SS 2017 Pattern Recognition

Exercise 1a - 20 February 2017
First Classification Task with KNN

Deadline: 06.03.2017

The aim of this exercise is to implement your first classifier and apply it to the popular MNIST dataset,[1] which contains images of handwritten digits.

In this exercise, your task is to implement the KNN algorithm and test it on the dataset.

Data and Features

On Ilias you find two csv files (train.csv, test.csv). The csv-files contain the labels (drawn number) and pixel values for 28×28 grayscale images with pixel values in the range of [0-255]. The images are represented as one-dimensional arrays so you need to reshape the data if you want to display the images on the screen.

As features we will use the pixel values and will not compute any other features (although you certainly can if you like to).

KNN

Create your own implementation of the KNN algorithm (see lecture notes). Classify the entries of the dataset using KNN. Experiment with different values of K. Aim for a fast implementation as the dataset is of considerable size.

Implement at least two different distance metrics (e.g. Euclidean and Manhattan) and evaluate which one works best for you.

Expected Output

- Source Code of your implementation.
- Accuracy of your classification for K = {1, 3, 5, 10, 15} using the best distance metric.

---

[1] http://yann.lecun.com/exdb/mnist/