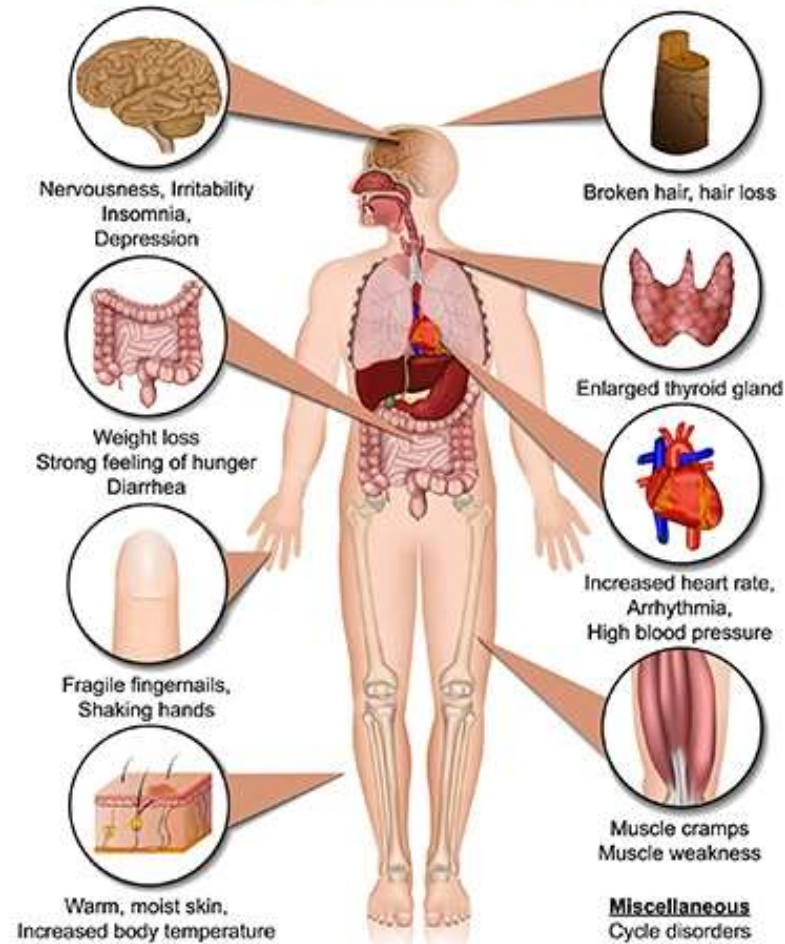


UNSUPERVISED LEARNING

THE HYPOTHYROIDISM DATASET

A.Y. 2023-2024

HYPOTHYROIDISM



Dataset

- 21 attributes
 - 15 binary
 - 6 continuous
- 7,200 data objects

Tasks

- Develop a software code for preprocessing the dataset, i.e., for computing the distance between different data objects. Notice that you have to cope with different types of attributes!!!
- Develop software code for anomaly detection by at least two different algorithms on the given dataset

UNSUPERVISED LEARNING — A.Y. 2023-2024

What must be delivered

- All software codes associated with the Tasks, use the standard practice you learnt during the lab lectures to describe and comment the software code.
- A technical report consisting of no more than 6 pages (4 additional for figures, tables, ...) where you describe how you analyzed the given dataset. Concrete examples of a technical report will be made available on the Moodle platform of the course.
- A dedicate link will be made available on Moodle (upload 3 days before the official data of the exam)
- You need also to deliver the data file with the last column containing the probbaility of the data object being anomalous (use the prediciton for the model you think is the best)

What is allowed

- You can develop the project yourself or you can develop the project with one of your colleagues
- You can access any source of high quality to develop your project, other courses, papers, books, ...

What is not allowed

- No plagiarism in any form, included chatGPT and similar (you must add a statement to your report that clearly exclude the use of any form of plagiarism)

UNSUPERVISED LEARNING — A.Y. 2023-2024

ASSESSMENT

- **Project** 20pts
- **Interview** 13pts

UNSUPERVISED LEARNING — A.Y. 2023-2024

Project Grading

- **Software code** (up to 10 points)
 - Correctness (up to 4 points)
 - Efficiency (up to 3 points)
 - Clarity of comments (up to 3 points)

- **Technical report** (up to 10 points)
 - Correctness, rigour, accuracy of the methodological approach (up to 4 points)
 - Clarity of expression and communication of ideas (up to 3 points)
 - Tables, figures, captions, are they appropriate? Well motivated? Explained? (up to 3 points)

UNSUPERVISED LEARNING — A.Y. 2023-2024

Interview

— about 20 minutes

- 10 minutes for the oral presentation of the project, slides to be used, both members must present
- 10 minutes of interview on methodological aspects of the project and of what presented in the course

Rejecting the Grade

- A new version of the project, substantially different from the previous one, must be prepared and submitted
- Lab Reports expire when the new course starts, 2025 March 1st
- New interview