

Multilayer Perceptron

Comprehensive Demo and Analysis

*From-scratch NumPy implementation with backpropagation,
mini-batch gradient descent, and multiple activation functions.*

Seed: 42 | NumPy only | Validated against sklearn

Generated by demo.py

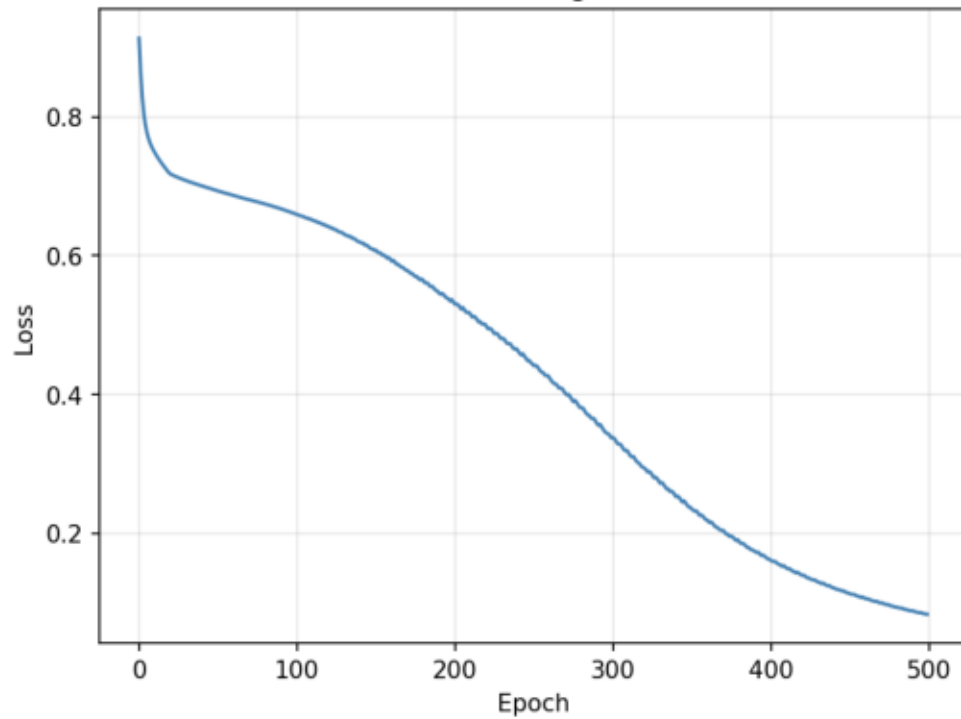
SUMMARY OF EXAMPLES

1. XOR Problem
The classic nonlinear classification problem that proves hidden layers can learn representations where classes become linearly separable.
2. Decision Boundary Visualization
Shows how MLP complexity (width and depth) affects the ability to carve nonlinear decision boundaries on moons and circles datasets.
3. Training Convergence
Demonstrates the effect of learning rate and architecture depth on convergence speed and final loss.
4. Weight Initialization Comparison
He vs Xavier initialization -- activation variance stability across layers and impact on training dynamics.
5. Sklearn Comparison
Validates our from-scratch MLP against sklearn's MLPClassifier, showing comparable accuracy on the moons dataset.
6. Real Dataset -- Breast Cancer
30-feature medical dataset with multiple architecture sizes. Demonstrates practical classification performance.
7. Activation Function Comparison
ReLU vs GELU vs SiLU vs Tanh -- same architecture, different activations, comparing convergence and decision boundaries.
8. Mini-batch vs Full-batch
Compares SGD (batch=1), mini-batch (32, 128), and full-batch training to show convergence noise vs speed tradeoffs.

