



UNIVERSITÀ  
DEGLI STUDI DI MILANO-BICOCCA

## SYLLABUS DEL CORSO

### Machine Learning M

2122-1-F8204B006

---

#### Aims

The student will learn the most effective Machine Learning techniques, understanding the theoretical foundations of each technique and acquiring the know-how to successfully apply them to solving practical problems. An overview of the most innovative solutions for the identification of the best Machine Learning algorithm and its optimal configuration, given a dataset (Automated Machine Learning - AutoML), will also be provided. The reference tool for the course will be R, but some equivalent solutions will also be presented in Python (for example scikit-learn) and Java (for example WEKA, KNIME).

#### Contents

Machine Learning basics: types of data, instances, features, tasks and scenarios, parameters and hyper-parameters, performance measures

-----

-----

-----

-----

#### Detailed program

## Introduction

- \_\_\_\_\_
- Types of data and problems: tabular, streams, text, time-series, sequences, spatial, graph, web, social

## Unsupervised Learning

- \_\_\_\_\_
- Clustering
- Outlier detection

---

- \_\_\_\_\_
- Model-free/instance-based approaches, a simple algorithm: the k-nearest neighbors (KNN)
- Model-based approaches: Support Vector Machine (lineari)

-----

- \_\_\_\_\_
- Decision Tree and Random Forest
- Kernel-based learning: kernel-SVM and Gaussian Processes, for classification and regression
- Dimensionality reduction: Principal Component Analysis (PCA) and kernel-based PCA (kPCA)

---

- \_\_\_\_\_
- Deep Learning: *"a fraction of the connectionist tribe"*

-----

---

## Prerequisites

Basic knowledge on computer science, applied math, probability calculus and statistics

## Teaching form

Teaching is achieved by classes: lectures will be video-recorded in order to make them digitally available. Lectures will address both theory and hands-on, specifically the adoption of open data and software libraries.

## **Textbook and teaching resource**

- Reference textbook: Mehryar Mohri, Afshin Rostamizadeh and Ameet Talwalkar (2018). Foundations of Machine Learning.
- Slides and materials provided by the lecturer

## **Other suggested texts:**

- Charu C. Aggarwal (2015). Data Mining – the Textbook
- Carl Edward Rasmussen and Christopher K. I., Williams (2006). Gaussian Processes for Machine Learning.
- Robert B. Gramacy (2020). Surrogates – Gaussian Processes Modeling, Design, and Optimization for the Applied Statistics.

## **Semester**

Secondo semestre

## **Assessment method**

Assessment is organized on to tests:

- the development of a project along with the preparation of an associated technical report, similar to a scientific paper,
- an oral examination aimed at assessing the degree of understanding of the course's topics .

The project can be performed by working in *team* (max 3 students per group) and the datasets to adopt, in agreement with the lecturer, will be selected among those available on open platforms such as OpenML, Kaggle or UCI Repository.

The project amounts for 60% of the final mark, while the oral examination amounts for the remaining 40%.

## **Office hours**

On appointment

---