CMTH 642
Homework 2
Deadline: Nov 20th, 2015

PROBLEM DEFINITION

Computers cannot beat humans on handwriting recognition yet. Handwriting recognition is an important research area in AI research with many applications. See these pages for further information about the problem.

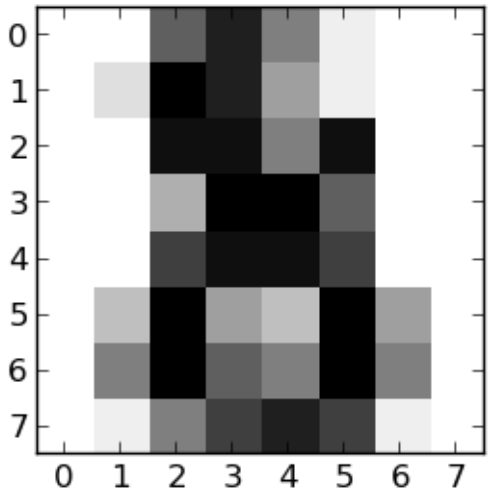
1. http://en.wikipedia.org/wiki/Handwriting_recognition
2. http://archive.ics.uci.edu/ml/datasets/Pen-Based+Recognition+of+Handwritten+Digits (original link to the dataset)
3. http://yann.lecun.com/exdb/mnist/ (this is a larger dataset. Check the performance of various algorithms.)

DATA

You are given a training and test dataset.

1. Pendigits.tra.csv
2. Pendigits.tes.csv

Each row contains 16 features (a 4x4 bitmap of a handwritten digit) and **its correct label** on the last column. Every feature is a numeric gray-scale pixel value. 255 is white and 0 is black. Here is an example of a 8x8 scanned handwritten digit:



You do not need to do any preprocessing to use this dataset.

TASK

1. Import the data to R (no cleanup necessary).
2. Use k nearest neighbor to predict the digit labels on the test dataset (you may use a library).
3. Optimize the value of k to get the best prediction performance.

OUTPUT

A zip file containing the following

- A report with:
    - Model
    - Results
- All the source code