Statistical Learning

F8204B015
F8204B033M

Aldo Solari

Docente

E-mail	:	aldo.solari@unimib.it
Pagina web:	:	https://aldosolari.github.io/
Ricevimento	:	Mercoledì, 17:00-18:00
		stanza 2030, edificio U7, II piano

Prerequisiti

- Probabilità e Statistica Computazionale M
  - Probabilità Applicata
  - Statistica Computazionale
- Statistica Avanzata M
  - Statistica Multivariata
  - Teoria dell'Inferenza Statistica

Obiettivi del corso

Acquisizione di metodi inferenziali all'avanguardia per l'analisi di dati complessi e perfezionamento delle abilità di *problem solving* e di programmazione in R tramite l'utilizzo di casi studio. In particolare:

- Formulate quantitative models to address scientific questions
- Apply a range of statistical methods for inference
- Organize and perform a complete data analysis, from exploration, to analysis, to synthesis, to communication

Workload

- $\approx 1/3$ applicato
- $\approx 2/3$ teorico

Materiale didattico

Il materiale didattico è scaricabile dalla pagina web del corso

https://github.com/aldosolari/SL

Si precisa che:

- è soggetto a modifiche, quindi si prega di controllare l'ultima versione aggiornata;
- verrà caricato progressivamente, in parallelo allo svolgimento delle lezioni.

Riferimenti bibliografici

EH Efron and Hastie (2016) Computer-Age Statistical Inference: Algorithms, Evidence, and Data Science, Cambridge University Press

GI Giraud C. (2015) Introduction to High-Dimensional Statistics. Chapman and Hall/CRC

NL Nolan and Lang (2015) Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving. Chapman and Hall/CRC

...

Prova d'esame

(A) Data analysis assignment
(B) Prova orale
   1 domanda su (A) + 3 domande di teoria

Data analysis assignment

- Each student will be required to perform a data analysis project
- Before the exam, each student must submit
    1. A write-up of their data analysis in a synthesized format, with numbered figures and references. (You may also include supplementary material for detailed additional calculations/analyses)
    2. A reproducible .Rmd file that produces all of the numbers, figures and results in your write-up.

Data analysis assignment

The grades will be broken down according to the following characterization of your data analysis:

❶ Did you answer the scientific question?

❷ Did you use appropriate statistical methods?

❸ Was your write-up simple, clear, and precise?

❹ Was your code reproducible?

Keep in mind that this is a data science class. In some cases standard methodology will be sufficient to answer the question of interest, in some cases you will need to go beyond the course, and in general the goal is to answer the question and provide an estimate of uncertainty. You may speak to your fellow students about specific statistical questions related to the projects, but the overall idea, analysis, and write-up should be your own individual work.

One data set, many analysts

Do soccer referees give more red cards to dark skin toned players?