Project 3: Ridge Regression

Ridge regression refers to the use of linear regression with regularization. Most specifically, using the L2 norm as a regularizer.

Objectives
By executing this project, you will be able to
1. Gain deep knowledge about the implementation of ridge regression
2. Improve your understanding of the role of regulatization
3. Strengthen the concept of model selection
4. Continue gaining practical expertise on the use of python, Jupyter, libraries used in machine learning and programming, in general.

Task 1. You will implement ridge regression from scratch in a file that you will name `ridge.py`. You may use the implementation of linear regression showcased in the course as a starting point. You cannot use scikit-learn. It is expected that you code the algorithm from scratch.

Make sure you debug your code, to verify that it works accordingly. You may consider comparing it against scikit-learn implementation for validation.

Task 2. Check your implementation using the Olympics 100 m dataset that has been showcased in several lectures. Are your results similar? How do you choose λ? Define a strategy for this.

Report:
You need to prepare a 1-page report explaining in detail the strategy you used to implement ridge regression. You should illustrate the performance of your implementation in the 100 m Olympics dataset and comment on the choice of λ. Optionally, you can compare your implementation to that of scikit-learn.

Deliverables:
Upload a zip file named <group-name>.zip containing the report, the ridge.py file, any notebooks you produced and any instructions required to run it.

Evaluation:

Criteria	Score
Your code runs and works as expected on the data	8
Your code runs on any dataset	2
Your report	10
Scikit-learn comparison	1

Important:
- Failing to submit a report leads to a mark of zero (0).
- If ChatGPT is used, failing to report it and explaining its use leads to a mark of zero (0).
- A group will be chosen at random to present their solution during the lecture. Failing to justify the submitted solution leads to a mark of zero (0).