Machine learning - 2020-2021
http://math.unice.fr/~rubentha/cours.html

Homework (deadline: 01/02/2021)

We want to try various algorithms on the handwriting recognition problem. Here we want our algorithm to recognize digits.

The original dataset is in a format that is difficult for beginners to use (it is made of pictures in grey scale). This dataset uses the work of Joseph Redmon to provide the MNIST dataset in a CSV format. The `mnist_train.csv`[1] file contains the 60,000 training examples and labels. The `mnist_test.csv`[2] contains 10,000 test examples and labels. Each row consists of 785 values: the first value is the label (a number from 0 to 9) and the remaining 784 values are the pixel values (a number from 0 to 255). You can plot the original pictures following the hints given on `https://www.askpython.com/python/examples/load-and-plot-mnist-dataset-in-python`

(1) Try a KNN classifier (trained on the train set). Give what is the error rate on the test set.
(2) Try a linear SVM classifier (trained on the train set). Give what is the error rate on the test set.
(3) Try a kernelized SVM classifier (trained on the train set), with the kernel you want. Give what is the error rate on the test set. Hopefully, this should be the smaller of the threes.

Write a report explaining your approach and including the code you produced (in python or something else). The report has to be in pdf and sent by e-mail on the deadline (or before). You can work with other students but you have to give me one report per student.

[1] https://math.unice.fr/~rubentha/enseignement/mnist_train.csv
[2] `https://math.unice.fr/~rubentha/enseignement/mnist_test.csv`