07 (↓ 65%) Accuracy: 95% → 98% (↑ 3.2%) **Setup 2** Loss: 0.20 → 0.11 (↓ 45.0%) Accuracy: 96% → 97% (↑ 1.0%)

Performance Comparison: Both setups improved accuracy, but Setup 2 (PSO + GA) achieved greater loss reduction, indicating better generalization.

DATASETS	METRICS
IRIS - MLP	Loss - 0.56 Accuracy - 90%
Wine - MLP	Loss- 0.16 Accuracy- 97%
CIFAR- 10 - CNN	Loss - 0.03 Accuracy - 99%

Benchmarking : Adam achieved similar results with more epochs, confirming gradient-based methods' efficiency and better fit for neural network training.