

Introduction in Machine Learning - WBAI056-2023

Assignment 1

Total points: **40**
Starting date: 7 September 2023
Submission deadline: **23:59, 24 September 2023**

Lecturers Dr. Matias Valdenegro, Dr. Andreea Sburlea, Dr. Marco Zullich,
Dr. Tsegaye Tashu, Juan Cardenas.

General guidelines:

- The tasks are targeted at groups of three students. Please make sure that the load is well divided: every student should contribute.
- Please take advantage of the practical sessions to ask your questions about the tasks.
- Provide a (short but comprehensive) explanation of what you are doing for each task.
- A reviewer should be able to understand plots independently; be sure to label axes, a legend for colors, use an easily readable font size, etc.
- Refer to all plots, tables, code blocks, etc. in your report.
- For the report: you can use a jupyter notebook or write a PDF in a word processor of your preference. Please include code as `.py` or `.ipynb` files as attachments in Brightspace.

Part I - Feature Types

Maximum obtainable points: **10**

Explain for each of the following if they are continuous, discrete, or categorical features:

- A random number in range $[0, 1]$.
- Number of bikes stores in Groningen.
- Handedness.
- Current types of Machine Learning methods.
- Duration of your trip from home to campus Zernike.

Part II - Classifiers vs Regressors

Maximum obtainable points: **10**

Consider the following features:

1. Number of coffee units drank per day.
2. Number of study hours.
3. Number of sleeping hours.
4. Country where a student obtained their high school diploma.

5. Age.
6. Number of times student attended lectures.

Answer these two questions, considering the problem of (individual)student success rate.

1. Which of these features is/are useful to predict the success rate of students passing a course? Please explain your choice.
2. Is predicting success rate of students passing a course a classification or a regression task? Please explain your reasoning.

Part III - Feature Transformations and Linear Separability

Maximum obtainable points: **20**

Use the wine recognition dataset available in scikit-learn (https://scikit-learn.org/stable/datasets/toy_dataset.html#wine-recognition-dataset) and perform the following:

- Load the dataset in python or jupyter.
- Choose four to eight features (up to you) that you think are best to predict wine quality.
- Train a simple linear model (logistic regression) and decide if your results show if the features you selected are linearly separable or not. Compare your results with training the same model on the full set of features.
- Try a single simple transformation (like different mathematical functions for each feature) of your selected features and evaluate if your new transformed features are linearly separable.

Discuss and analyze your results for each bullet point.