MALIS Project 1:
Leave One Out Cross-Validation and KNNs

In this project, you will be studying the problem of validation to identify the best hyperparameters in a given model. In particular, we will focus on K-nearest neighbors for regression.

Objectives

By executing this project, you will be able to:

1. Learn to read, understand, and implement a research paper.
2. Gain practical understanding of the concept of model validation.
3. Strengthen knowledge on the KNNs algorithm and its strengths and weaknesses.
4. Improve proficiency in the usage of machine learning libraries and in programming.

Part I – Leave One Out Cross-Validation

Task 1. Investigate the concept of cross-validation and the special case of leave-one-out cross-validation (LOOCV). What are these used for? Why may one choose to use LOOCV? In which scenarios?

Part II – Implementing Fast LOOCV for KNNs

Task 2. Read the paper on fast LOOCV for K-nearest neighbors regressors by Motonobu Kanagawa (link).

Task 3. You will need to implement his proposed method to accelerate LOOCV in the context of regression using KNNs. The fast LOOCV strategy should be implemented within a Python file, following the skeleton provided (function do_fast_loocv). Document your code.

Task 4. Implement as well, within the same Python file, a standard LOOCV (function do_normal_loocv).

Part II – Evaluation

Task 5. You will now use your code to run experiments. For this matter, create a Jupyter notebook denoted experiment.ipynb.

You should implement the necessary routines to reproduce Figures 2 and 3 in the paper using the code developed in Part II. Failing to use the Python file developed in Part II (tasks 3 and 4) will result in a penalty. Document all the steps clearly. Any additional experiments you perform should also be documented. Report all your results.

Report:

You need to prepare a 500-750-word (max) report containing the following:
- A summary of your findings on what cross-validation is and, in particular, leave-one-out cross-validation

- A paragraph explaining why cross-validation can be problematic in KNNs
- A study comparing the results of standard cross-validation and the results you obtained (i.e., present the results of task 5).

Contributions: Please include a section that describes the work each team member contributed to the project. This is to ensure that team members are carrying a fair share of the work on projects. This section does not count towards the total word count.

Deliverables:

Upload a zip file containing the report, all code files, and any instructions required to run your code.

Important:
- Failing to submit a report leads to a mark of zero (0).
- Failing to run your code 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).