model-card-example.md

Model card for MNIST Machine Learning Course Project

Model details

- Developed by David Drexel and other teaching assistants of the bachelor's course in Machine Learning at the University of Innsbruck
- Classifies images of the digits 0 to 9
- Convolutional Neural Network
- Developed in 2023, v1
- See report for full details on training methods and parameters

Intended use

- Intended for classifying images of single digits
- Not intended to classify multi-digit numbers
- Should to serve as an example for the final project of the bachelor course in Machine Learning
- Not intended for commercial use

Factors

- Model trained on black and white images
- Some digits coincide signficantly in appearance (e.g. 3, 5) depending on writing style; error rates for these digits were higher

Training data

- Images are from the MNIST database [1]
- Frequency for each class ranging from 9.0% (5) to 11.3% (1)
- Training was done on 50000 images, 10000 were split off for a validation dataset
- Each image was symmetrically padded to 32x32 and whitening was applied to the images

Evaluation data

- Test dataset of 10000 digits is also available in the MNIST dataset
- Digit frequencies ranged from 8.9% (5) to 11.4% (1)
- Padding applied to the test images similarly to the training set, but no whitening applied

Metrics

- Accuracy of 98.69% on the test dataset
- Precision and recall metrics captured for each class

Quantitative analyses

Precision and Recall of Predicting Classes of Digits 0-9
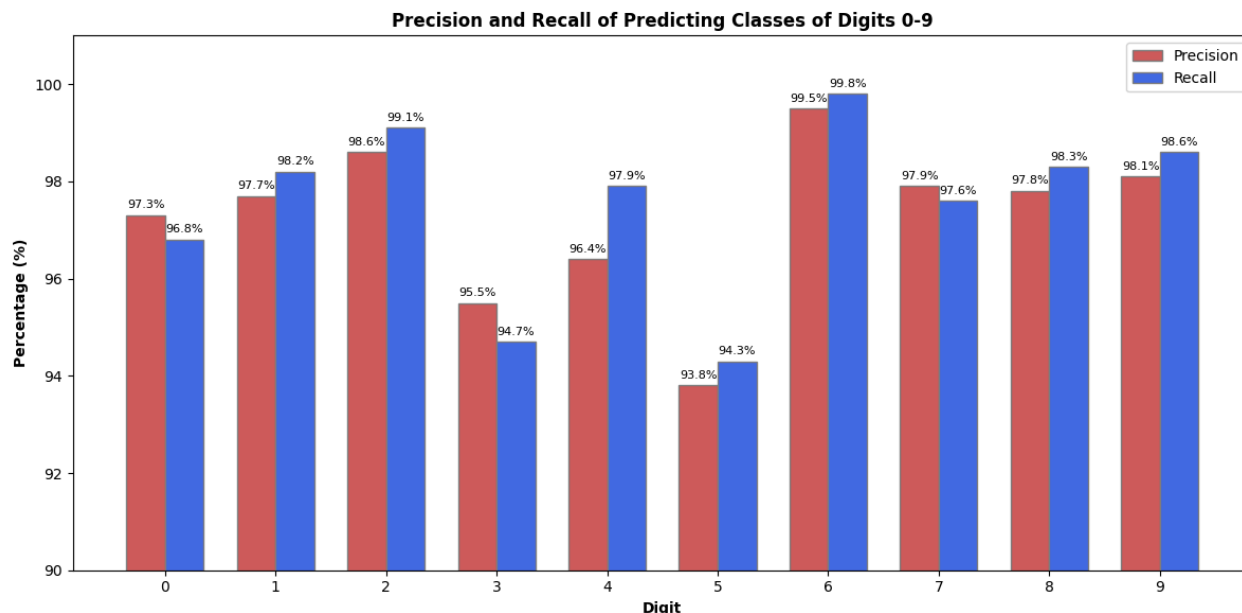
Ethical considerations

- The dataset was gathered in the the United States, so writing style reflects American conventions and may perform worse outside of this region. For example, Europeans tend to add a dash across the stem of a 7 to differentiate it from their 1, which contains a second stroke. The digit 1 in this dataset tends to only have one stroke, which is more typical in the US. This model may misclassify a more typically European styled 1 as a 7.

Caveats and recommendations

- This model was designed with the purpose of experimenting with different machine learning methods, rather than to optimize correct classifications. While the model was trained to improve its performance, other models may be better suited in making predictions.