UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Data Mining M

2122-2-F8204B014

Learning objectives

The course aims to provide an introduction to deep learning and to improve predictive modelling skills by using the R software environment for statistical computing. The course is divided in two parts:

1. **Introduction to Deep Learning** (3CFU, Prof. Borrotti);
2. **Applied Predictive Modelling** (3CFU, Prof. Solari).

Contents

Introduction to Deep Learning

TBD

Applied Predictive Modelling
The goal is to elucidate a framework for constructing models that generate accurate predictions. This framework includes pre-processing the data, splitting the data into training and testing sets, selecting an approach for identifying optimal tuning parameters, building models, and estimating predictive performance.

Detailed program

Introduction to Deep Learning
Introduzione al Deep Learning
Reti Neurali
Reti Neurali Shallow e Deep

Recurrent Neural Network
Convolutional Neural Network

Applied Predictive Modelling
Important concepts: overfitting, bias and variance tradeoff, optimism
Ensemble methods
The model and the modelling process

Prerequisites

Knowledge of topics covered in the courses *Probability and Statistics M* and *Advanced Statistics M* is highly recommended.

Teaching methods

Lessons are held both in classroom and in lab, integrating theoretical principles with practicals of data analysis and programming in R.

Assessment methods

Introduction to Deep Learning

TBD

Applied Predictive Modelling

Data analysis with the goal of predicting the observations in the test set. In addition to the predictions, a report containing the description of the analysis and the code used must be submitted by the deadline (at least one week before the exam session). It will be possible to submit the predictions only once per Academic Year. If the student (or the teacher) requests the oral exam, the final grade is an average of the marks from the data analysis and the oral exam. The data analysis aims at verifying the predictive modelling skills.

Textbooks and Reading Materials

Introduction to Deep Learning

- Efron, Hastie (2016) *Computer-Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press.
- Bishop (2009) *Pattern Recognition And Machine Learning*. Springer.
- Goodfellow, Bengio, Courville (2017) *Deep Learning*. MIT.
- Chollet, Allaire (2018) *Deep Learning with R*. Manning

Applied Predictive Modelling

- Azzalini, Scarpa (2004). *Analisi dei dati e data mining*. Springer-Verlag Italia
- Kuhn, Johnson (2019). *Feature Engineering and Selection*. Chapman and Hall/CRC
- Kuhn, Silge (2021+). *Tidy Modeling with R.* (in progress)
- Lewis, Kane, Arnold (2019) *A Computational Approach to Statistical Learning*. Chapman And Hall/Crc.

Semester

First semester, second period.

Teaching language

The lessons are held in Italian, but the textbooks are in English.