Comprehensive Machine Learning Experiment Analysis: MNIST and Titanic Datasets
1. Introduction

This comprehensive report analyzes machine learning experiments across two distinct datasets - MNIST (handwritten digit recognition) and Titanic (passenger survival prediction) - comparing various activation functions, optimizers, and their performance characteristics.

2. Experiment Configurations

2.1 Common Activation Function Configurations

1. ReLU → Sigmoid → Tanh → Sigmoid/Softmax
2. ReLU → ReLU → ReLU → Sigmoid/Softmax
3. Sigmoid → Sigmoid → Sigmoid → Sigmoid/Softmax
4. Tanh → Tanh → Tanh → Sigmoid/Softmax

2.2 Optimizers Used

- ADAM
- SGD (Stochastic Gradient Descent)
- RMSPROP

3. MNIST Dataset Analysis

3.1 Performance Breakdown

1. **Experiment 2 (ReLU → ReLU → ReLU → Softmax)**
    - Best Performance
    - Peak Accuracy: 96.30%
    - Lowest Loss: 0.1300
    - Rapid and stable learning
2. **Experiment 3 (Sigmoid → Sigmoid → Sigmoid → Softmax)**
    - Second Best Performance
    - Peak Accuracy: 95.42%
    - Loss: Gradually decreasing
3. **Experiment 1 (ReLU → Sigmoid → Tanh → Softmax)**
    - Moderate Performance
    - Accuracy fluctuating around 91-93%
4. **Experiment 4 (Tanh → Tanh → Tanh → Softmax)**
    - Poorest Performance

    - Accuracy range: 83-85%
    - Highest and most unstable loss

4. Titanic Dataset Analysis

4.1 Optimizer and Activation Performance

1. **ADAM Optimizer**

    - Fastest convergence
    - Most stable loss reduction
    - Best performance across configurations
    - Lowest final test losses
2. **RMSPROP Optimizer**

    - Moderate performance
    - Faster initial learning compared to SGD
    - Slight performance variations
3. **SGD Optimizer**

    - Slowest convergence
    - Most stable but least aggressive learning
    - Consistently high initial losses

4.2 Activation Function Insights

- **ReLU-based configurations**: Faster learning
- **Sigmoid-only configurations**: Slower, more stable convergence
- **Mixed activation functions**: Showed nuanced performance characteristics

5. Comparative Performance Metrics

5.1 MNIST Dataset

- **Best Configuration**: ReLU → ReLU → ReLU → Softmax
    - Peak Accuracy: 96.30%
    - Lowest Loss: 0.1300
    - Rapid and consistent learning

5.2 Titanic Dataset

- **Best Configuration**: ADAM with ReLU → Sigmoid → Tanh → Sigmoid
    - Lowest final test loss
    - Most stable learning curve
    - Consistent performance across epochs

6. Key Observations

6.1 Activation Function Impact

- ReLU shows superior performance in complex classification tasks
- Mixed activation functions provide flexibility
- Tanh tends to perform less consistently

6.2 Optimizer Differences

- ADAM: Adaptive, fast convergence
- SGD: Stable but slow learning
- RMSPROP: Balanced performance

7. Recommendations

7.1 Model Selection

- For image classification (MNIST):
    1. Prefer ReLU-based deep networks
    2. Use ADAM optimizer
    3. Consider mixed activation functions
- For structured data (Titanic):
    1. Use ADAM optimizer
    2. Experiment with ReLU and mixed activations
    3. Monitor loss and accuracy closely

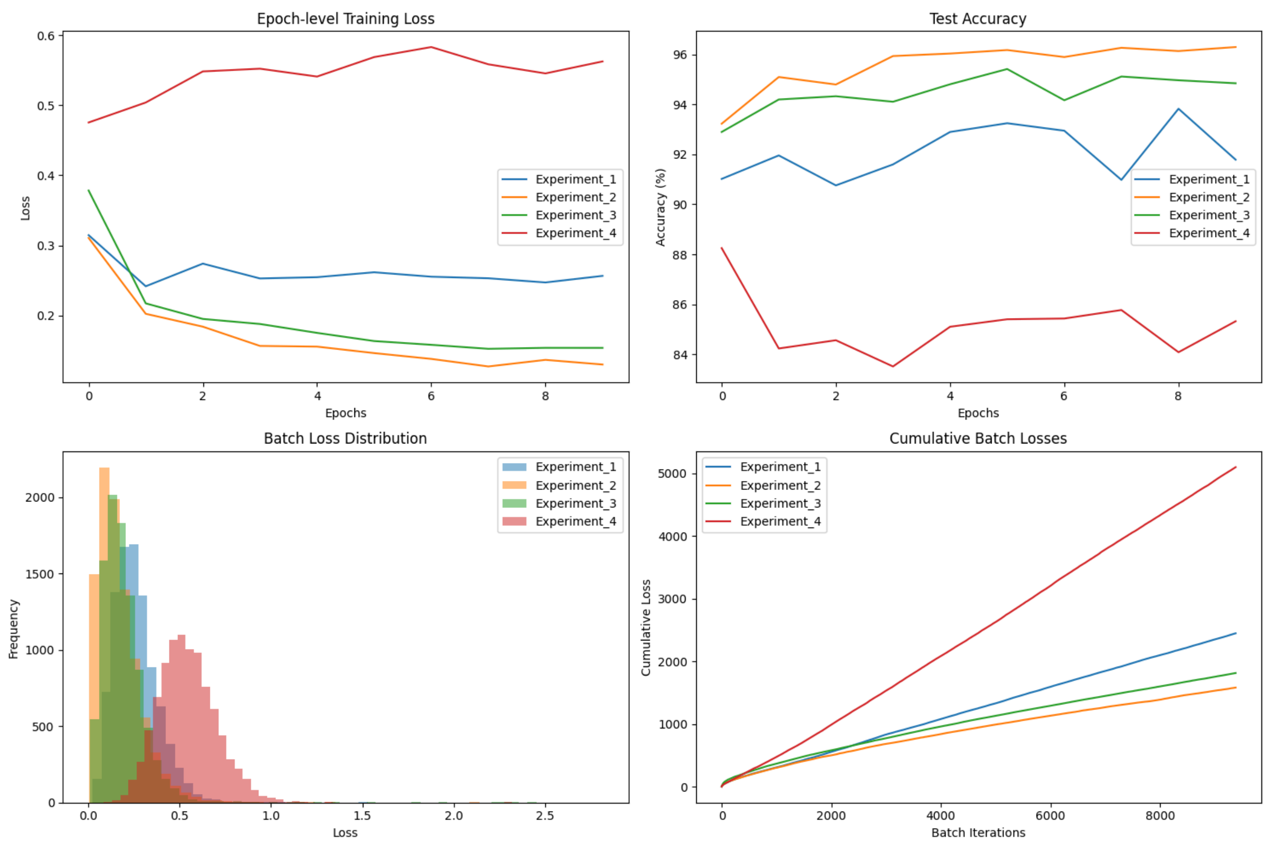
8. Limitations and Future Work

- Results specific to current datasets
- Recommend cross-validation
- Explore more activation function combinations
- Investigate hyperparameter tuning

9. Conclusion

The experiments demonstrate the critical role of activation functions and optimizers in machine learning model performance. Careful selection and configuration can significantly impact learning efficiency and predictive accuracy.

Mnist:



Titanic: