Project 1: Classification

ELEN0062-1 | Introduction to Machine Learning

Academic Year 2025-2026

Goals of the project

- Put some classical classification ML algorithms into practice.
- Understand and illustrate some concepts, such as over and underfitting.
- Get accustomed with Python programming.

What we give you

- Statement
- Partial code
- Two binary datasets: one synthetic, one real

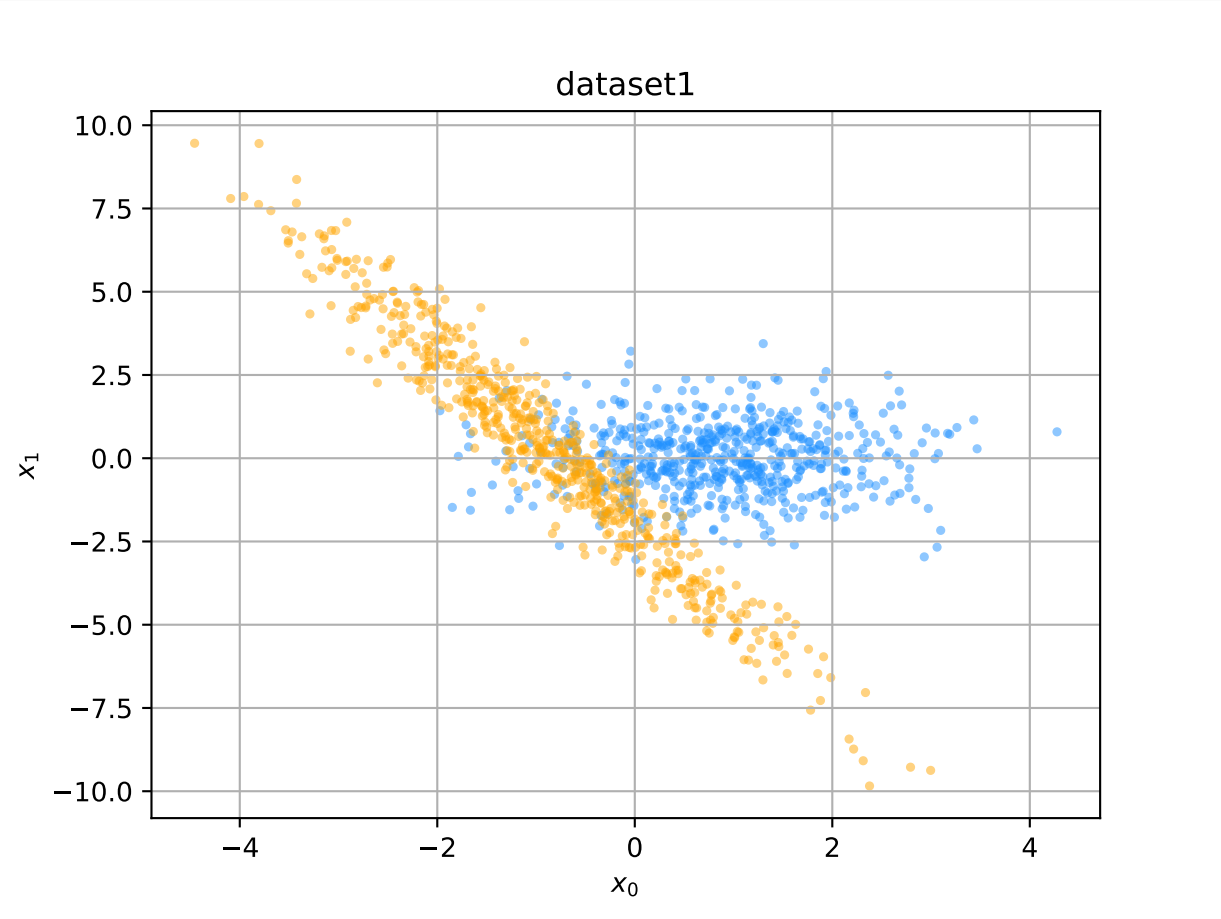
For all projects, resources are published at
https://iml.isach.be

There will also be published statement corrections or updates.

Datasets (1/2)

The first dataset is synthetic and consists of two Gaussian distributions.

We ask you to generate 1,200 samples, 900 of which for training.



Datasets (2/2)

The second dataset is real and contains breast cancer detections based on 30 physiological features obtained by a medical examination.

It contains 569 samples, 426 of which for training. 357 samples are benign, and 212 are malignant.

The dataset contains 30 continuous features: mass radius, texture, perimeter, ...

Questions

1. Study the performance of decision trees on the synthetic dataset.
2. Study the performance of kNN on the synthetic dataset.
3. Implement yourself another method (QDA) and study its performance on the two datasets.
4. Tune and compare the different methods.

Code

- *data.py*: Generates datasets.
- *plot.py*: Plots boundaries.
- *dt.py*: Decision trees experiments.
- *knn.py*: kNN experiments.
- *qda.py*: QDA/LDA experiments.
- *mc.py*: Method comparison.

The first two files are already completed.

Resources to learn about Python programming and setting up your environment for the project are available at *https://iml.isach.be*.

Report

In addition to the code to be filled, we ask you to write a report of up to 4 pages, excluding figures.

Please use LaTeX to write your report. A template and helpful resources are available at *https://iml.isach.be*.

Submission

Deadline: 24th October 2025 at 23:59.

Make groups of 2 or 3 on Gradescope (course code in the statement).

You must submit 2 parts. Each one has a dedicated Gradescope assignment.

- A 4-page report (*`report.pdf`*).
- The code for the 4 uncompleted files.

Please pay attention to the presentation of your results! Read the statement carefully.

Need help?

Please read the resources on the website thoroughly first. Please prioritize using eCampus before contacting us by mail.

Otherwise, feel free to contact us if you need help!

Teaching assistants:

- Sacha Lewin (*sacha.lewin@uliege.be*).
- Yann Claes (*y.claes@uliege.be*).

Questions?