

## AML course rules

### Prerequisites

This course blends methodological and computer science skills. Students are expected to be comfortable with Python programming, and with common libraries used in the context of data science and machine learning problems. Moreover, students are assumed to be comfortable with machine learning methodologies.

The skills above are acquired mostly in the MALIS and the Deep Learning courses. In such courses, students gain familiarity both with Python, Jupyter Notebooks, machine learning libraries such as sk-learn and TensorFlow, on the computer science side. Additionally, students are exposed to most of the important machine learning concepts, methods and theory. Optionally, the ASI course can give a special twist to address problems in a probabilistic manner, but it is not required.

If you are enrolled to the AML course, but **didn't follow MALIS, it could be very problematic**. If you **didn't follow the Deep Learning course, your modeling approaches could be limited**.

Another underlying prerequisite to participate to AML is **familiarity with the Kaggle platform**. This platform uses "kernels", which are the equivalent of Jupyter notebooks, hosted on the Kaggle platform, which is backed by Google Cloud Engine.

In case you do not fulfill the prerequisites, you are invited to contact Prof. Pietro Michiardi.

### Expected workload

A good indication of the effort you should put in each of the challenges is in the order of **10-15 hours per activity maximum**, spread throughout the duration allocated for each of them.

If you feel like you need much more time to complete an activity, there may be many reasons:

- You may be in a position in which you aim for quantity rather than quality. Please, refer to the grading rules below to better dose the effort you put in each activity.
- Your group requires more balanced skills. Please, refer to the section "working in groups" with some suggestions about how to form groups.
- You do not have the prerequisites necessary for the course. Please, contact Prof. Pietro Michiardi if this is your case.

---

### *What is a challenge in this course*

The concept of a "data science" or "machine learning" challenge or competition is nowadays widespread. Typically, as done in Kaggle, a competition involves a team cooperate toward a submission, which takes the form of a set of predictions for a test set for which the ground-truth is undisclosed to participants. Then, an automatic system computes a ranking based on a given performance metric, which is used to compile a leaderboard, together with the attribution of "honor" badges and even monetary prizes.

The goal of the AML course is to make sure students develop good methodologies for data science problems, in particular those that also involve machine learning. As a consequence, we will both adopt the Kaggle philosophy, and rank teams based on performance score, as well as we will expect more "academic" kind of submissions, which we detail below.

A **successful team** is not one that submits a notebook with the highest performance, nor the one that submits the longest notebook, with the largest number of plots, code, models, or whatever else. Grading criteria are given below.

---

### *Working in Groups (a.k.a. teams)*

The activities to be completed within the AML course, are conceived to be carried out in groups. This has several benefits, including:

- Spreading the workload across students in a group.
- Learning how to work in teams for a data science project.
- Reduce the number of submissions, that the TAs will need to evaluate.

As a consequence students must form groups/teams **using Moodle**, with the following rules:

- Minimum 2 students per group
- Maximum 4 students per group

A good advice is to mix and match skills: someone with a good "coding experience" but a thin background on machine learning, could match with someone with less "coding experience" but with a substantial background on machine learning and statistics.

Since we will use the official Kaggle platform to host the challenges, you will need to **create a team on Kaggle** with the same composition of your Moodle group, and the same team name as your group on Moodle. So if you are in "Group 18" on Moodle, then your Kaggle team should be called "Group 18". This way we will be able to associate your Kaggle notebook (a.k.a. kernel) to your group and your submission on Moodle.

Note that for you to access the Kaggle platform, an identity must be created, e.g. you could use your Google credentials to log-in.