Name: Tatiana Prado Santos

Datasheet for the MNIST Dataset

1. Motivation

- Why was the dataset created? The MNIST dataset was created as a benchmark for evaluating machine learning models on the task of digit recognition. It serves educational purposes and helps in pushing the development of better, more accurate models.

2. Composition

- What do the instances represent? Each instance is a 28x28 pixel grayscale image of handwritten digits from 0 to 9.
- How many instances are there? There are 60,000 training images and 10,000 test images in the dataset.

3. Collection Process

- How was the data collected? The images were collected from a mix of Census Bureau employees and high school students.
- Who was involved in the collection process? The dataset was created by the National Institute of Standards and Technology (NIST).

4. Data Preprocessing

- What preprocessing/cleaning was done? The images were normalized in size and centered. Each image is encoded as a 28x28 matrix where each cell contains grayscale pixel value.

5. Uses

- What tasks should the dataset be used for? The MNIST dataset is intended for use in training and testing in the field of machine learning for the task of digit recognition.
- What tasks should it not be used for? This dataset should not be used for tasks involving personal identification or decisions that could impact people's lives or livelihoods, as it does not contain any personal information.

6. Distribution

- Who can access the dataset? The dataset is publicly available and can be accessed by anyone for academic and educational purposes.
- Are there any restrictions? No, there are no restrictions on the usage of this dataset.

7. Maintenance

- How is the dataset maintained? The dataset is static and does not require ongoing maintenance. Updates and modifications are not typically made.

Name: Tatiana Prado Santos

Model Card for the Digit Recognizer Model

1. Model Details

- Who developed the model, and when? The model was developed by Tatiana Prado Santos as part of the capstone project for the ICL certification course in 2024.
- What type of model is it? It is a convolutional neural network designed for image classification tasks.

2. Intended Use

- What are the intended applications? The model is intended for educational purposes and to demonstrate the capabilities of neural networks in digit recognition from image data.
- Who are the intended users? Students, educators, and researchers in machine learning.

3. Factors

- What factors does the model consider? The model considers the pixel values of the images as input features without any demographic or ethical implications.

4. Metrics

- What metrics are used to measure performance? The model's performance is measured using accuracy, specifically the percentage of correctly predicted images.

5. Evaluation Data

- Describe the datasets used for evaluation. The model is evaluated using a split from the original MNIST dataset, which includes a separate set of 10,000 test images.

6. Training Data

- Describe the training data. The training data consists of 60,000 images from the MNIST dataset, each labeled with the correct digit.

7. Ethical Considerations

- Discuss potential ethical concerns in model application. There are minimal ethical concerns as the data does not contain personal or sensitive information.

8. Caveats and Recommendations

- Any additional considerations or usage recommendations. Users should ensure that the model is only used for educational and research purposes due to its limitations in understanding complex or ambiguous handwritten digits.