Exam Assignment

Machine Learning (BSc DS)
IT University of Copenhagen

Fall 2019

1 Introduction and formalities

This is the project description for the exam project in the Machine Learning course for the BSc program in Data Science at the IT University of Copenhagen. The exam must be submitted electronically via LearnIT no later than 14.00 on 13th December.

Groups You have to work in groups of 2–3 people. The group formation happens during the first week of the project, and one person from each group must send a list of group members for approval to Therese Graversen (`theg@itu.dk`) by Wednesday 13th November. The course manager reserves the right to modify the grouping, if necessary.

What should be handed in? You must hand in both a report and the source code you have developed to solve the project. The report must be in PDF format and have a front page that meets the ITU requirements.[1] Your code must be handed in as a single file (either a zip or tar archive). Only one person for each group should hand in the project (report and code).

Length The report must contain no more than 2,500 words and be no longer than a total of 12 pages including figures, tables, code snippets, references and appendixes, but excluding the front page. The project must be typeset with at least 11pt font size and margins (both horizontal and vertical) of at least 2cm. The number of words and the methods used for counting must be stated at the beginning of the report.

2 Problem and data set

In this project we will explore different methods for determining the type of clothing from an image of the item. Zalando has collected a dataset of 70,000 28x28 grayscale images of clothing and labelled them according to ten different categories (Xiao et al., 2017). For this project, we use a subset of 15,000 images that may be either that of a t-shirt/top, trousers, a pullover, a dress, or a shirt (see Figure 1).

The images are divided into a training set of 10,000 images and a test set of 5,000 images. The images and associated labels are available in NPY format as:

`fashion_train.npy` and `fashion_test.npy`.

Each line describes a piece of clothing. The first 784 columns are the pixel values of the 28x28 grayscale image, each taking an integer value between 0 and 255. It is recommendable that you normalise these values for further analysis. The last column, number 785, is the category of clothing and takes values in $\{0, 1, 2, 3, 4\}$ (see Table 1).

[1] Found at `https://itustudent.itu.dk/study-administration/exams/submitting-written-work`

Table 1: Categories of clothes

Type of clothing	T-shirt/top	Trouser	Pullover	Dress	Shirt
Label	0	1	2	3	4

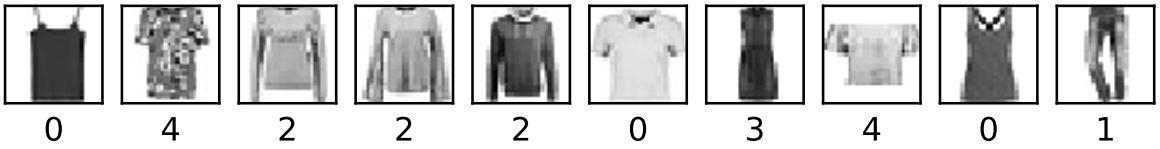

Figure 1: A random sample of 10 images from the training dataset.

3 Methods and report

You have to implement and compare a number of different methods for this classification problem. In particular, you have to implement the following methods:

1. A linear discriminant analysis (LDA).
2. A multilayer perceptron with the size and other hyperparameters of your own choice.
3. One (or more) classification method(s) of your own choice.

The models should be trained and validated on the training set. You are (of course) not allowed to use the test set for training or validating the models.

The report should describe the choices you have made, your solutions and your results. In particular, you have to address the following:

- You must visualise the data set using a suitable dimensionality reduction based, for instance, on PCA or LDA projections.
- You are to document and describe the methods and models, you have implemented, and how you have implemented them (e.g. with relevant mathematical expressions, pseudocode or code snippets).
- You must describe and discuss relevant aspects of model selection for each of the models you have implemented.
- For each of the models, you must report on their performance on both the training and the test set. You must compare and discuss these results. You should use at least one of the two performance metrics *average accuracy* and *macro-F1 score* defined in Appendix A.
- You have to document which considerations and precautions you have made to ensure that your models have not been fitted to the test data.

Make sure to use correct references to works of other people in your report, including references to the course book, Alpaydin (2014).

4 Implementation and code

Your implementation has to be in Python. You may make use of any standard Python libraries and the numerical libraries NumPy and SciPy. *The only machine learning library you may make use of is TensorFlow* (without using Keras, or any other similar high-level API).

Your code should be organised such that is easy to read, i.e. you have to use descriptive names for files, functions, variables, etc. The code may be organised in regular Python source files (.py files) or in Jupyter notebooks.

If you take inspiration from or copy code developed by other people, it is important that you document this in your report.

A Average accuracy and macro-F1

The average accuracy for multiclass classification is the generalized form of binary accuracy:

$$\text{AAC} = \frac{1}{C}\sum_{i=1}^{C}\frac{\text{TP}_i + \text{TN}_i}{\text{TP}_i + \text{TN}_i + \text{FP}_i + \text{FN}_i}$$

where C is the total number of classes.

The F1-score is a function of recall and precision:

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

and macro-F1 is the average of the per-class F1-score:

$$\text{Macro-F1} = \frac{1}{C}\sum_{i=1}^{C}\text{F1}_i$$

References

Alpaydin, E. (2014). *Introduction to Machine Learning*. The MIT Press.

Xiao, H., Rasul, K., and Vollgraf, R. (2017). Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms.