

## Introduction in Machine Learning - WBAI056-2024.2025

### Assignment 2

Total points: **40**  
Starting date: 29 November 2024  
Submission deadline: **23:59, 13 December 2024**

**Lecturers** Dr. Matias Valdenegro, Dr. Andreea Sburlea.

**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.
- In your report, please include a statement from all group members, indicating what was their contribution to this assignment.

## Part I - Overfitting Curves

Maximum obtainable points: **10**

In the lectures we have seen many examples of train/validation loss curves, in this part you should draw train/validation loss curves denoting specific characteristics related to overfitting, underfitting, and neither of them, as follows:

1. Two curves showing overfitting.
2. Two curves showing underfitting.
3. One curve showing neither overfitting, neither underfitting.

Example curves can be given from real machine learning training runs or drawn by you, but should not be examples from the lectures. And just a reminder, do not use GenAI to generate these examples. The train/val loss curves should be made by your group.

For each of your examples, include a short statement on why that example shows the relevant condition (underfit, overfit, neither).

## Part II - To Split, or Not to Split

Maximum obtainable points: **10**

Given knowledge from the lectures, answer the following question:

What are the consequences of **not** having a dataset split?

For the above question, consider that a validation set is not available (50% of the points), and a test set is not available (50% of the points), each independently.

## Part III - Multi-Task Regularization.

Maximum obtainable points: **20**

Select a machine learning task (some examples can be object detection, image classification, segmentation, reinforcement learning, predicting the next word, sequence-to-sequence prediction, etc), and propose one additional task (also called auxiliary task) within a multi-task learning setup that would regularize the task and overall improve performance. For this you can do one of the following:

1. Propose the additional task yourself.
2. Find an additional task defined in the literature for this purpose, and remember to cite any sources you use.
3. Derive your additional task based on the lecture examples.

In any of the selected cases above, you should describe the additional task (50% points), and answer the question: **how does your additional task regularize the original task?** (50 % points).

To answer the question, you can make theoretical arguments, or build intuitions, but there is no need to do experiments or make a convoluted arguments.