

SYLLABUS DEL CORSO

Streaming Data Management and Time Series Analysis

2122-2-F9101Q017

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 lineari (reti neurali, support vector machine).

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 (Pelagatti)

- Teoria della previsione statistica (miglior previsore e miglior previsore lineare).
- Processi stazionari e integrati
- Modelli ARIMA
- Modelli VAR e cointegrazione
- Modelli a componenti non osservabili (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 componenti e identificazione di anomalie).
- Molte applicazioni a dati reali usando R e SAS

Seconda parte (Candelieri)

- La natura delle serie temporali
 - Modalità di rappresentazione: raw data, features extraction, modelling
 - Dati storici versus dati streaming
 - Strumenti di gestione: time series databases
- Principali task di time-series mining
 - Similarità e Clustering
 - Classificazione, regressione e previsione
- Approcci non parametrici basati su Machine Learning
 - Reti Neurali
 - Macchine a Vettori di Supporto

Prerequisiti

Per seguire l'insegnamento con successo è necessario conoscere l'inferenza statistica e R.

Modalità didattica

Lezioni teoriche in aula e pratiche in laboratorio informatico. Sarebbe utile se gli studenti portassero con se un computer con installato R anche durante le lezioni in aula.

Materiale didattico

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

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

Ulteriore materiale sarà reso disponibile all'interno della piattaforma 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 i docenti dovranno essere analizzate e previste per mezzo di modelli ARIMA, UCM e non-lineari (reti ricorrenti, SVM, ecc.). Lo studente illustrerà l'elaborato durante

l'orale 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 domande teoriche 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).

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