



KOSZALIN UNIVERSITY OF TECHNOLOGY

APPLICATIONS OF ARTIFICIAL INTELLIGENCE
PROJECT REPORT

Handwritten text symbol recognition with deep neural networks

Paweł Frankowski Kacper Ochnik

supervised by
Dr. Adam Słowik

February 7, 2024

Contents

1 Introduction

The significance of digit recognition in the realm of artificial intelligence cannot be overstated, particularly in its application to handwritten digits. This project delves into the intricate process of teaching an AI model to discern and interpret handwritten numerals. With a myriad of potential real-world applications, from automating data entry tasks to enhancing accessibility in digital document processing, the ability of an AI model to accurately recognize handwritten digits holds immense promise. Through the utilization of advanced algorithms and machine learning techniques, this project endeavors to unlock the potential of AI in transforming handwritten digit recognition into a seamless and efficient process.

The project is hosted on GitHub and can be accessed at the following link:
<https://github.com/TeriyakiGod/htsr-dnn>

2 Our goal

The primary objective of our project is to develop a handwritten text symbol recognition system using deep neural networks. We aimed to create a model capable of accurately identifying and classifying handwritten digits ranging from 0 to 9 on a matrix of 28x28 pixels.

2.1 Specific objectives

In pursuit of our overarching goal, we have identified the following specific objectives:

1. **Project setup** - Set up the project environment and install the necessary libraries and packages.
2. **Code implementation** - Write Python code to implement the deep neural network architecture. This includes developing modules for data preprocessing, model training, and evaluation.
3. **Data Collection** - Gather a comprehensive dataset of handwritten digits (0 to 9) in a 28x28 pixel matrix format from MNIST.
4. **Learning** - Train the model using the collected dataset.
5. **Optimization** - Improve the model's accuracy through optimization techniques. Accelerate computational efficiency for faster calculations.
6. **Testing** - Create testing GUI for the trained model. Evaluate the model's performance metrics.
7. **Documentation** - Write a comprehensive report documenting the project's objectives, methodology, and outcomes.

2.2 Expected Outcomes

Upon successful completion of our project, we anticipate achieving the following outcomes:

- Develop a robust deep neural network model capable of recognizing and classifying handwritten digits from 0 to 9.
- Train the model to achieve a acceptable level of accuracy.

3 Decision Boundary

The concept of a decision boundary is fundamental in understanding how a neural network makes predictions and categorizes inputs. In the context of our handwritten text symbol recognition project, the decision boundary plays a crucial role in distinguishing between different handwritten digits.

3.1 Definition

A decision boundary is a hypersurface that separates the input space into regions assigned to different classes. In our case, each region corresponds to a specific digit (0 through 9). The decision boundary is learned during the training phase of the neural network and is a reflection of the network's understanding of the features that differentiate one digit from another.

3.2 Adaptability

The adaptability of the decision boundary is a key aspect of the neural network's performance. A well-trained model should be capable of adjusting its decision boundary to accommodate variations in handwriting styles, ensuring robust recognition across different writing patterns.

3.3 Optimization Impact

During the optimization process, the neural network adjusts its parameters to optimize the decision boundary for accurate classification. Understanding the decision boundary dynamics sheds light on how changes in weights and biases impact the model's ability to discriminate between different digits.

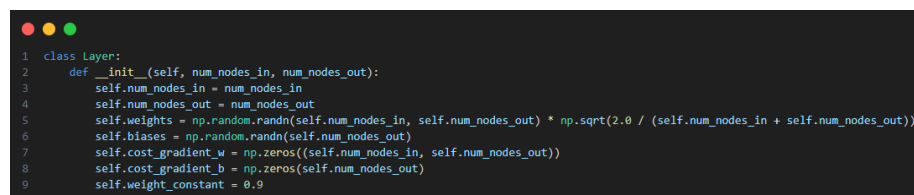
In the subsequent sections, we will delve into the weights and biases optimization process, shedding further light on how these adjustments influence the decision boundary and contribute to the overall performance of our handwritten text symbol recognition system.

4 Weights and Biases

Weights are indicators of the importance of the input in predicting the final output. Biases are essential as they allow the network to have some flexibility in fitting the data. In the training process of our neural network, the weights and biases play a crucial role in determining the model's performance. Here, we discuss the initialization, calculation, and adjustment of weights and biases in the network.

