UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Unsupervised Learning

2627-1-F9103Q045-F9103Q04502

Aims

To develop skills for solving real worls unsupervised learning problems.

The goal is achieved by;

- Teaching how to design, train, deploy and monitor unsupervised learning models. **(DdD 1, DdD 2)**
- Exploiting open source platforms, languages and software, **(DdD 1, DdD 2)**
- Stimulating team working. **(DdD 4)**
- Reading and discussing scientific papers made available by the teacher **(DdD 3, DdD 5)**

Contents

The course contents are the following;

- **Data Types**; to list different types of data and to learn hw they must be used for unsupervised learning.
- **Data Preprocessing**; to preprocess data in such a way it can be used by unsupervised learning tasks,
- **Clustering Learning**; to form homogeneous groups of observations and/or attributes using a given proximity measure,
- **Clustering Validation**; to evaluate and compare diferent clusteirng solutions to select the one to deploy.
- **Anomaly Detection**; to find anomalous observations, to discover outliers observations, under different theoretical settings.
- **Bayesian Networks**; to learn probabilistic/causal structure from data and to make decisions under uncertainty.

You will learn how to design, train, validate and deploy unsupervised learning models using Python.

Detailed program

1. Data
1.1 Data types and attributes
1.2 Proximity measures for nominal, ordinal and continuous attributes
1.3 Data Pre-Processing

2. Cluster Analysis
2.1 Introduction
2.2 Clustering algorithms
2.2.1 Partitioning
2.2.2 Hierarchical
2.2.3 Graph-based
2.2.4 Density-based
2.2.5 Time-series
2.3 Comparing clustering solutions
2.3.1 Performance measures
2.3.2 Evaluation
2.3.3 Comparison

3. Anomaly Detection
3.1 Introduction
3.2 Anomaly detection algorithms
3.2.1 Statistical approaches
3.2.2 Proximity-based approaches
3.2.3 Clustering-based approaches
3.2.4 One-class classification
3.2.5 Information theoretic approaches

4. Bayesian Networks
4.1 Introduction
4.2 Bayesian network models
4.2.1 Discrete variables
4.2.2 Continuous variables
4.2.3 Mixed variables
4.3 Learning
4.3.1 Parameters
4.3.2 Structure
4.4 Inference
4.4.1 Exact
4.4.2 Approximate

Prerequisites

Basic knowledge on: probability theory, statistics, mathematics.
Good skills to design and develop computer programs.

Teaching form

The course is organized as follows:

- 16 lectures of 2 hours each of theory in physical presence of erogative nature
- 12 lectures of 2 hours each of hands-on in physical presence of interactive nature

Textbook and teaching resource

- **Introdution to Data Mining** (https://www-users.cse.umn.edu/~kumar001/dmbook/index.php)
- **Bayesian Networks and Decision Graphs** (https://link.springer.com/book/10.1007/978-0-387-68282-2)

Semester

Spring Semester

Assessment method

The exam consists of:

- **Project work;** The student is asked to apply and/or develop one or more algorithms for analizing a case study assigned by the teacher. **(Awards a maximum of 13 points)**.
- **Interview on project work;** An interview on those topics that have been presented in classes and connected to the project works **(Awards a maximum of 20 points)**.

No interim assessments are scheduled.

Office hours

To be agreed on by mail message fabio.stella@unimib.it

Sustainable Development Goals