UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Machine Learning M

2627-2-F8206B014

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 python, specifically the suite "scikit-learn", but some equivalent solutions will also be presented in R (for example mlr and mlrmbo) and Java (for example WEKA, KNIME, SMAC3).

Contents

Machine Learning basics: types of data, instances, features, tasks and scenarios, parameters and hyper-parameters, performance measures
Similarity and distance (between data points, between data sets, between probability distributions)
Unsupervised learning techniques (based on distances)
Supervised learning techniques (based on distances): classification and regression
Modeling non-linearity in data: kernel-based techniques
The connectionists approach: artificial neural networks
Automated Machine Learning: automatic configuration of a Machine Learning model
Usage of Python libraries for implementing Machine Learning tasks

Detailed program

Introduction

- Machine Learning scenarios & tasks, useful notations
- Types of data and problems: tabular, streams, text, time-series, sequences, spatial, graph, web, social, immagini, distribuzioni

Similarity and distance

- Distance between data points
  - over continuous spaces: p-order Minkowski (and special cases: Manhattan, Euclidea, Chebyshev)
  - over discrete spaces: Hamming and Jaccard
- DIstance between datasets
  - "linkage" between sets of data points
  - the Monge's problem
- Distance between probability distributions
  - Optimal Transport theory: distance between two continuous, two discrete, or one continuous and one discrete probability distributions
  - Datasets as empirical discrete probability distributions: Monge's and Kantorovitch problem
  - Wasserstein distance and differences with divergences
  - Distance between datasets laying into two different spaces: Gromov-Wasserstein

Unsupervised Learning

- Clustering: deterministic vs probabilistic approaches; flat vs hierarchical; distance/similarity vs density -based
- Outlier and anomaly detection

Supervised Learning

- Foundations of "learning": binary classification, data generation process, concept vs hypothesis, empirical vs generalization error
- Classification and regression: metrics and validation techniques (hold-out, k fold-cross, leave-one-out)
- Model-free/instance-based approaches, a simple algorithm: the k-nearest neighbors (KNN)
- Model-based approaches: Support Vector Machine (linear)

Supervised Learning for non-linear data

- Non-linearity, VC dimensions, kernel-trick
- A brief recall on Decision Tree and Random Forest
- Kernel-based learning: kernel-SVM and Gaussian Processes, for classification and regression
- Dimensionlity reduction: Principal Component Analysis (PCA) and kernel-based PCA (kPCA)

The connectionist approach

- Artificial Neural Networks: perceptron, shallow networks e deep networks (Recurrent Neural Networks, Long-Short Term Memory, Gated Recurrent Unit)
- Generative neural models: Auto-Encoder (AE) and Variational-AE (VAE), Generative Adversarial Network (GAN) and Wasserstein-GAN (WGAN), Transformer

Automated Machine Learning (AutoML): an overview

- Hyeprparameter optimization of a Machine Learning algorithm
- Algorithm selection and (simultaneous) hyperparameter optimization of a Machine Learning algorithm via Bayesian Optimization

Exercises and examples

Prerequisites

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

Teaching form

Teaching is provided in-person. 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:

- Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). Mathematics for machine learning. Cambridge University Press.
- Charu C. Aggarwal (2023). Neural Networks and Deep Learning – A Textbook
- Bishop, C. M., & Bishop, H. (2023). Deep learning: Foundations and concepts. Springer Nature.
- Robert B. Gramacy (2020). Surrogates – Gaussian Processes Modeling, Design, and Optimization for the Applied Statistics.
- Charu C. Aggarwal (2015). Data Mining – the Textbook

Semester

First semester - First period

Assessment method

Assessment is organized on two tests:

- the development of a project along with the preparation of an associated technical report, similar to a scientific paper,
- an oral examination (individual and mandatory) 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 quality of the project will be assessed according to the correct adoption of ML algorithms and the analysis of the results. Oral examination is devoted to assess the understanding of theoretical and methodological aspects of ML.
The project amounts for 60% of the final mark, while the oral examination amounts for the remaining 40%.

There are no mid-term review(s)

Office hours

On appointment

Sustainable Development Goals

QUALITY EDUCATION