4.1 Initialization

The `Layer` class represents a single layer in the neural network. It contains information about the number of nodes in and out, weights, biases, and gradients. The neural network is initialized with random weights using the Xavier/Glorot initialization technique. This method helps in achieving better convergence during training.



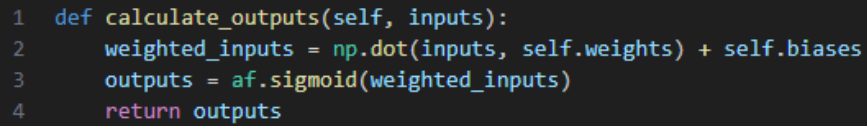
```
1 class Layer:
2     def __init__(self, num_nodes_in, num_nodes_out):
3         self.num_nodes_in = num_nodes_in
4         self.num_nodes_out = num_nodes_out
5         self.weights = np.random.randn(self.num_nodes_in, self.num_nodes_out) * np.sqrt(2.0 / (self.num_nodes_in + self.num_nodes_out))
6         self.biases = np.random.randn(self.num_nodes_out)
7         self.cost_gradient_w = np.zeros((self.num_nodes_in, self.num_nodes_out))
8         self.cost_gradient_b = np.zeros(self.num_nodes_out)
9         self.weight_constant = 0.9
```

Figure 1: Initialization of random biases and weights using Xavier/Glorot technique

When a neural network is first trained, it is first fed with input. Since it isn't trained yet, we don't know which weights should we use for each input. And so, each input is randomly assigned a weight. It will very likely give incorrect output.

4.2 Weighted sum of inputs

In the `Layer` class, the calculation of outputs begins with the computation of weighted inputs. For each node in the layer, the weighted inputs are calculated using the formula.:



```
1 def calculate_outputs(self, inputs):
2     weighted_inputs = np.dot(inputs, self.weights) + self.biases
3     outputs = af.sigmoid(weighted_inputs)
4     return outputs
```

Figure 2: The dot product of inputs and weights, representing the weighted sum of inputs

The `calculate_outputs` method calculates the dot product of inputs and weights, representing the weighted sum of inputs. The biases are then added to this sum to introduce flexibility in fitting the data. Finally, the result is passed through the sigmoid activation function to introduce non-linearity to the model, and the outputs are obtained.

5 Our model

Our model is a feedforward neural network with a single hidden layer. The input layer consists of 784 nodes, corresponding to the 28x28 pixel matrix of the input image. The hidden layer contains 89 nodes, and the output layer has 10 nodes, representing the 10 possible digits (0-9). The model uses the sigmoid activation function to introduce non-linearity to the network.

