HANDWRITING RECOGNITION

-Project documentation-

Demir Korać

Due to Pythons simplicity and overall coverage with libraries a very simple script is needed for implementation.

The dataset being used in the project is the famous Modified National Institute of Standards and Technology database of handwritten digits, as it is the most reliable and largest database of this type.

The MNIST database of handwritten digits has a training set of 60,000 examples, and a test set of 10,000 examples. The digits have been size-normalized and centered in a fixed-size image.

The database and the files such as the training set images and labels, and the test set images and labels can be found at: <http://yann.lecun.com/exdb/mnist/>

For implementation of the main part of the project the scikit-learn python library has been used along with numpy, matplotlib and pandas

The documentation for the libraries used can be found here:

NumPy: <http://www.numpy.org/> (used for scientific computing with Python)

Matplotlib: <https://matplotlib.org/> (used for producing publication quality figures in Python)

Pandas: <https://pandas.pydata.org/> (used for providing high-performance, easy-to-use data structures and data analysis tools for Python, but here concretely used for loading the training data set)

Scikit-learn: <http://scikit-learn.org/stable/> (used for implementing machine learning in Python)

Program process of operation:

After importing the needed libraries, the user is asked to enter the order of the number from the database that they wish.

Then the dataset is loaded and separated into two parts, the training set and the test set.

The data is converted from row vectors into 28x28 matrices and analyzed and predicted, and finally outputed.

Optionally the predicitons can be compared to the dataset for an accuracy review.