ASSIGNMENT 1 NEURAL NETWORKS: SUPERVISED LEARNING

Ismael Peruga, Álvaro González

January 2017

You can compile this file by using `pandoc main.md -o main.pdf`.

Data

To perform this assignment, 4 different sets of data are given. The first 3 consists of 2-dimensional features (the first and the second have 2 classes and the third have 3), the 4th is composed with OCR digits that have 64 features and 10 different classes (from 0 to 9).

$a^2 = \sqrt{b^2 + c^2}$

Set data 1

Set data 2

Set data 3

Set data 4

This set of data is OCR. It represents handwritten numbers. Each element of this set has 64 different features (which are the 8x8 matrix, where each element is between the 0 and 16 range). This set has been down-sampled from 32x32 to 8x8 by dividing the big matrix in blocks of 4x4. It allows to reduce the dimensions and give less variance to small distortions.

K-Nearest Neighbours (kNN)

In the first task, the kNN algorithm has to be implemented. This is a method of classification which consist in a comparison of the input data and all the training examples. The output of this algorithm is the estimated class of the input data.

To implement this, the euclidean distance is calculated between all training data and the input. All these distances are stored in a vector and the k smallest distances are taking in account to decide the output class. It is a vote system where the winner is the class with more neighbours between the k-nearest. In case of draw, the algorithm takes the smallest class (since the classes are represented as numbers). %we have to improve this

Backpropagation

The back propagation algorithm

Discussion and conclusion