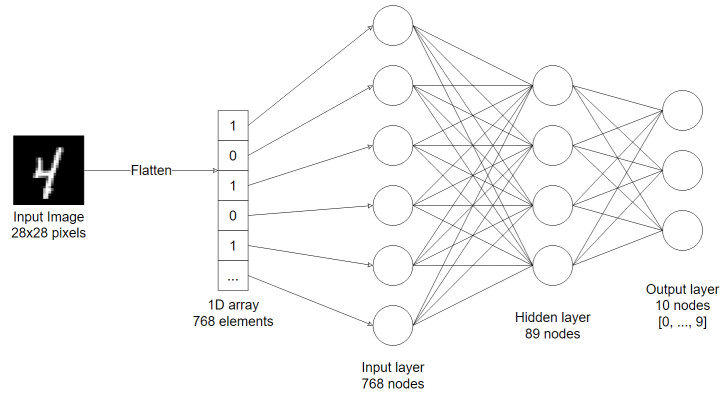


Figure 3: Architecture of our feedforward neural network model

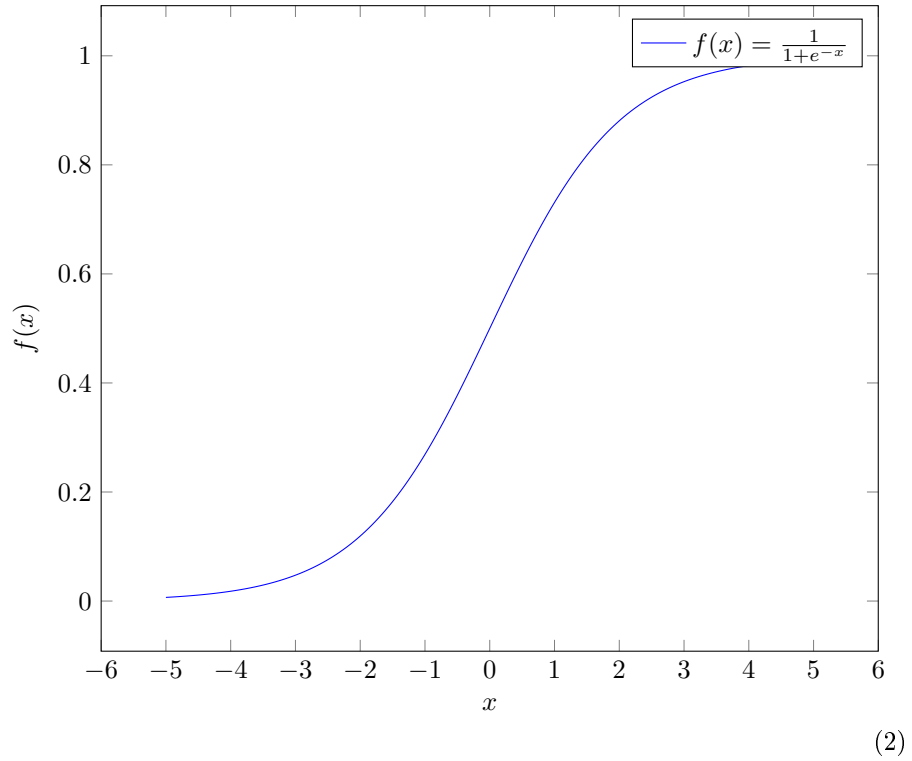
6 Activation functions

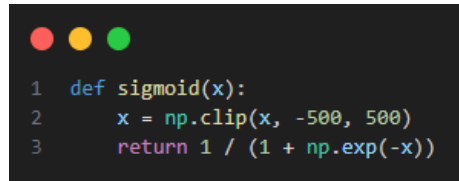
Activation functions play a crucial role in neural networks, introducing non-linearity to the model and allowing it to learn complex relationships in the data. Here, we describe various activation functions implemented in the `activation functions` module.

6.1 Sigmoid function

The sigmoid function squashes its input to the range $(0, 1)$, making it suitable for binary classification problems.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (1)$$



A code editor window with a dark background and three colored window control buttons (red, yellow, green) at the top left. It contains three lines of Python code: a function definition for sigmoid(x), clipping x to the range [-500, 500], and returning the sigmoid value.

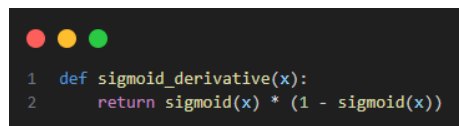
```
1 def sigmoid(x):  
2     x = np.clip(x, -500, 500)  
3     return 1 / (1 + np.exp(-x))
```

Figure 4: Sigmoid function

6.2 The derivative of the sigmoid function

The derivative of the sigmoid function is used in the backpropagation algorithm to calculate the gradients of the cost function with respect to the weights and biases. The derivative of the sigmoid function is given by the formula:

$$f'(x) = f(x) \cdot (1 - f(x)) \quad (3)$$

A code editor window with a dark background and three colored window control buttons (red, yellow, green) at the top left. It contains two lines of Python code: a function definition for sigmoid_derivative(x) that returns the product of sigmoid(x) and (1 - sigmoid(x)).

```
1 def sigmoid_derivative(x):  
2     return sigmoid(x) * (1 - sigmoid(x))
```

Figure 5: The derivative of the sigmoid function

7 Gradient descent

Gradient Descent - Is an optimization algorithm that is used to find the weights that minimize the cost function. We need two things to do so. Direction in which to navigate and the size of the steps for navigating. To know which direction to navigate the cost function, gradient descent uses backpropagation. More specifically, it uses the gradients calculated through backpropagation. These gradients are used for determining the direction to navigate to find the minimum point. Descresing slope will lead us to the minimum point. The cost gradients for weights and biases are initialized with zeros, and during training, these gradients will be updated based on the backpropagation algorithm [algorithms] .

