Unsupervised Machine Learning Course Work

Option B

Unsupervised and Reinforcement Learning (URL)

Master in Artificial Intelligence (MAI)

2020 Spring Semester

Chapter 1

The course work

1.1 The assignment

A task that is usual in research is to compare the methods/algorithms that we propose to other similar/competing methods in order to show empirically the benefits of our approach.

This means to look into the state of the art for the more recent methods in our area and, if there is not an implementation available, to implement the competing algorithms to be able to test and compare them with ours.

The goal of the course is not to develop new algorithms, but we can pick some papers describing current unsupervised machine learning methods, develop an implementation and then test its properties against some well accepted methods. Basically the task is to reproduce and test the results of a paper.

The main tasks of the assignment are:

- Read and understand the references that describe the algorithms that you have chosen

- Implement the algorithms in python 3 following the API conventions used by the library `scikit-learn`

- Write a report:

  – Summarizing the characteristics of the algorithms

  – Comparing experimentally the algorithms to similar algorithms using artificial and real datasets

In order to obtain a fair comparison, you should implement the algorithms **efficiently**. This means to use extensively the numpy/scipy libraries and their array operations. You can use also the optimized datastructures and functions implemented in these libraries like for instance spatial indexes for nearest neighbor search or distance matrix computations.

The deadline for delivering the report is **June 1st 2020**. The paper has to be in PDF format and has to be uploaded to the Raco using the assignment delivery application for the course.

1.2 The algorithms

You have available a list of algorithms in:

http://www.cs.upc.edu/~bejar/URL/courseworkB.html

Each option has one paper describing a new clustering algorithm or a variation of an existing one. You <u>can pick any</u> of the options, but a **maximum of two students** is allowed for option, so you have to send and e-mail (`bejar@cs.upc.edu`) declaring the one that you prefer as soon as possible.

If you want to propose another algorithms related to the course feel free to contact me and we will talk about it.

1.3 Evaluation

The evaluation of the work will consider these elements:

- Quality of the description of the implemented algorithms
- Quality of the experiments performed:
    - Types and relevance of the datasets used
    - Quality of the analysis of the experiments (what is tested, how is tested)
    - Quality of the explanations of the results and their conclusions
- Quality of the implementation, mainly the coverage of the characteristics of the reference algorithms and the efficiency of the implementation