UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Streaming Data Management and Time Series Analysis

2324-2-FDS01Q023

Obiettivi

Il corso fornisce gli strumenti per gestire, analizzare e prevedere dati in forma di serie storica anche quando generati in tempo reale.

Oltre alle metodologie per la gestione del dato in tempo reale, il corso copre sia modelli lineari (ARIMA, VAR, state-space/filtro di Kalman) sia modelli non parametrici (machine learning).

Lo studente che avrà seguito l'insegnamento con successo saprà gestire dati in tempo reale, scegliere e identificare il modello di serie storiche più adatto al problema e produrre scomposizioni e previsioni delle serie storiche sotto analisi.

Contenuti sintetici

Gestione dei dati in tempo reale, modelli basati su filtri lineari (ARIMA, VAR), modelli basati su componenti non osservabili (state-space/Kalman filter), modelli non lineari (reti neurali, support vector machine, vicini più vicini).

Programma esteso

Prima parte

- Teoria della previsione statistica (miglior previsore e miglior previsore lineare).
- Processi stazionari e integrati
- Modelli ARIMA
- Modelli VAR e cointegrazione

- Modelli a componenti non nosservabili (UCM)
- Forma state-space
- Kalman filter e stime di massima verosimiglianza di modelli in forma state-space
- Smoothing delle variabili di stato e dei disturbi (estrazione comonenti e identificazione di anomalie).
- Applicazioni a dati reali usando R (o Python)

Seconda parte

- Principali task di time-series mining
- Classificazione, regressione e previsione
- Approcci non-parametrici statistici
- Approcci non-parametrici basati su Machine Learning
- Reti Neurali

Prerequisiti

Per seguire l'insegnamento con successo è necessario conoscere l'inferenza statistica, R o Python (useremo R nelle lezioni, ma potete utilizzare anche Python se preferite).

Modalità didattica

Lezioni teoriche e pratiche in laboratorio informatico.

Materiale didattico

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

Altro materiale sarà reso disponibile sulle pagine elearning.

Periodo di erogazione dell'insegnamento

Primo semestre

Modalità di verifica del profitto e valutazione

L'esame è organizzato in due parti. Entro la data dell'esame ciascuno studente dovrà produrre e inviare ai docenti un elaborato dove una o più serie storiche concordate con il docente dovranno essere analizzate e previste per mezzo di modelli ARIMA, UCM e machine learning. Lo studente illustrerà l'elaborato durante l'orale in una quindicina di minuti e i docenti potranno fare domande sul contenuto. Nel medesimo giorno dell'orale vi è anche uno scritto dalla durata di un'ora che prevede la risposta a cinque tra domande teoriche ed esercizi su modelli ARIMA e UCM.

Per superare l'esame entrambe le parti dovranno essere sufficienti e il voto finale sarà calcolato come media aritmetica semplice delle votazioni delle due parti.

La valutazione della parte teorica sarà basata sull'esattezza e la completezza delle risposte alle domande proposte (ogni risposta ha lo stesso peso nel voto). La valutazione dell'elaborato sarà basata sulla qualità dei modelli costruiti e, in particolare, sulle features costruite e sulla selezione dei modelli finali.

Orario di ricevimento

Pelagatti: su appuntamento (matteo.pelagatti@unimib.it).

Sustainable Development Goals