

# Assignment 2

Fundamentos del Aprendizaje Automático  
Curso 2025-2026

This second practical assignment of the course aims at expanding and clarifying the concepts introduced in the modules studied so far during the course (mainly, **T2**, **T3**, and **T4**).

You will follow this script to develop the required programs and experiments to eventually deliver a report summarizing the main insights and conclusions obtained.

## Objectives

- Assimilate the evaluation methodology of machine learning methods.
- Implement practical learning techniques and schemes.
- Critically compare different learning strategies under different scenarios.
- Explain and justify the reasons for the decisions taken.
- Learn to report results and insights in a scientific manner.

## Tasks

- T1. Dataset:** Select a *classification* dataset from a data repository and load it into your program.
- T2. Prepare the assortment:** Create the necessary data partitions and arrange a cross-validation scheme.
- T3. Classifiers:** You will compare the performance of Bayesian classifiers (MLE for multivariate, Naïve Bayes), nonparametric density estimators (histogram, Parzen, and  $k_n$ -Nearest Neighbor),  $k$ -Nearest Neighbor rule.
- T4. Hyperparameters:** Optimize the target metric by adjusting the different hyperparameters of the models.
- T5. Results:** Report and discuss the results obtained.
- T6. Analyze:** Derive insights and conclusions about the results obtained.

## Considerations

- The *dataset* must have been devised for *multiclass* classification tasks.
- It should not contain *more than one feature* and *without missing* values (for simplicity).
- A set of questions will be provided in the next sessions that *must* be addressed in the report.

## Report format

The report must follow these points:

- Must use the ICML template.
- Clearly state the assignment, your name, and identification details (ID and email).
- The report must be *3 pages maximum* (double column).

**Delivery:**

Submit a single ZIP file via Moodle. It must include the report and it may also contain the developed code.

**Delivery date: November 25, 2025**

## Tips and suggestions

- You may obtain a dataset from the UCI repository: <https://archive.ics.uci.edu/datasets>.
- Code quality is not a requirement, but the conclusions obtained.
- A Jupyter Notebook (or similar) may be adequate for the task at hand.