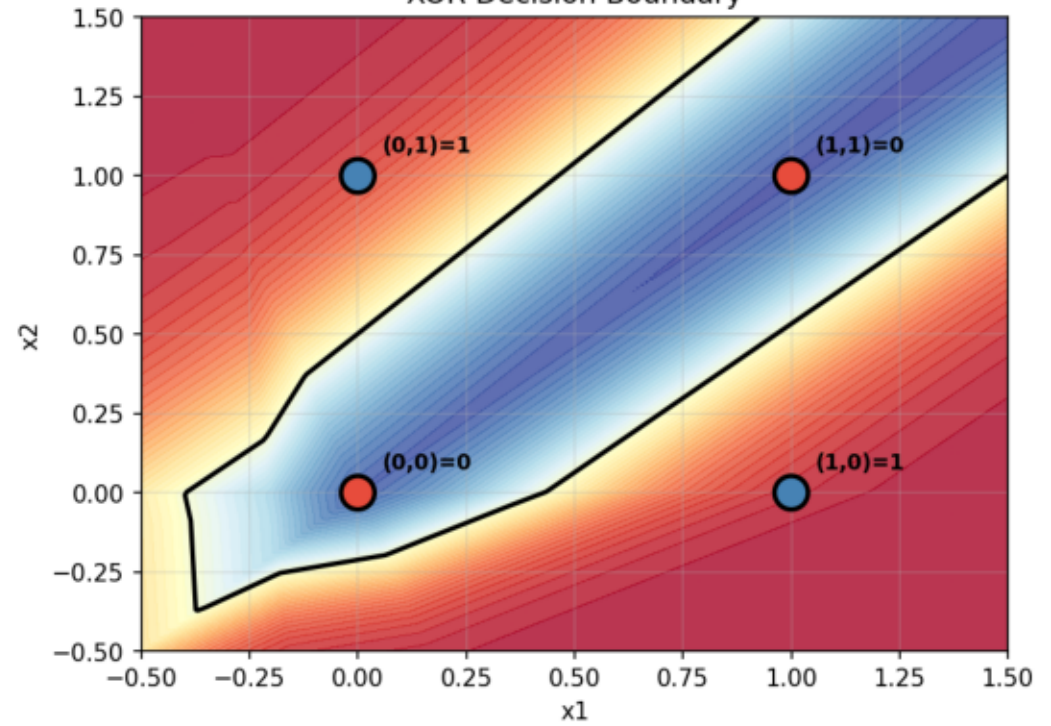
XOR Problem -- Classic Nonlinear Classification

