UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Machine Learning

2627-1-FDS02Q002-FDS02Q00201

Aims

To train the data analysis expert according to the machine learning methodology.
The goal is achieved by;

- teaching how to design, develop and present machine learning projects,
- exploiting open source platforms, languages and software,
- stimulating the team working methodology.

At the end of the course the student will gain skills and knowledge so as to design, develop, document and present a machine learning study. **(DdD 1, DdD 2)**.
Moreover, the organization of the exam, through a group project to be developed and documented, will allow the candidate to develop critical thinking skills **(DdD 3)** and communication skills **(DdD 4)**, while the availability of reading materials suggested by the instructor will enable the candidate to continue studying and deepening their understanding of the subject independently **(DdD 5)**.

Contents

The course contents are the following;

- Data Exploration to inspect and summarize the available data and to design and develop a pre-processing workflow,
- Supervised Classification, to learn a mapping from input attributes to output or target attributes to be classified or predicted,
- Unsupervised Classification, to form homogeneous groups of observations and/or attributes using a given proximity measure,

The course also hosts 2 vertical topics

- **From Statistical to Neural Language Models;** This focused module introduces the evolution of language models, from statistical approaches based on N-grams to neural architectures and modern Large Language Models (LLMs). Students will gain an overview of the key concepts underlying language representation, neural language modeling, and the pretraining–fine-tuning paradigm that powers today's AI systems.
- **A gentle introduction to Causal Bayesian Networks;** This vertical module introduces causal Bayesian networks, a probabilistic graphical model which is widely applied to several application domains as economy, healthcare and medicine.

You will learn how to develop machine learning workflows using the KNIME open source software platform. You are not required to code any programs while if you want KNIME allows to use powerful and professional open source programming languages and commercial software environments; R, Weka, Matlab, Python, Java, ...

Detailed program

Data Exploration and Preprocessing

- Data types and attributes
- Graphical and tabular data exploration
- Missing data treatment
- Data Pre-Processing

Supervised Classification

- Introduction
- Techniques, models and algorithms; artficial neural nets, Bayesian classifiers, decision trees, ...
- Performance measures to evaluate and compare classifiers
- Unbalanced classes and non binary classification

Unsupervised Classification

- Introduction
- Proximity measures for nominal, ordinal and continuous attributes
- Techniques, models and algorithms; partitioning, hierarchical, graph based, density based, ...
- Performance measures to evaluate and compare clustering solutions

Prerequisites

Basic knowledge on; informatics, probability calculus and statistics.

Teaching form

Teaching is achieved by classes. The entire course is also available in digital form consisting of video lectures for theory and hands-on. The course material is organized through learning paths where lecture modules consist of theoretical lecture, hand-on lecture and self-evaluation sessions. Self-evaluation session offers a powerful and

effective resource to online learning, i.e. after the class has taken place, The course makes available 170 quizzes to allow students to fairly assess their understanding level and to train for the exam.

- 14 lectures of 2 hours each of erogative nature delivered in physical presence.
- 9 hands-on lectures of 2/3 hours each of erogative nature delivered in physical presence.

Textbook and teaching resource

Video-lectures, slides, datasets and workflows designed and developed by the teacher.

https://www-users.cse.umn.edu/~kumar001/dmbook/index.php
https://github.com/kerasking/book-1/blob/master/ML%20Machine%20Learning-A%20Probabilistic%20Perspective.pdf

Semester

First semester

Assessment method

The examination consists of a 3 to 4 hours test during which the candidate is required to answer 6 quizzes (1 point for each correct answer) and 1 open?ended question (up to a maximum of 9 points) **(DdD 5)**. In addition, a machine learning problem will be assigned. Specifically, a problem will be presented, a dataset will be provided, and the candidate will be required to design and develop a KNIME workflow **(DdD 3, DdD 4)** (each node of the workflow must be commented using the functionalities offered by the KNIME platform; up to 10 points are awarded). The candidate must also prepare slides following a template provided by the instructor (up to 5 points). Finally, upon request, the candidate may take an oral exam, which assigns a score between –3 and +3 points.
No midterm assessments are scheduled.

The workflow and the slides will be evaluated according to the following criteria:

1. Technical merit, accuracy, and correctness of the workflow.
2. Clarity of presentation and communication of ideas, including readability and critical insight of the slides.

Office hours

By appointment, send and email to fabio.stella@unimib.it

Sustainable Development Goals