UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Streaming Data Management and Time Series Analysis

2122-2-F9101Q017

Aims

The course illustretes methods and applications for managing, analysing and forecasting - possibly streaming - time series.

Beside data managing applications, our lessons cover both linear (ARIMA, VAR, state-space/Kalman filter) and nonlinear (neural networks, support vector machine) methods.

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

Contents

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

Detailed program

First part (Pelagatti)

- Theory of statistical prediction (best predictor, best linear predictors).
- Stationary and integrated processes
- ARIMA models
- VAR models and cointegration
- Unobserved Component Models (UCM)
- State-space form
- Kalman filter and maximum likelihood estimation of model in state-space form

- State and disturbance smoothing
- Many applications to real data using R and SAS

Second part (Candelieri)

- Nature of time series data
    - Representing time series: raw data, features extraction, modelling
    - Historical versus streaming data
    - Managing time series data: time series databases
- Main time series mining tasks
    - Similarity and Clustering
    - Classification, regression and forecasting
- Non-parametric approaches based on machine Learning
    - Artificial Neural Networks
    - Support Vector Machines

Prerequisites

Attending students should know statistical inference and R.

Teaching form

Theoretical lessons in class and computer applications in lab. It would be helpful if the students could bring their computers (with R installed) also during the lesseons in classroom.

Textbook and teaching resource

Pelagatti M. (2015) Time Series Modelling with Unobserved Component Models. Chapman and Hall/CRC (freely available under Bicocca's IP addresses).

Galit Shmueli, Kenneth C. Lichtendahl Jr. “Practical Time Series Forecasting with R: A Hands-On Guide” [2nd Edition] (Practical Analytics) – July 19, 2016

Further material will be available in the elearning platform.

Semester

First semester

Assessment method

The examination is organized in two parts. By the date of the examination each student has to produce and send to

the lecturers a paper in which he/she has 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 and the lecturers will ask questions about its content. On the same day of the oral exam, there will be also a one-hour written assessment, which consists in answering to five theoretical questions on ARIMA and UCM models.

In order to pass the exam both parts must have a positive valuation and the final grade will be computed as simple 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 modelling. We will pay particular attention to feature engineering and model selection procedures.

Office hours

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

Candelieri: Tuesday 10:00-12:00 (U14)