XOR Problem -- Classic Nonlinear Classification

XOR Training Loss



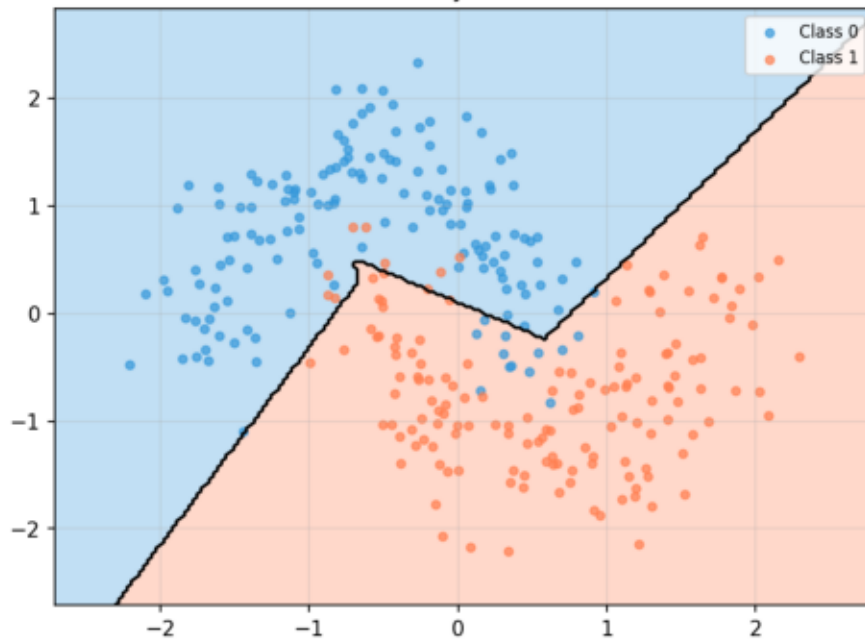
XOR Decision Boundary



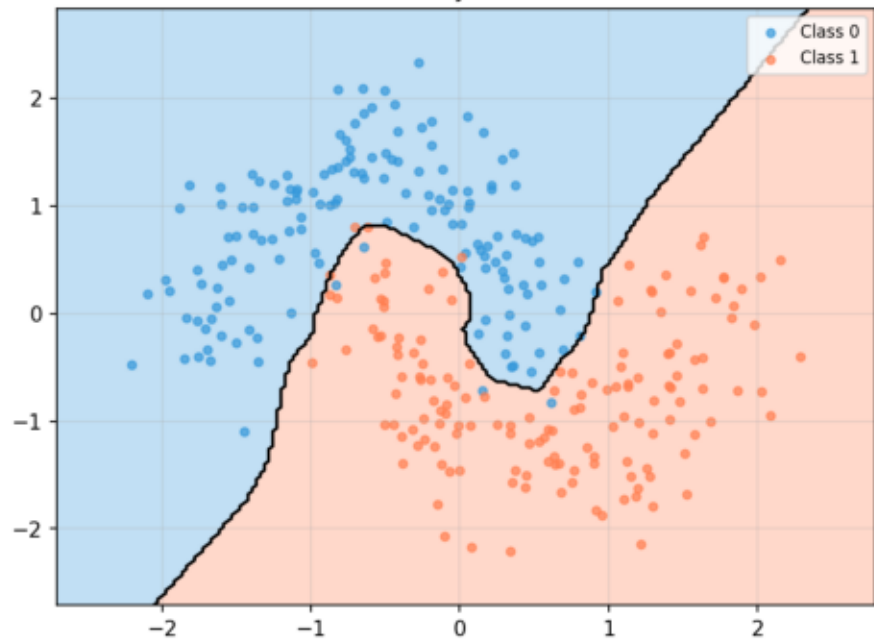
Decision Boundaries -- Dataset vs Architecture Complexity

Decision Boundaries -- Dataset vs Architecture Complexity

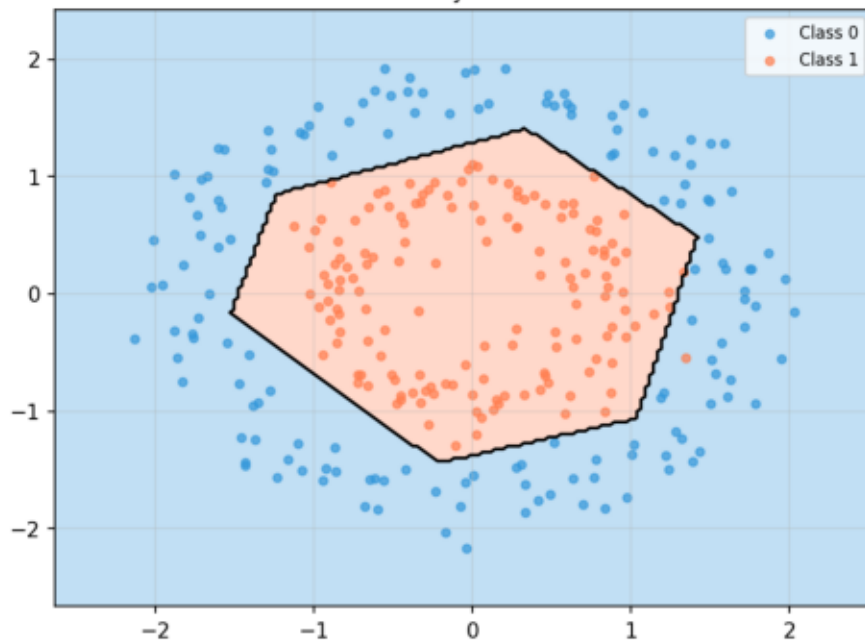
Moons -- Small (2-4-2)
Accuracy: 92.67%



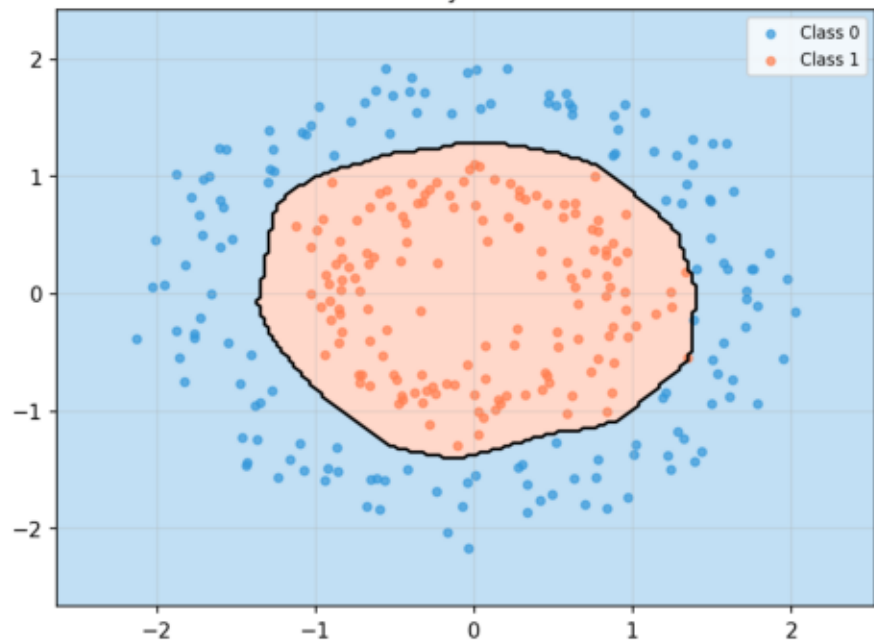
Moons -- Large (2-32-16-2)
Accuracy: 97.67%



