

Course overview

Modern Inference

Aldo Solari

Contact

E-mail	<code>aldo.solari@unimib.it</code>
Course page	<code>https://aldosolari.github.io/MI</code>
Course repo	<code>https://github.com/aldosolari/MI</code>
Skype	<code>aldo.solari</code>

Calendar

1	27 March	13:30-16:30 (3h)
2	30 March	13:30-16:30 (3h)
3	31 March	9:30-11:30 (2h)
4	3 April	11:30-16:30 (3h)
5	6 April	13:30-16:30 (3h)
6	7 April	9:30-11:30 (2h)
7	10 April	11:30-16:30 (3h)
8	17 April	13:30-15:30 (2h)

Topics

The main goal of this course is to expose students to modern ideas in statistical inference:

- The Crisis of Modern Science
- The Law of Selection
- High-Dimensional Statistics
- Multiple Testing (FWER and FDR)
- Post-Selection Inference (Uniform bounds on FDP)
- High-Dimensional Inference (Sample-Splitting)
- Modern Methods for Variable Selection (Stability Selection and the Knockoff Filter)

To explain or to predict?

- Explanatory modelling and predictive modelling (what you've seen in DM and IntroDL) are two different goals
- True model

$$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- Underspecified model

$$Y = \gamma_1 X_1 + \epsilon$$

- Explanation aims at estimating the coefficients β_1 and β_2 of the true model, but a wrong model can sometimes predict better (when predictors X_1 and X_2 are highly correlated, when the data are very noisy, when β_2 is small, etc.)

Reference

Shmueli (2010)

To explain or to predict?

Statistical science

Exam

Two options:

- ❶ Oral examination
- ❷ Homework and (shorter) oral examination

For the homework:

- Students can participate in groups (max. 3 persons)
- There will be 5 homework; the lowest score will be dropped in the final grade
- Each homework has a deadline: solutions should be submitted via e-mail (late submissions will *not* be accepted)

Homework

N	Date	Deadline
HW1	27/03/2020	06/04/2020
HW2	31/03/2020	10/04/2020
HW3	03/04/2020	10/04/2020
HW4	06/04/2020	17/04/2020
HW5	10/04/2020	17/04/2020

Homework rules

- 1 You will have to submit your solution by the deadline
- 2 Feel free to choose the format for your solution (.txt, .tex, .pdf etc.), the nicer the better, and the number of files in attachment (one or more, but not too many)
- 3 Answer with clarity and precision. All R code must be reproducible
- 4 For theoretical questions, try to provide a well-reasoned mathematical argument. Simulations can help form your intuition; but a purely empirical answer will only receive partial points
- 5 It is encouraged to discuss the problem sets with others, but every group needs to turn in a unique write-up. Use of sources (people, books, internet and so on) without citing them in homework sets results in failing grade
- 6 At the end of the homeworks, each member must fill a peer rating form of team members. The individual score will be the team score adjusted based on peer ratings

Forum

If you have any questions or are in need of clarification,
please post a message to the forum (e-learning).

I'll try to answer on the forum.

Only if necessary, write me an email.