Unsupervised Learning

Javier Béjar

URL - 2024 Spring Term

CS - MIA

- Javier Béjar
- **e-mail:** bejar@cs.upc.edu
- Omega Building - Office 204

- Web for the course: https://sites.google.com/upc.edu/mai-url/ul-mai

- To introduce to main algorithms in Unsupervised Learning
  - Preprocessing and data cleaning, Dimensionality reduction, Clustering
- To give an overview of the main methods for scaling Clustering algorithms for Big Data
- To introduce to advanced topics on unsupervised learning: Ensembles/Consensus
- To introduce to Unsupervised Deep Learning
  - Generative models
  - Representation learning

Unsupervised Learning

1. Knowledge Discovery in Databases
2. Unsupervised Data Preprocessing
3. Clustering Algorithms
4. Clustering Validation/Model Selection
5. Scalable Unsupervised Learning in Data Mining
6. Consensus Clustering

Unsupervised Deep Learning

7. Deep Autoregressive Models
8. Flow models
9. Deep Latent Variables Models (VAEs)
10. Deep Implicit models (GANs)
11. Denoising Diffusion Models
12. Self-Supervised and Contrastive Learning

The evaluation for this part of the course will consist of:

- A test exam about the topics of the course (20%) **(June 14th)**
- Evaluation of Clustering Topics (40%) **(May 6th)**
  - Implementation of unsupervised learning algorithms (in python)
- Evaluation of Unsupervised Deep learning Topics (40%) **(June 10th)**
  - Record a video presentation about one current research paper

- You will have to pick one of the options of unsupervised learning methods from the list proposed (2 student maximum per algorithm)
- The implementation must be in the python 3 language, following the conventions of the scikit-learn library
- You will have to write a report about the algorithms comparing them with other similar algorithms
- The list of possible algorithms to implement is in: https://sites.google.com/upc.edu/mai-url/ul-mai/coursework1
- The deadline for delivering this report, and the code is May 6th 2024

- You will have to pick one of the selected papers about unsupervised deep learning (2 student maximum per paper)
- Record a short video presentation (between 8-12 minutes) explaining the content of the paper (follow the evaluation rubric)
- We will use peer evaluation: Two presentations will be assigned to each student that will grade them according to the rubric criteria
- I will evaluate the presentations too, the grade will be a combination of both
- The list of possible papers is in: https://sites.google.com/upc.edu/mai-url/ul-mai/coursework2
- The deadline for delivering the video is June 10th 2024, peer grading due June 17th 2024

- Course lecture notes, slides and papers
  https://sites.google.com/upc.edu/mai-url/ul-mai/materials
- Other papers not included on the web page can be found through Google Scholar
  http://scholar.google.es/
- Software and datasets used during the semester
  https://sites.google.com/upc.edu/mai-url/ul-mai/software

- We will use Python notebooks to illustrate some topics and play with the algorithms
- For using yourself the notebooks, one option is to install Python and all the packages used in the course (see the software webpage)
- Other options are:
  - Google Colab
  - Install Docker and the data-science notebook image