

Machine Learning — Programming Assignment

June 2024

1 Problem definition and data

A warehouse specialized in recycling old clothes needs a system to group items of the same kind automatically. A total of 70000 images have been taken with a low-resolution grayscale camera. Ten different kinds of clothes have been identified:



- 0. T-shirt;
- 1. Trouser;
- 2. Pullover;
- 3. Dress;
- 4. Coat;
- 5. Sandal;
- 6. Shirt;
- 7. Sneaker;
- 8. Bag;
- 9. Ankle boot;

Images have been already divided in training (60000) and test (10000) sets. All images have been resized to 28×28 pixels. They have been archived in the 'train.npz' and 'test.npz' numpy archives. To load the data you can do the following:

```
Xtrain , Ytrain = np.load("train.npz").values()  
Xtest , Ytest = np.load("test.npz").values()
```

2 Assignment

We want to build a classifier that is able to predict the kind of clothing given an image. For the programming assignment you are expected to:

1. analyze and comment the data;
2. implement, train and evaluate one or more classification models;
3. use suitable data processing and visualization techniques to analyze the behavior of the trained models.

All the above should be implemented as scripts in the Python programming language. Any machine learning library (included `pvm1`) can be used. Data and code used during the course can be used for the assignment if needed.

3 Report

Prepare a report of three to five pages documenting all your work. Provide detailed instructions on how to reproduce the results. The report must be in the PDF format. Include your name in the report and conclude the document with the following statement: “I affirm that this report is the result of my own work and that I did not share any part of it with anyone else except the teacher.”

Make a ZIP archive with the report and the Python scripts, and submit it from the course web page. To keep the size of the submission manageable, **do not include files containing the original data, the features etc.**