hugomcp@di.ubi.pt, 2023/24

Practical Project 2

Logistic Regression

1. Consider the "MNIST" dataset, (in "csv" format) available at "Kaggle.com" (https://www.kaggle.com/datasets/oddrationale/mnist-in-csv).

   The "mnist_train.csv" file contains the 60,000 training examples and labels.

   The "mnist_test.csv" 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).

Develop and implement a "logistic_regression.py" script that contains a model able to distinguish between the "0"..."9" classes in this dataset.

You should consider…

a) different feature normalization strategies:

   - Min-max
   - Z-score

b) different model regularization values.

c) different stopping criteria and learning rates for your model.