*Comparative Analysis of Image Classifiers on MNIST Digits Dataset*

*Abstract*— This report aims to compare the performance of various image classifiers on the MNIST digits dataset. We construct a three-layer model with varying sizes for the second layer (hidden-units) and explore three different optimizer functions: SGD, Adam, and RMSprop. The activation function for the first two layers is ReLU, while the last layer utilizes SoftMax. The models are evaluated based on their accuracy and loss using the negative log likelihood (NLL) as the loss function.

# Introduction

Image classification is a fundamental task in computer vision, and selecting an appropriate model architecture and optimizer plays a crucial role in achieving high accuracy. This report investigates the impact of varying the size of the hidden units in the second layer and utilizing different optimizer functions on the accuracy and loss of image classifiers.

# Methodology

## Dataset

The MINST[1] digits dataset is a widely used benchmark dataset for image classification. It consists of 60,000 training examples and 10,000 testing examples, where each image is a 28x28 grayscale image of a handwritten digit from 0 to 9.

## Model Architecture

We construct a three-layer model for image classification with the following specifications:

* The first two layers use the rectified linear unit (ReLU[2]) activation function.
* The hidden layer (second layer) has varying sizes: 128, 256, 512, 1024, and 2048 units.
* The final layer uses the SoftMax[3] activation function.
* The loss function employed is the negative log-likelihood (NLL[4]) loss.

## Optimizer Functions

Three different optimizer functions are employed to train the models: Stochastic Gradient Descent (SGD)[5], Adam[6], and RMSprop[7]. The learning rate (lr) for each optimizer function is varied among [0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.5].

# Model Generation and evaluation

We generate multiple models (105 model) by combining the variations in the hidden layer size, optimizer function, and learning rate. For each combination, we train the model on the MINST digits dataset using the specified parameters.

We evaluate the performance of each model using two primary metrics:

* **Accuracy:** The percentage of correctly classified digits in the test set.
* **Loss:** The value of the NLL loss function on the test set.

# Conclusion

## Performance Comparison of Optimizers

The first observation focuses on the performance comparison of the three optimizers: SGD, Adam, and RMSprop. The (Figure 1) illustrates the accuracy and loss metrics for each optimizer across different learning rates. It is observed that SGD exhibits poor performance in both accuracy and losses compared to the other two optimizers. The accuracy achieved by SGD is consistently lower, and the losses are higher throughout the different learning rates. On the other hand, Adam and RMSprop show better accuracy and lower losses overall.

Additionally, an interesting finding is the high sensitivity of Adam and RMSprop to the learning rate. As the learning rate varies, there is a noticeable impact on both accuracy and losses. The choice of an appropriate learning rate becomes crucial for achieving optimal performance with these optimizers.

## Impact of "hidden\_units" Size and Learning Rate

The second observation focuses on the impact of "hidden\_units" size and learning rate on the model fitting progress over each training cycle (epoch). The (Figure 2) visualizes the accuracy trends for different optimizers and learning rates, considering varying "hidden\_units" sizes. It is observed that for SGD with different learning rates, there is a steady progress in accuracy as the training cycles increase. On the other hand, for Adam and RMSprop, the accuracy mainly depends on the learning rate, with a smaller influence of the number of training cycles.

In conclusion, the analysis of the image classifiers on the MINEST digits dataset reveals important insights. The comparison of optimizers indicates that SGD performs poorly compared to Adam and RMSprop. Furthermore, Adam and RMSprop demonstrate sensitivity to the learning rate, highlighting the importance of selecting an appropriate learning rate for optimal performance. The impact of "hidden\_units" size and learning rate on model fitting progress shows that SGD exhibits steady progress with increasing training cycles, while accuracy in Adam and RMSprop models is primarily influenced by the learning rate, with less significance given to the number of training cycles. These findings provide valuable guidance for configuring image classifiers and optimizing their performance on similar datasets in the future.

­­­A picture containing text, diagram, line, plot

Description automatically generated

Figure 1: Optimizer accuracy & Losses vs Learning Rate

A picture containing text, diagram, parallel, number

Description automatically generated

Figure 2: Accuracy vs Epoch

##### References

1. https://en.wikipedia.org/wiki/MNIST\_database
2. https://pytorch.org/docs/stable/generated/torch.nn.functional.relu.html#torch.nn.functional.relu
3. https://pytorch.org/docs/stable/generated/torch.nn.functional.softmax.html#torch.nn.functional.softmax
4. https://pytorch.org/docs/stable/generated/torch.nn.NLLLoss.html
5. https://pytorch.org/docs/stable/generated/torch.optim.SGD.html#torch.optim.SGD
6. https://pytorch.org/docs/stable/generated/torch.optim.Adam.html#torch.optim.Adam
7. https://pytorch.org/docs/stable/generated/torch.optim.RMSprop.html#torch.optim.RMSprop