UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Streaming Data Management and Time Series Analysis

2526-2-FDS01Q023

Aims

The course illustrates methods and applications for managing, analyzing, and forecasting - possibly streaming - time series.

Besides data managing applications, our lessons cover linear (ARIMA, state-space/Kalman filter) and nonparametric (machine learning) methods.

The student who successfully follows this course will be able to manage streaming data and select, identify, and implement the time series model to fit the data and address the problem under analysis.

The course is part of the statistics learning area of the master's degree program in Data Science.

Contents

Streaming data management, linear-filter-based models (ARIMA, VAR), unobserved component models (state-space form/Kalman filter), nonparametric methods (nonparametric regression, tree-based methods, neural networks, support vector machines, nearest neighbors, etc.).

Detailed program

First part

- Theory of statistical prediction (best predictor, best linear predictors).
- Stationary and integrated processes

- ARIMA models
- VAR models and cointegration (basic concepts)
- Unobserved Component Models (UCM)
- State-space form
- Kalman filter and maximum likelihood estimation of the model in state-space form
- State and disturbance smoothing
- Many applications to actual data using R (or Python)

Second part

- Main time series mining tasks
- Similarity and Clustering
- Classification, regression, and forecasting
- Non-parametric approaches based on statistical methods
- Non-parametric approaches based on machine Learning
- Artificial Neural Networks

Prerequisites

Attending students should know statistical inference, matrix algebra, and R, or Python (we will adopt R in class, but you can use Python if you prefer).

Teaching form

Theoretical lectures and hands-on sessions in the computer lab. Each lesson will be a mix of lecture-based and interactive teaching. Every time a new theoretical concept or tool is introduced, a practical application carried out together with the students will demonstrate its implementation. In fact, theory serves practice, and practice helps to understand theory.

Textbook and teaching resource

Rob J Hyndman and George Athanasopoulos, *Forecasting: Principles and Practice* (2nd ed): https://otexts.com/fpp2/

Pelagatti M. (2015) *Time Series Modelling with Unobserved Component Models*. Chapman and Hall/CRC (il libro è scaricabile gratuitamente sotto indirizzo IP di Bicocca).

Abhijit Ghatak (2019) Deep Learning with R. Springer

Further material will be available in the elearning platform.

Semester

First semester

Assessment method

The examination is organized in two parts. First, by the date of the examination, each student must produce and send to the lecturer a paper in which they have to analyze and predict one or more time series (in agreement with the lecturers) using linear (ARIMA, UCM) and non-linear methods (RNN, SVM, etc.). The student will illustrate the paper during the oral examination in ca. 15 minutes, and the lecturers will ask questions about its content. On the same day of the oral exam, there will also be a one-hour written assessment, which consists in answering five theoretical questions on ARIMA and UCM models.

To pass the exam, both parts must have a positive valuation, and the final grade will be computed as the arithmetic mean of the grades of the two parts.

The evaluation of the theoretical part is based on the exactness and completeness of the answers (each answer is equally weighted). The assessment of the prediction exercise is based on the quality of the modeling. We will pay particular attention to feature engineering and model selection procedures.

No in-course assessments will take place.

Office hours

Pelagatti: by appointment (matteo.pelagatti@unimib.it).

Sustainable Development Goals