Unsupervised Machine Learning Projects Submission Guide

General description

The projects consist of three exercises. Each exercise requires to implement unsupervised machine-learning algorithms in Matlab, Octave, R or Python, test them using surrogate or real data, discuss the results obtained and their relation to the underlying theory, and write a report. These skills are fundamental for unveiling the benefits and limitations of existing algorithms, as well as possible customizations for real applications. The course also teaches scientific reporting skills, fundamental for communicating the results obtained with existing methods and the rationale behind improvements and breakthroughs.

Instructions for solving the exercises

1. Each exercise can be solved by students working individually or in pairs (not more than two students). When working in pairs, only one report and group of source codes per exercise per pair is enough.

2. The submission must include a report in pdf, source codes with the implementations of the machine-learning algorithms, and a file per exercise with all the calculations necessary for reproducing the computations and figures.

3. Results mostly consist of a figure or a short program output. In both cases, add a short paragraph explaining what is being shown and why the results are the answer to the questions.

4. All questions must be answered.

5. Whenever the exercises ask you to discuss a result, the result must be discussed. If you have no idea what to discuss, please state that in the report or ask me for some hints before submitting.

6. Whenever the exercises ask you to test whether an algorithm works, always give an example which shows that the algorithm indeed works (or not).

7. Source codes must include comments reflecting the connection between the code lines and the math underlying the implemented method.

Instructions for writing the report

1. Reports must be written in English. Aim to write in either American or British spelling, but avoid using both in the same report, such an inconsistence is generally considered bad practice.

2. For each exercise, students are required a separate report as a single pdf containing both results and figures. For each project, there must be as many reports as exercises.

3. The font size must be 12, preferably Arial, Helvetica or Times New Roman. Font size in figures must be at least 10.

4. A report must contain at most as many pages with text as items in each exercise. For example, in project 1 exercise 1, the number of pages with text must not exceed 6.

5. Reports must include a title page with the following information:

    a) Project title

    b) Exercise number

    c) Name, student ID number, affiliation (for example, Department of Computer Science) and academic level (for example, Bachelor student, Master student or PhD student), of each student involved in the project.

6. Figures must be numbered, include a short caption summarizing the results shown, positioned immediately after they are first mentioned in the text, and their axes must be labelled. To speed up the writing process, however, labels can be instead denoted in the caption of the figure.

7. All abbreviations and acronyms must be defined the first time they are used in the text.

Instructions for writing file with calculations

1. There must be a file per exercise with all the code involved in the computations of the results and the figures.

2. Each piece of code corresponding to each question in each item must be separated by a comment indicating what that code is answering.

3. FIX the seeds of the random numbers at the beginning of the file so that results in your report are reproducible.

4. The provided code must be bug free, ready to run on a windows or linux machine.

5. Indicate what platform, packages, language and versions have you employed.

Completion of the project

1. The project will be completed once all exercises are submitted to the following email address: hugo.eyherabide@helsinki.fi.

2. All exercises must be submitted before the deadline indicated above. Failure to do so will yield a reduction of 10 points per day.

3. Each exercise must be submitted as a single zip file containing both the pdf file with the report and the source codes of the implementation.

4. Exercises can be submitted individually after completion.

5. Each exercise in the project is assigned 30 points. They will, at the end, determine your grade for the computer project.

6. Writing and coding style is assigned 10 points. Reports must be clear and enjoyable to read.