# Neural Network Experiment Report

## 1. Introduction

This report summarizes the findings from various neural network experiments conducted on the Kuzushiji-MNIST (KMNIST) dataset. The objective was to evaluate different hyperparameter configurations and determine the best-performing model based on validation accuracy. Additionally, the best model was further evaluated using cross-entropy and squared error loss functions.

## 2. Experimentation and Observations

We trained eight different models with various configurations of hyperparameters, including number of epochs, hidden layers, weight decay, learning rate, optimizer, batch size, and activation function. The table below presents the validation and test accuracy for each configuration:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Epochs | Hidden Layers | Weight Decay | Learning Rate | Optimizer | Batch Size | Activation | Validation Accuracy |
| 1 | 5 | [32, 32, 32] | 0.0005 | 0.001 | adam | 32 | relu | 0.922167 |
| 2 | 10 | [64, 64, 64, 64] | 0.0 | 0.0001 | sgd | 16 | sigmoid | 0.100667 |
| 3 | 5 | [128, 128, 128, 128, 128] | 0.5 | 0.001 | rmsprop | 64 | relu | 0.097833 |
| 4 | 10 | [32, 32, 32] | 0.0005 | 0.0001 | nadam | 32 | sigmoid | 0.7575 |
| 5 | 5 | [64, 64, 64, 64] | 0.0 | 0.001 | nesterov | 16 | relu | 0.922333 |
| 6 | 10 | [128, 128, 128, 128, 128] | 0.5 | 0.0001 | momentum | 64 | sigmoid | 0.101833 |
| 7 | 5 | [32, 32, 32] | 0.0005 | 0.001 | sgd | 32 | relu | 0.713833 |
| 8 | 10 | [64, 64, 64, 64] | 0.0 | 0.0001 | adam | 16 | sigmoid | 0.874833 |

## 3. Best Model and Confusion Matrix

The best model was identified based on validation accuracy. The model with the highest validation accuracy achieved 92.23%, using the following configuration:

- Epochs: 5  
- Hidden Layers: [64, 64, 64, 64]  
- Weight Decay: 0.0000  
- Learning Rate: 0.0010  
- Optimizer: Nesterov  
- Batch Size: 16  
- Activation Function: ReLU

The test accuracy for this model was 83.36%. Below is the confusion matrix for the best-performing model.

## 4. Comparison of Loss Functions

To further evaluate the performance of the best model, we compared cross-entropy loss with squared error loss. The results are as follows:

- Test Accuracy with Cross-Entropy Loss: 83.36%  
- Test Loss with Cross-Entropy: 0.5577  
- Test Accuracy with Squared Error Loss: 53.58%  
- Test Loss with Squared Error: 0.0619

From the results, we observe that cross-entropy loss performs significantly better than squared error loss. Squared error loss tends to have a smoothing effect that can result in poor performance in classification tasks.

## 5. Recommendations for MNIST Dataset

Based on the experiments conducted on the KMNIST dataset, we derive the following recommendations for the standard MNIST dataset:

1. \*\*Use ReLU Activation with Adam or Nesterov Optimizer\*\*: Models with ReLU activation and either Adam or Nesterov optimizers showed strong performance, achieving validation accuracy above 92%.

2. \*\*Avoid High Weight Decay\*\*: The models with high weight decay (e.g., 0.5) showed very poor performance, indicating that excessive regularization hurts learning.

3. \*\*Prefer Cross-Entropy Over Squared Error Loss\*\*: The test accuracy with squared error loss was significantly lower than with cross-entropy loss, making cross-entropy the preferred loss function for classification tasks.