



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Machine Learning

2223-1-F5105P030

Learning area

Learning objectives

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.
- The student will be able to design, develop, document, and present machine learning projects solving real world problems.

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,
- Association Rules, to automatically extract rules hidden in the data with specific reference to transaction data.

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; artificial 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

Association Rules

Introduction and basic definitions

Item and itemsets

Apriori, principle and algorithm

Performance measures to evaluate and compare association rules

Prerequisites

Basic knowledge on; informatics, probability calculus and statistics.

Teaching methods

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.

Assessment methods

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 are 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.

The machine learning project gives a maximum of 21 points, assigned according to six criteria as follows:

- Technical merit: notably rigour, accuracy and correctness (maximum 5 points)
- Clarity of expression and communication of ideas; including readability and discussion of concepts (maximum 5 points)
- Appropriate referencing and the context of the present work (maximum 2 points)
- Overall balance and structure of report (maximum 3 points)
- Repetition; have significant parts of the manuscript already been published by other authors? (maximum 3 points)
- Diagrams, tables, captions; are they clear and essential (maximum 3 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.

Textbooks and Reading Materials

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

http://dsd.future-lab.cn/members/2015nlp/Machine_Learning.pdf

Sustainable Development Goals
