SPLEX Project

Life Sciences Data Exploration
Propose and Develop an Optimal Method

General information:

- The project is to be done by a team of 2 students (binôme). You are allowed to work alone but we will not accept teams with 3 or more students.

- Goals: 1) to conduct independent data exploration, and some research using machine learning methods, 2) to propose an optimal approach to process the data, and 3) to analyse and clearly explain the obtained results.

- Duration: until the end of the semester

- Presentation of your results: you will be given 10 minutes to present your results during the last TME (prepare some slides)

Project Schedule:

1. Week 1:

    - Download the data Choose and download a data set. You are free to choose any benchmark, but you should be unique in the group (no repeating data sets), e.g., you can look here : `https://archive.ics.uci.edu/ml/datasets.php?format=&task=cla&att=num&area=life&numAtt=10to100&numIns=&type=mvar&sort=nameUp&view=table`
      or here : `https://www.kaggle.com/datasets`
    - Do some data pre-processing if necessary
    - Test some standard machine learning methods relevant for your problem
    - Read the papers related to the scientific problem (some papers will be provided but feel free to read more)

2. Week 2:

    - Analyse the results obtained with the state-of-the-art methods
    - Propose your own approach (it can be a combination of the existing methods, or a particular pre-processing or post-processing step, or maybe you have found a state-of-the-art method which is optimal)
    - Implement it in Python. You are not asked to re-implement existing algorithms, just the part you propose.
    - Explain why the chosen method is the best in your case (it can be the most accurate, or the simplest, or the fastest, or ...)

3. Week 3: Oral presentation. You are about 10 minutes to clearly present your results. I will ask you to send me your presentation (as a .pdf file). It is not needed to write a report. Your slides should include:

    - Problem description
    - Methods you tested, why you choose them
    - The approach you consider to be an optimal solution
    - Implementation details, and difficulties you met
    - Conclusions and perspectives

Note: you are free to propose your own plan of work, however, you are supposed to present the obtained results during the last practical session.

If you are particularly interested in a specific application, let me know. We may accept that you work on a data set which is not listed, but we need to discuss it before.

Project Evaluation

- You should demonstrate that you well understood the methods you used and the results you obtained

- It is possible that the optimal approach is an existing method. In this case, you have to prove it by comparing the performance of several machine learning methods

- Clarity of the presentation