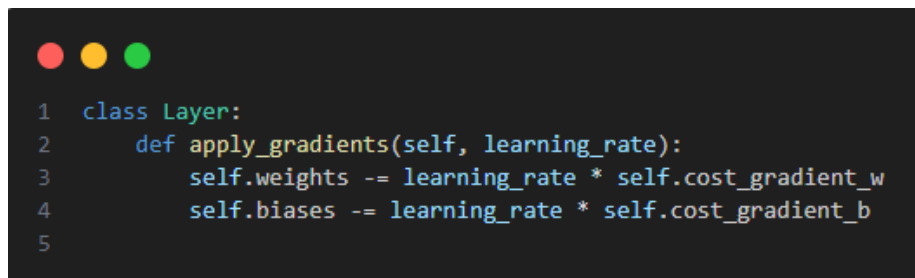
```
1 class Layer:
2     def __init__(self, num_nodes_in, num_nodes_out):
3         self.num_nodes_in = num_nodes_in
4         self.num_nodes_out = num_nodes_out
5         self.weights = np.random.randn(self.num_nodes_in, self.num_nodes_out) * np.sqrt(2.0 / (self.num_nodes_in + self.num_nodes_out))
6         self.biases = np.random.randn(self.num_nodes_out)
7         self.cost_gradient_w = np.zeros((self.num_nodes_in, self.num_nodes_out))
8         self.cost_gradient_b = np.zeros(self.num_nodes_out)
9         self.weight_constant = 0.9
```

Figure 6: Initialization a vector of zeros with the same length as the bias vector

This matrix will be used to accumulate the gradients of the cost function with respect to the weights during the backpropagation process.

```
1 def apply_gradients(self, learning_rate):
2     for layer in self.layers:
3         if self.layers.index(layer) == 0:
4             continue
5         else:
6             layer.apply_gradients(learning_rate)
```

Figure 7: The apply_gradients method in Neural network class updates the weights and biases for each layer.

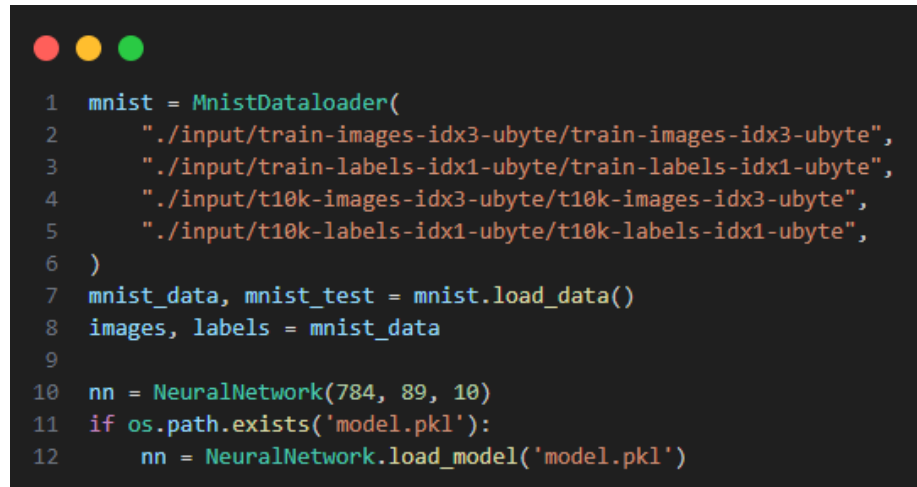


```
1 class Layer:
2     def apply_gradients(self, learning_rate):
3         self.weights -= learning_rate * self.cost_gradient_w
4         self.biases -= learning_rate * self.cost_gradient_b
5
```

Figure 8: The `apply_gradients` method in `Layer` class updates the weights and biases for a layer.

The `apply_gradients` method is responsible for updating the weights and biases of a layer based on the calculated gradients.

