UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Computational Statistics

2122-3-E4102B085-E4102B087M

Obiettivi formativi

L'obiettivo principale del corso è introdurre strumenti software avanzati e di alta complessità computazionale per disegnare ed eseguire analisi di dati e modellazione statistica complessa.Alla fine del corso lo studente ha la possibiltà di proporre i principali algoritmi , discernendo pregi e difetti, essendo in grado di sperimentare ed applicare le conoscenze acquisite su dati reali.

Contenuti sintetici

Il corso affronta lo studio di tecniche modellistiche algoritmiche e le principali problematiche e tecniche statistiche di statistica complessa

Programma esteso

(1) SAS language and R (overiview)

(2) Interpretazione di Modelli lineari complessi (Anova, Ancova, GLM) con interazioni, trasformate,

(3) Robust methods (Bootstrap, Jacknife, Robust Regression, IRLS, WLS, nonparametric regression,

loess smoothing and splines)

(4) Passi per costruzione di un modello Robusto

(5) missing data mechanism, missing imputation, (y, X)-transformation, misure di Influenza, diagnostiche, heteroschedaticità, model selection.

(6) Binary and multinomial logistic regression

Prerequisiti

Superamento esame di Analisi statistica Multivariata

Metodi didattici

Lezione frontale e sessioni di laboratorio

Modalità di verifica dell'apprendimento

PROVA SCRITTA

PROJECT WORK (Sviluppo di un progetto originale a partire da una semplice idea o dall'analisi di un caso esistente)

Due Lavori applicativi da svolgere autonomamente o in gruppo di max 3 persone su un dataset scelto dallo studente (R o SAS) su cui applicare i principali argomenti svolti a lezione .

L'elaborato (doc, html, pdf, R markdown) va spedito via mail al docente (piergiorgio.lovaglio@unimib.it) una settimana prima dell'appello con file nominato con i cognomi degli appartenenti al gruppo di lavoro

Di seguito le analisi da svolgere per le due analisi applicative (Sas base o R):

PROJECT WORK completo con

1 Analisi di regressione e sviluppo del modello robusto con target quantitativo

(analisi descrittive, collin, missing data, heteroscedasticità, dati influenti, trasformazioni, diagnostiche, costruzione modello robusto, bootstrap)

1 Analisi di regresione logistica con target binario (discretizzare il target dell'analisi precedente)

(modello logistico con tutte le covariate e con model selection, controllare solo collinearity e separation)

I principali output del PROJECT WORK sul modello robusto e sull'analisi logistica vanno stampati e portati all'orale (se in presenza)

Portali WEB per la scelta del dataset:

https://archive.ics.uci.edu/ml/datasets

www. kaggle.com

PROVA ORALE

COLLOQUIO DI DISCUSSIONE SULLO SCRITTO

L'esame orale consta di domande sulla TEORIA affrontata a lezione e sul commento degli output dei lavori applicativi per verificare la comprensione dei principali strumenti adottati e il conseguente "modus operandi" dell'analisi statistica svolta.

Testi di riferimento

Carter Hill, William E. Griffiths, Guay C. Lim. Principles of Econometrics (chapters 2, 4 ,5, 6, 7, 8, 16 fino sezione 16.3 compresa)

Gareth, Witten, Hastie, Tibshirani, An Introduction to Statistical Learning with Applications in R Chapter 3 (no section 3.5), Chapter 4, 6,7

Lucidi sul moodle

Consigliati

Principles of Econometrics associate R book https://bookdown.org/ccolonescu/RPoE4/

Periodo di erogazione dell'insegnamento

I semestre, I ciclo

Lingua di insegnamento

Italiano