

Introduction to Machine Learning. Classification

October 2019

The project should be uploaded in the ENSIIE Project website MALTP1 before october 31th 2019, 23H59. The projet zip file will contain the markdown and the pdf file. The length of the pdf report should not excess 5 pages.

Handwritten digit recognition

The aim of this project is to study the ability of several learning machines to build a handwritten digit classifier. In this application, the dataset contains handwritten digits written on envelopes to specify the postal code. Each digit is described by an image of $16 * 16$ pixels in normalized grayscale. The value for each pixel is a floating number between -1 and +1. The data are stored in the `ElemStatLearn` library of R and can be uploaded in the R environment using the following instructions:

```
library(ElemStatLearn);  
data(zip.train);  
data(zip.test);
```

Information on both datasets can be obtained using the instructions 'help(zip.train);' 'help(zip.test)'. The data in the 'zip.train' file are used to train the models and the data of the 'zip.test' file are used to evaluate the performances of the previous calibrated models.

-Check the number of observations in the zip.train and the zip.test files ($n_{train} = 7291$ and $n_{test} = 2007$).

-Plot the value of the first four examples of the zip.train file using the following instructions:

```
par(mfrow=c(1,4));  
image(zip2image(zip.train,1), col=gray(256:0/256), zlim=c(0,1), xlab="", ylab="",xaxt="n",yaxt="n");  
image(zip2image(zip.train,2), col=gray(256:0/256), zlim=c(0,1), xlab="", ylab="",xaxt="n",yaxt="n");  
image(zip2image(zip.train,3), col=gray(256:0/256), zlim=c(0,1), xlab="", ylab="",xaxt="n",yaxt="n");  
image(zip2image(zip.train,4), col=gray(256:0/256), zlim=c(0,1), xlab="", ylab="",xaxt="n",yaxt="n");
```

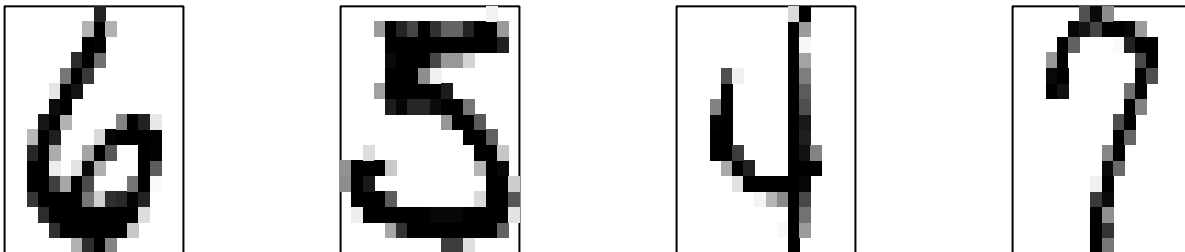


Figure 1: Illustration of the first digits in the zip.train dataset

-Note that the label of each observation is stored in the first position of each raw vector and that the remaining $16 * 16$ numbers of this raw vector provides the image (vector) of the digit.

Description of the work

- Chose two digits, for example '0' and '8' and don't forgett to tell in the beginning of your report your choice!
- Train, test and compare different binary machine learning classifiers to automatically recognize the two digits you chosed.
- Conclusion. Draw conclusions about the performances and the models used.