Circles -- Small (2-4-2)
Accuracy: 99.67%



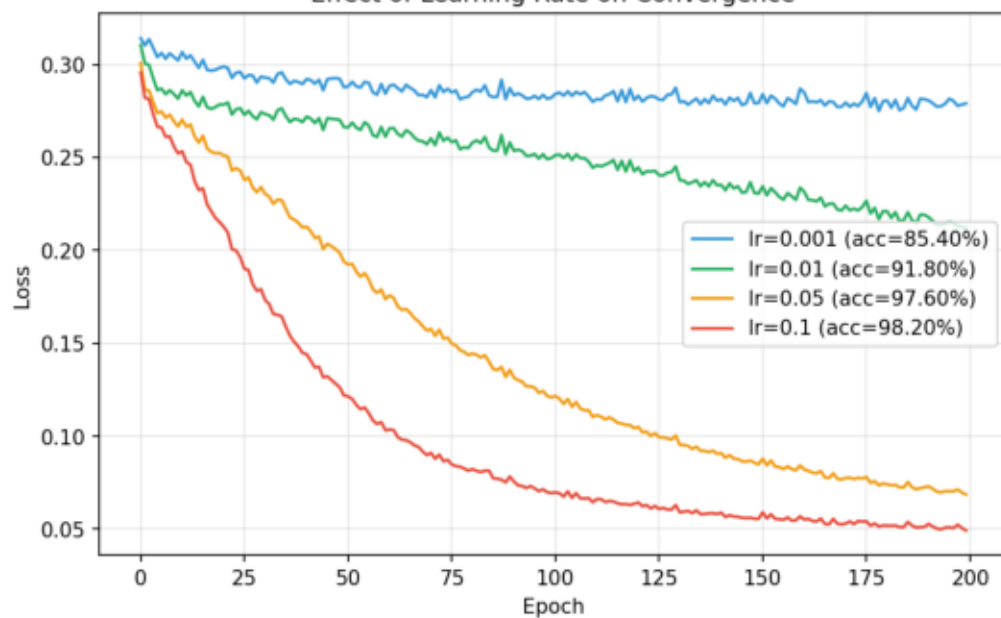
Circles -- Large (2-32-16-2)
Accuracy: 100.00%