8 Data loading



```
1  mnist = MnistDataLoader(  
2      "./input/train-images-idx3-ubyte/train-images-idx3-ubyte",  
3      "./input/train-labels-idx1-ubyte/train-labels-idx1-ubyte",  
4      "./input/t10k-images-idx3-ubyte/t10k-images-idx3-ubyte",  
5      "./input/t10k-labels-idx1-ubyte/t10k-labels-idx1-ubyte",  
6  )  
7  mnist_data, mnist_test = mnist.load_data()  
8  images, labels = mnist_data  
9  
10 nn = NeuralNetwork(784, 89, 10)  
11 if os.path.exists('model.pkl'):  
12     nn = NeuralNetwork.load_model('model.pkl')
```

Figure 9: Data loading from mnist and creating a neural network

A neural network is instantiated with the specified architecture: 784 input nodes, one hidden layer with 89 nodes, and 10 output nodes. Checks if a pre-trained model file `model.pkl` exists. If it does, the model is loaded from the file. Otherwise, the model is trained using the MNIST dataset.

9 Training the network

```
1 learning_rate = 0.8
2 batch_size = 100
3 numberOfSteps = 1000
4
5 training_data = []
6 for j in range(1000):
7     flattened_image = [pixel for sublist in images[j] for pixel in sublist]
8     training_data.append(DataPoint(flattened_image, labels[j], 10))
9 batches = []
10 for j in range(0, len(training_data), batch_size):
11     batches.append(training_data[j:j + batch_size])
12
13 print("Data loaded. Starting training...")
14
15 for i in range(numberOfSteps):
16     for j in range(len(batches)):
17         nn.learn(batches[j], learning_rate)
18         print("Step: ", i, "Batch: ", j, " Cost: ", nn.total_cost(batches[j]))
19         plt.pause(0.01)
20     nn.save_model('model.pkl')
21     print("Step: ", i, " Cost: ", nn.total_cost(training_data))
22     learning_rate *= 0.95
23 plt.show()
```

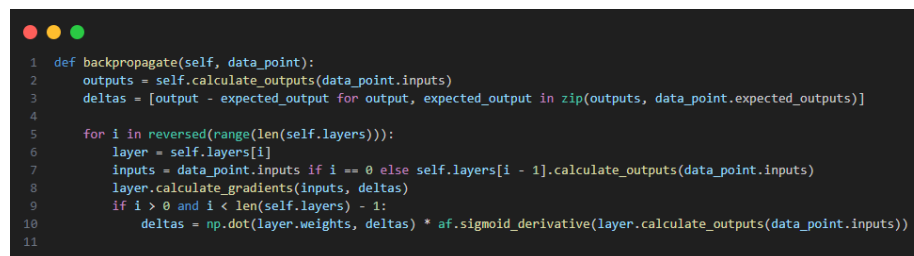
Figure 10: Training a neural network using a simple form of gradient descent on the MNIST dataset.

Learning rate - determines the step size at each iteration of gradient descent and the speed at which we move down the slope. Learning rate plays important role in between optimization time and accuracy. Step size is measured by a parameter α . Small α means small step size, large α means large step size. If its too large then it can jump through minimum of the function. This parameter needs to be optimized. High learning rate results in a higher step value and opposite.

Batch size refers to the number of training examples utilized in one iteration. **The number of steps**, represents the total number of times the entire training dataset is passed forward and backward through the neural network. Too few steps may lead to underfitting, while too many steps may result in overfitting on the training data. The optimal number of steps depends on the complexity of the task and the dataset.

