



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Statistica Non Parametrica M

2425-2-F8204B041

Obiettivi formativi

Il corso intende esplorare i principi fondamentali e le metodologie proprie della statistica nonparametrica moderna. I metodi considerati permetteranno di affrontare un'ampia gamma di problemi applicati e la loro implementazione avverrà mediante l'utilizzo del software open-source R. Alla fine del corso lo studente sarà in grado di: i) individuare le applicazioni in cui la metodologia nonparametrica risulti opzione più conveniente di quella parametrica, ii) effettuare un'analisi di dati completa con metodi nonparametrici tramite l'utilizzo di R, iii) valutare con adeguati strumenti teorici la performance delle diverse procedure disponibili.

Contenuti sintetici

Il corso si propone di affrontare in modo rigoroso il problema di stimare una distribuzione senza fare assunzioni parametriche. Lo stesso problema verrà formulato in termini di stima di funzioni di ripartizione e di stima di densità. Diversi strumenti e metodi saranno considerati, con attenzione a vantaggi e svantaggi, teorici e computazionali, di ciascuno. In base al tempo a disposizione, altri temi che potranno toccati essere riguarderanno i modelli di regressione non parametrici e i test per il goodness-of-fit.

Il corso sarà diviso in due parti dove gli stessi temi verranno trattati seguendo un approccio prima classico e poi bayesiano. Il corso si occuperà sia dello studio delle proprietà teoriche dei modelli nonparametrici che della loro implementazione per l'analisi dati. I laboratori richiederanno l'utilizzo degli strumenti numerici più adatti all'implementazione dei modelli nonparametrici, strumenti che includono algoritmi di ottimizzazione e algoritmi di simulazione di tipo Markov chain Monte Carlo (MCMC).

Programma esteso

Parte 1 (approccio classico)

- stima di una funzione di ripartizione
- bande di confidenza per una funzione di ripartizione
- stima di funzionali statistici
- il metodo bootstrap
- smoothing
- istogramma
- stima di densità attraverso kernel
- modelli di regressione kernel (in base al tempo a disposizione)
- test per goodness-of-fit (in base al tempo a disposizione)

Parte 2 (approccio bayesiano)

- scambiabilità e prior nonparametriche
- il processo di Dirichlet
- Proprietà del processo di Dirichlet
- distribuzione predittiva del processo di Dirichlet
- stima di densità tramite misture di processo di Dirichlet
- problemi di clustering tramite misture di processo di Dirichlet
- modelli di regressione tramite misture nonparametriche (in base al tempo a disposizione)
- test bayesiano per goodness-of-fit (in base al tempo a disposizione)

Prerequisiti

Si richiede la conoscenza dei corsi base di Statistica ad un livello di laurea triennale. E' inoltre consigliata la conoscenza della statistica bayesiana parametrica, in tal senso è **fortemente consigliato** il corso di "Statistica bayesiana".

Metodi didattici

Lezioni frontali accompagnate da alcune sessioni di laboratorio. Nello specifico sono previste:

- 6 lezioni frontali da 3 ore (modalità erogativa)
- 6 lezioni frontali da 2 ore (modalità erogativa)
- 6 attività di laboratorio da 2 (modalità interattiva)

E' previsto che l'80% siano erogate in presenza e il 20% da remoto.

Modalità di verifica dell'apprendimento

L'esame consiste di una prova orale composta di due parti:

1. (obbligatoria) colloquio sugli argomenti svolti a lezione;
2. (opzionale) colloquio su argomenti di approfondimento non trattati a lezione.

La parte obbligatoria (parte 1) è individuale ed è atta alla valutazione della preparazione dello studente su tutti gli argomenti trattati durante le lezioni e i laboratori.

La parte opzionale dell'esame orale (parte 2) prevede lo svolgimento e la presentazione di un progetto di gruppo su un tema da concordare con il docente. Questa parte è volta a valutare la capacità degli studenti di studiare, approfondire, utilizzare strumenti avanzati di statistica nonparametrica.

Non sono previste prove in itinere.

Testi di riferimento

- Wasserman, L., 2006. All of nonparametric statistics. Springer Science & Business Media.

Ulteriori materiale e referenze verranno forniti durante il corso.

Periodo di erogazione dell'insegnamento

Primo semestre, primo ciclo.

Lingua di insegnamento

inglese

Sustainable Development Goals

ISTRUZIONE DI QUALITÀ