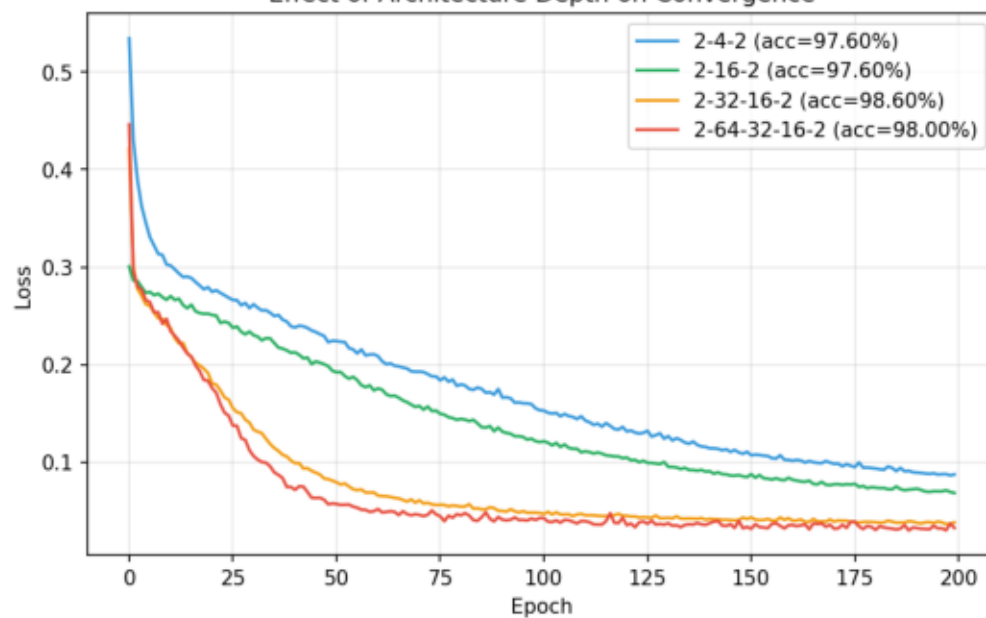
Training Convergence Analysis

Training Convergence Analysis

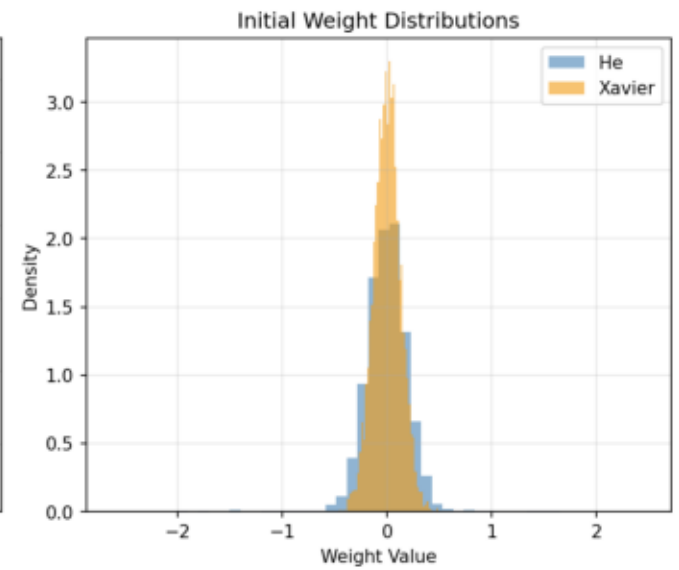
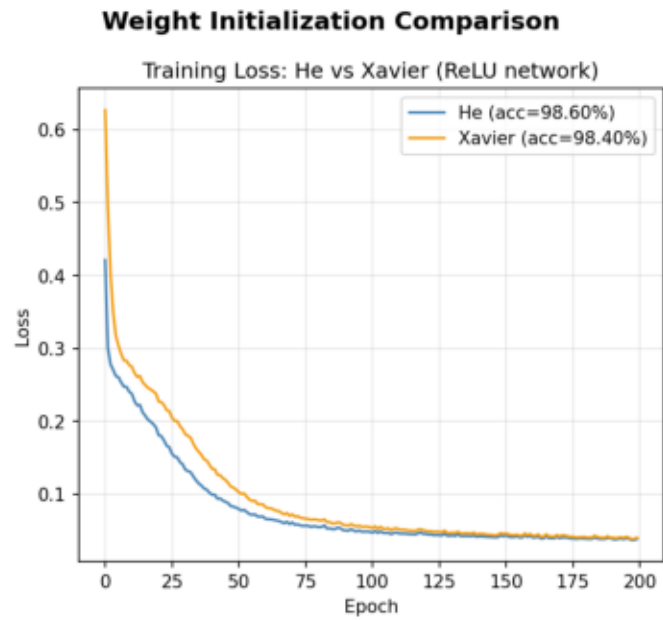
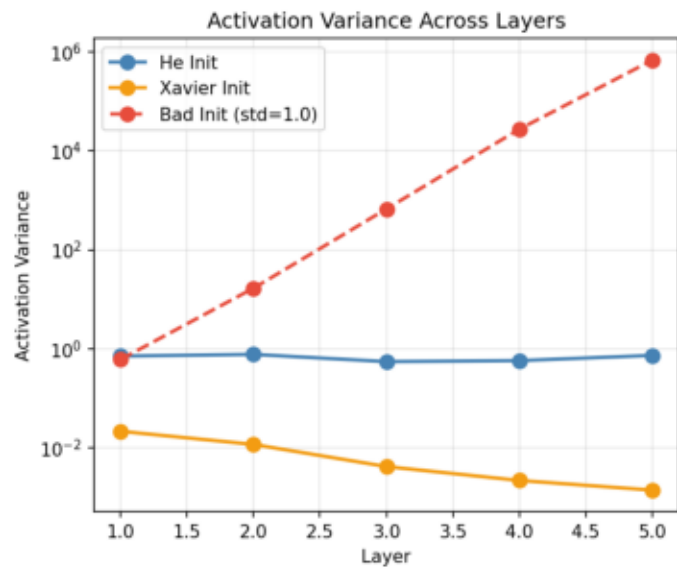
Effect of Learning Rate on Convergence



