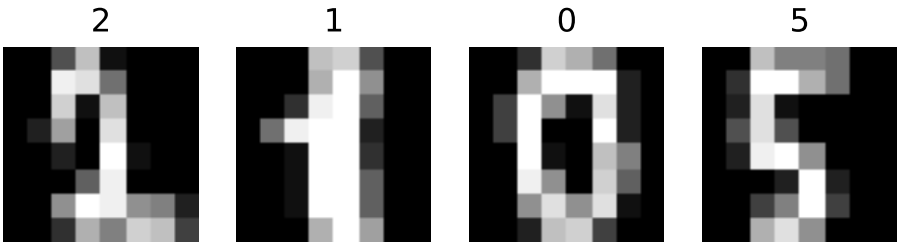
Machine learning for classification. Project

M. Mougeot

ENSIIE, October 2020

The aim of this project is to study the ability of several learning machines for building a handwritten digit classifier. The dataset contains a sample of handwritten digits. Each digit is described by an image of $8 * 8$ pixels in normalized grayscale.



The data are stored in the ScikitLearn library and can be uploaded in Python using the following instructions:

```
from sklearn import datasets
digits = datasets.load_digits()
```

Following instructions store the input data in the X matrix, the output labels in the y vector and visualize the 10th observation

```
import matplotlib.pyplot as plt
import numpy as np
X=digits.data;
y=digits.target;
obs=X[10,:]; label=y[10];
image=np.array(obs, dtype='uint8').reshape(8,8);
plt.imshow(B, cmap='gray')
```

Description of the work

- Choose two digits, for example '0' and '8' and don't forget 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 have chosen.
- Conclusion. Draw conclusions about the performances and the models used.

Before November 3th 2020, 09h00.

The project must be carried out by groups of two students. The names should be written in the two first lines of the jupyter notebook and the pdf report file ($dont'tforget$!) The project zip file should contain a Python jupiter notebook and a pdf report file. The length of the pdf report should not exceed 5 pages. $\rightarrow$ **The project should be uploaded in the ENSIIE Project website, in the repository MALTP1Proj before November 3th 2020, 09h00.**