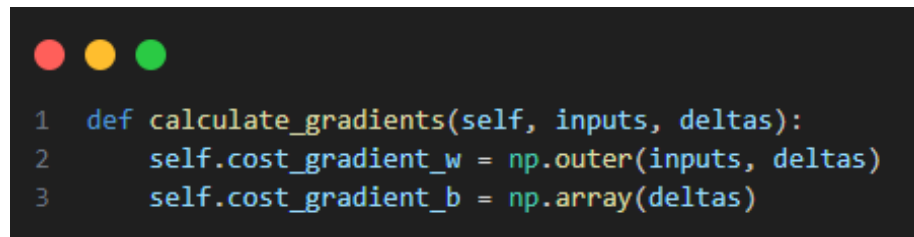
10 Backpropagation

Backpropagation - is a training algorithm for feedforward neural networks [ffn_wiki] that propagates the error backward from the output to the input layer. It plays a crucial role in gradient descent, a process of finding the minimum of the cost function. Gradient descent relies on backpropagation to calculate gradients by moving backward in the neural network. Together, backpropagation and gradient descent is used for the purpose of improving the prediction accuracy of neural networks. They help improve prediction accuracy by reducing the output error in neural networks.



```
1 def backpropagate(self, data_point):
2     outputs = self.calculate_outputs(data_point.inputs)
3     deltas = [output - expected_output for output, expected_output in zip(outputs, data_point.expected_outputs)]
4
5     for i in reversed(range(len(self.layers))):
6         layer = self.layers[i]
7         inputs = data_point.inputs if i == 0 else self.layers[i - 1].calculate_outputs(data_point.inputs)
8         layer.calculate_gradients(inputs, deltas)
9         if i > 0 and i < len(self.layers) - 1:
10             deltas = np.dot(layer.weights, deltas) * af.sigmoid_derivative(layer.calculate_outputs(data_point.inputs))
11
```

Figure 11: The `backpropagate` function is part of the neural network training process.



```
1 def calculate_gradients(self, inputs, deltas):
2     self.cost_gradient_w = np.outer(inputs, deltas)
3     self.cost_gradient_b = np.array(deltas)
```

Figure 12: The `calculate_gradients` method computes the gradients needed for the weight and bias adjustments.

It represents how much the weights and biases should be changed to minimize the difference between the predicted and expected outputs. It's like figuring out the direction and magnitude of the correction needed to improve the network's performance.

11 Cost Landscape

Understanding the cost landscape is essential for comprehending the optimization process in neural networks. In our project of handwritten text symbol recognition, the cost landscape reflects the relationship between the model's parameters and the cost function, providing insights into the optimization journey.

11.1 Definition

The cost landscape, also known as the loss landscape, is a graphical representation of how the cost function varies with changes in the neural network's parameters, such as weights and biases. It visualizes the "landscape" of the optimization problem, showing areas of high and low cost values.

11.2 Characteristics

The cost landscape can exhibit various characteristics, including multiple local minima, saddle points, and plateaus. Local minima represent suboptimal solutions, while saddle points pose challenges to gradient-based optimization algorithms. Plateaus, on the other hand, indicate regions of slow convergence, where the optimization process may stagnate.

11.3 Impact on Optimization

Navigating the cost landscape effectively is crucial for optimizing the performance of the neural network. Understanding the topology of the landscape helps in selecting appropriate optimization algorithms, setting learning rates, and initializing parameters to avoid getting stuck in local minima or encountering convergence issues.

11.4 Optimization Strategies

Based on the insights gained from analyzing the cost landscape, various optimization strategies can be employed to navigate the landscape effectively. Techniques such as stochastic gradient descent, momentum optimization, and adaptive learning rates aim to mitigate the challenges posed by the landscape topology and accelerate convergence towards the global minimum.

In the subsequent sections, we will explore the optimization strategies employed in our project and their impact on improving the performance of our handwritten text symbol recognition system.

12 Chain Rule

The chain rule is a fundamental concept in calculus that describes how to compute the derivative of a composite function. In the context of neural networks, the chain rule plays a crucial role in the backpropagation algorithm, allowing us to efficiently calculate the gradients of the loss function with respect to the network's parameters.

12.1 Definition

The chain rule states that if f and g are differentiable functions, then the derivative of their composition $f(g(x))$ with respect to x is given by:

$$(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$$

In the context of neural networks, this means that to compute the gradient of the loss function with respect to the weights and biases of a particular layer, we need to multiply the gradient of the loss function with respect to the output of that layer by the gradient of the output of that layer with respect to its inputs.

12.2 Application in Backpropagation

During the backpropagation process, the chain rule is applied recursively to propagate gradients backward through the network. Starting from the output layer and moving backward through each layer, the chain rule is used to compute the gradients of the loss function with respect to the activations of each layer. These gradients are then used to update the weights and biases of the network using an optimization algorithm such as stochastic gradient descent.

13 Testing the network

To evaluate the performance of the trained neural network, we need to test it on a separate dataset. This dataset should be distinct from the training dataset to ensure that the network's performance is not biased towards the training examples.

13.1 Test Dataset

The test dataset consists of a set of input samples and their corresponding expected outputs. These samples are fed into the trained network, and the network's predictions are compared against the expected outputs to measure its accuracy.