Effect of Architecture Depth on Convergence

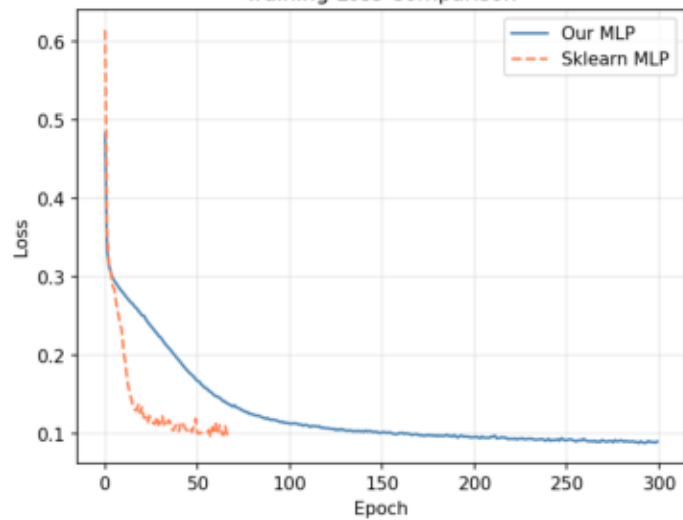


Weight Initialization Comparison

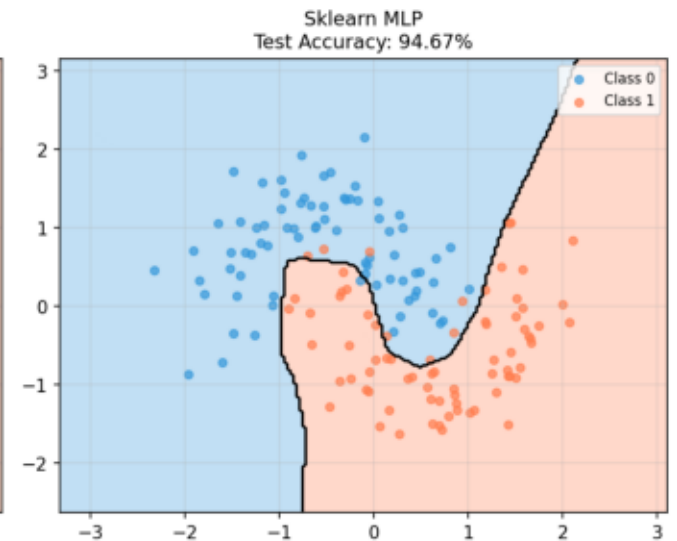
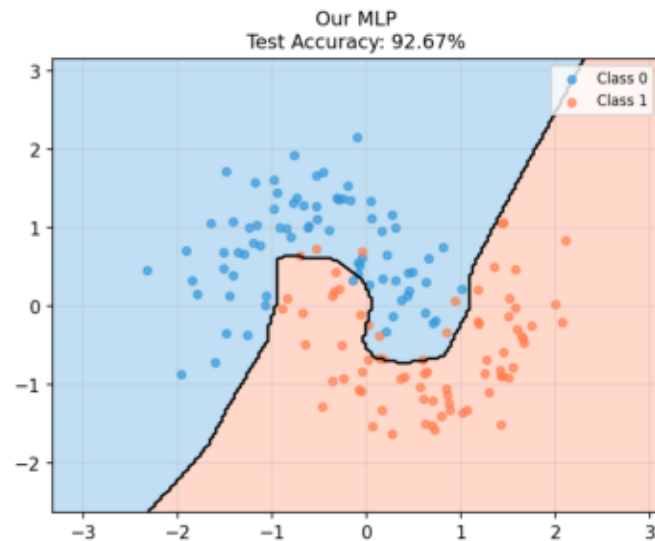


