SVM for digit recognition

Matteo Cavada

June 13, 2022

Contents

1 Declaration **1**

2 TODO Introduction and goals **1**

3 Dataset and pre-processing **1**

3.1 Dataset 1

3.2 Preprocessing 2

1 Declaration

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

2 TODO Introduction and goals

The goal of this project is to apply SVM techniques in order to obtain a reasonably good predictor for the task of *hand-written digit recognition*.

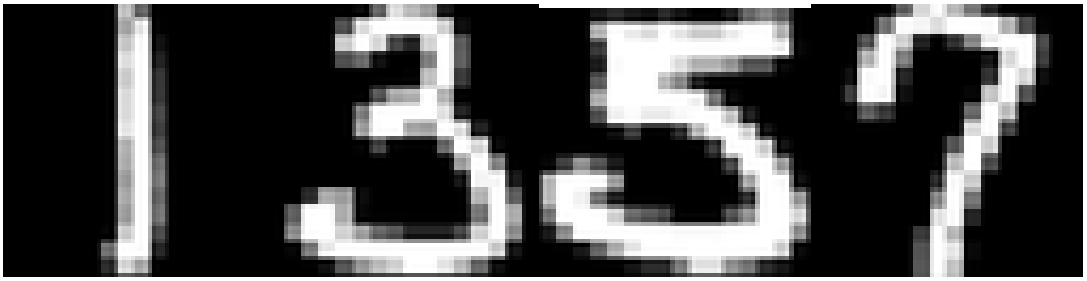
3 Dataset and pre-processing

3.1 Dataset

The dataset is available at kaggle.com.

The USPS dataset *"is a digit dataset automatically scanned from envelopes by the U.S. Postal Service"*.

The images are all in grayscale, with size 16x16. Each pixel is a value between 0 (black) and 1 (white).



3.2 Preprocessing

The data is stored in a single binary file using the format HDF5; the extraction of data from the file is achieved through the python library h5py.

The data is already normalized in the range $[0, 1]$, so no further action is required.