UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Machine Learning

2526-1-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,

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 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

Assessment is based on two components, a machine learning project and a methodology exam which is performed in the laboratory by using a computer. Students and encouraged to work in small teams to design, develop and document their data and/or text mining project. The data and/or text mining project is usually selected by the students team by exploiting the Kaggle platform (https://www.kaggle.com/) where Data Science requests and offers meet.
No interim assessments are scheduled.

- **Project work; (DdD 3, DdD 4)** The student is asked to design and develop a KNIME workflow to solve a problem similar to those presented during the course **(Awards maximum 21 points)**.The machine learning project gives a maximum of 21 points, assigned according to six criteria as follows: 1) Technical merit: notably rigour, accuracy and correctness (maximum 5 points). 2) Clarity of expression and communication of ideas; including readability and discussion of concepts (maximum 5 points). 3) Appropriate referencing and the context of the present work (maximum 2 points). 4) Overall balance and structure of report (maximum 3 points). 5) Repetition; have significant parts of the manuscript already been published by other authors? (maximum 3 points). 6) Diagrams, tables, captions; are they clear and essential (maximum 3 points).

- **Open and closed answer questions; (DdD 5)** on the topics presented during classes **(Awards maximum 11 points)**
  The methodology exam gives a maximum of 11 points. according to the following; 6 points for 6 quizzes, one point for each quiz (each quiz concerns concepts presented in the course) and a maximum of 5 points for an open ended question having the goal to evaluate the critical point of view of the candidate. The candidate can also ask to undergo oral examination, consisting of questions about reasoning and deduction about the concepts presented in the course, which gives a maximum of 3 points.

WARNING: It is possible to take the lab exam only after submitting the project. Furthermore, for each submitted project, only one lab exam attempt is allowed. If the grade is rejected, a new project must be submitted. Finally, the exam must be completed within the same exam date or, at most, within two consecutive exam dates.

Office hours

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

Sustainable Development Goals

QUALITY EDUCATION | REDUCED INEQUALITIES