Our MLP vs Sklearn MLPClassifier

Training Loss Comparison

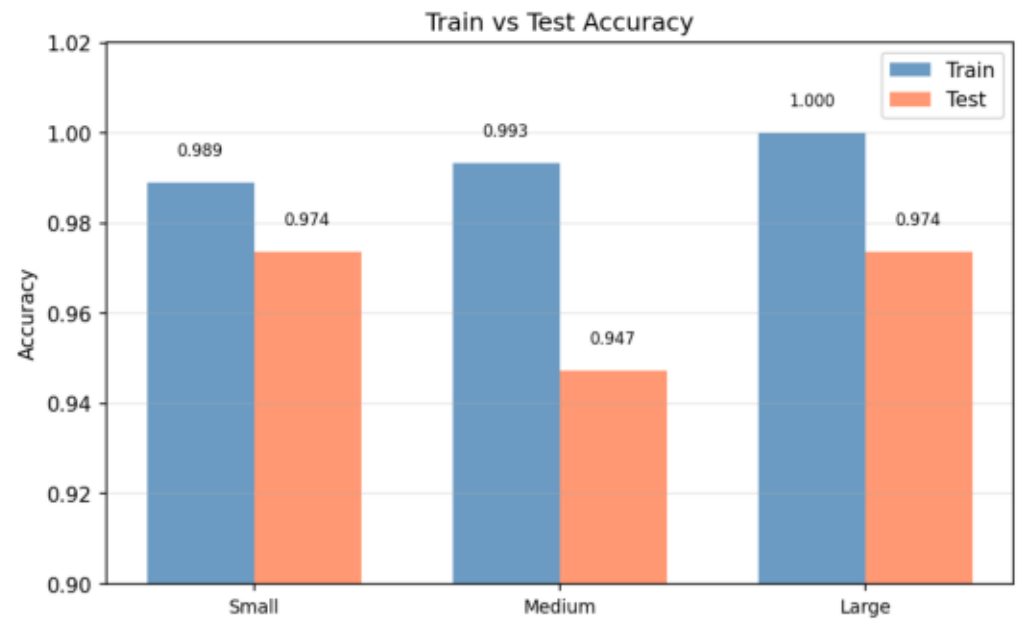
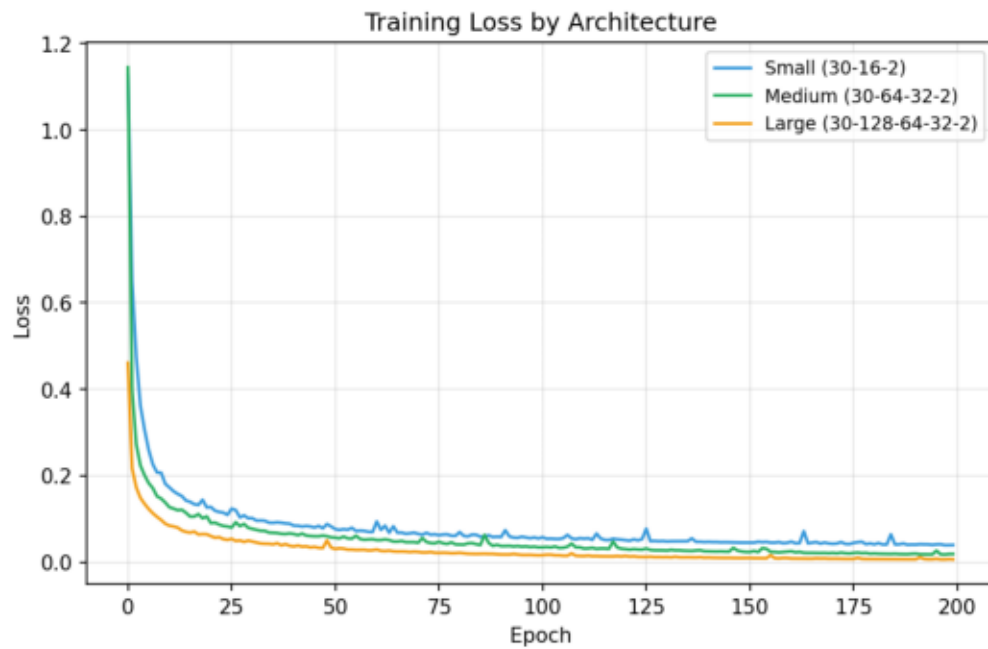


