Assignment I:
Selected Topics in Distributed Multimedia Systems

Deadline: September 30, 2022 (rolling evaluation[1])

The *MNIST Database*[2] (Modified National Institute of Standards and Technology) is a widely used data set of handwritten numbers consisting of grayscale images with a size of 28×28 pixels. In total, there are 60,000 training images and 10,000 test images, which often used to evaluate machine learning methods.

The goal of this assignment is to compare different machine learning approaches for multi-class classification. To this end, an executable R script is given (*mnist_svm.r*), reading the data and solving the problem using SVM (default parameters), where the following task need to be addressed:

(a) Analyze different parameter choices for *SVM* in terms of classification results and runtime.

(b) Run a PCA-prepossessing step and analyze the performance in terms of classification results and runtime.

(c) Replace *SVM* with *Random Forests* **or** with a simple *Neural Network* and analyze different parameter choices in terms of classification results and runtime.

The results need to be summarized and interpreted in a short report (approx. 3-4 pages). In particular, highlight the differences, the advantages and disadvantages, and describe which approach you would use in practice.

Even though the initial script is given as *R*-Code, also *Python* might be used for the implementation. The used (executable) scripts must be attached to the main document. Due to the large amount of data, it would be meaningful to reduce the number of samples for training and evaluation to speed up the experiments.

[1] The projects will be evaluated when submitted.

[2] http://yann.lecun.com/exdb/mnist/