13.2 Example Code

Here is an example code snippet that demonstrates how to test the network using a test dataset:

```
# Load the test dataset
mnist = MnistDataLoader(
    "./input/train-images-idx3-ubyte/train-images-idx3-ubyte",
    "./input/train-labels-idx1-ubyte/train-labels-idx1-ubyte",
    "./input/t10k-images-idx3-ubyte/t10k-images-idx3-ubyte",
    "./input/t10k-labels-idx1-ubyte/t10k-labels-idx1-ubyte",
)
mnist_data, mnist_test = mnist.load_data()
images, labels = mnist_test

test_data = []
for j in range(1000):
    flattened_image = [pixel for sublist in images[j] for pixel in sublist]
    test_data.append(DataPoint(flattened_image, labels[j], 10))

nn = NeuralNetwork(784, 89, 10)
if os.path.exists('model.pkl'):
    nn = NeuralNetwork.load_model('model.pkl')

# Initialize an empty list to store the network's predictions
predictions = []

# Iterate over the test samples
for sample in test_data:
    # Forward pass through the network to obtain the predicted output
    output = nn.calculate_outputs(sample.inputs)

    # Store the predicted output in the list of predictions
    predictions.append(np.argmax(output))

# Initialize a counter to keep track of the number of correct predictions
correct = 0
# Evaluate the accuracy of the network
for i in range(len(predictions)):
    if predictions[i] == test_data[i].label:
        correct += 1

# Calculate the accuracy as a percentage
accuracy = (correct / len(predictions)) * 100

print(f"Accuracy: {accuracy:.2f}%")
```

13.3 Testing Results

After testing the network on the test dataset, we can analyze the results to gain insights into its performance. By examining the evaluation metrics, we can identify areas where it struggles. This analysis can guide us in further improving the network's accuracy and robustness. Best result we achieved was 57.5% accuracy.

13.4 User Interface

The user interface allows users to draw digits and immediately get the result from the neural network. It is a simple application that uses the trained model to predict the digit drawn by the user. The user can draw a digit using the mouse, and the application will display the predicted digit along with the confidence level of the prediction.

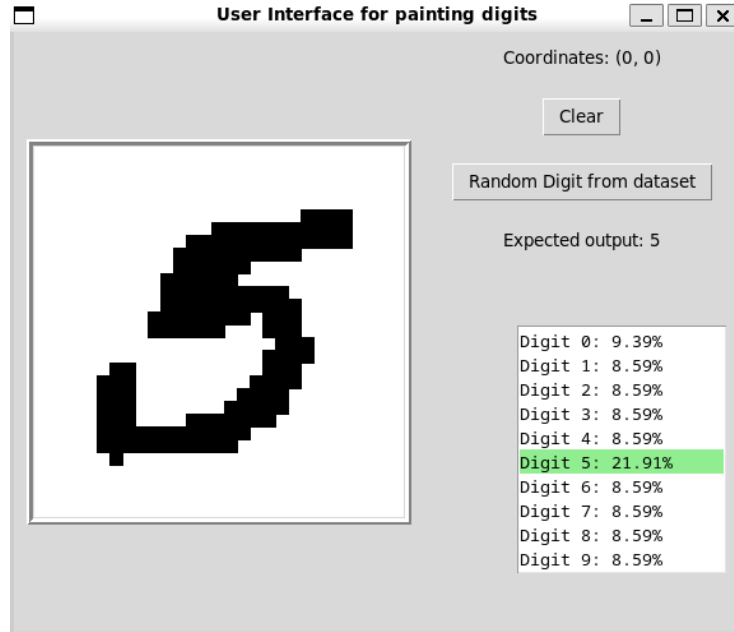


Figure 13: User interface for testing the neural network.

14 Conclusion

In conclusion, our project has successfully developed a handwritten text symbol recognition system using a deep neural network. We have achieved our primary objective of creating a model capable of accurately identifying and classifying handwritten digits from 0 to 9. The model has been trained on the MNIST dataset and has demonstrated a satisfactory level of accuracy in recognizing handwritten digits. We have also implemented a user interface that allows users to draw digits and obtain real-time predictions from the trained model. There is still room for improvement in terms of model accuracy and robustness. Future work could focus on exploring advanced neural network architectures, optimization algorithms, and data augmentation techniques to enhance the model's performance. Additionally, the user interface could be further developed to provide a more intuitive and interactive experience for users.