Our MLP vs Sklearn MLPClassifier



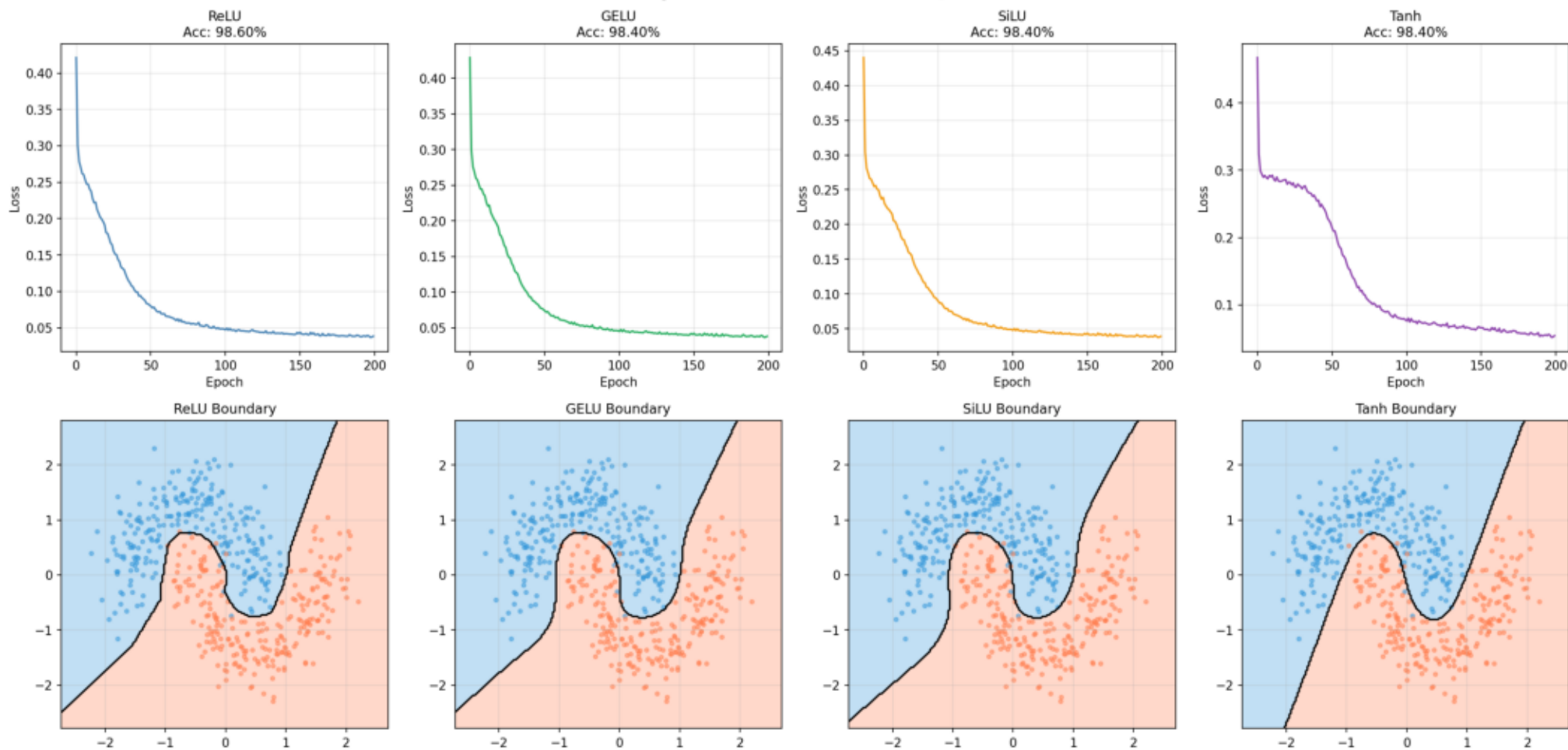
Breast Cancer Classification (30 features)

Breast Cancer Classification (30 features)



Activation Function Comparison -- Same Architecture, Different Activations

Activation Function Comparison -- Same Architecture, Different Activations



Mini-batch vs Full-batch Training

Mini-batch vs Full-batch Training

