Unsupervised Learning

Winter semester, 2025/2026

RULES OF ASSESSMENT
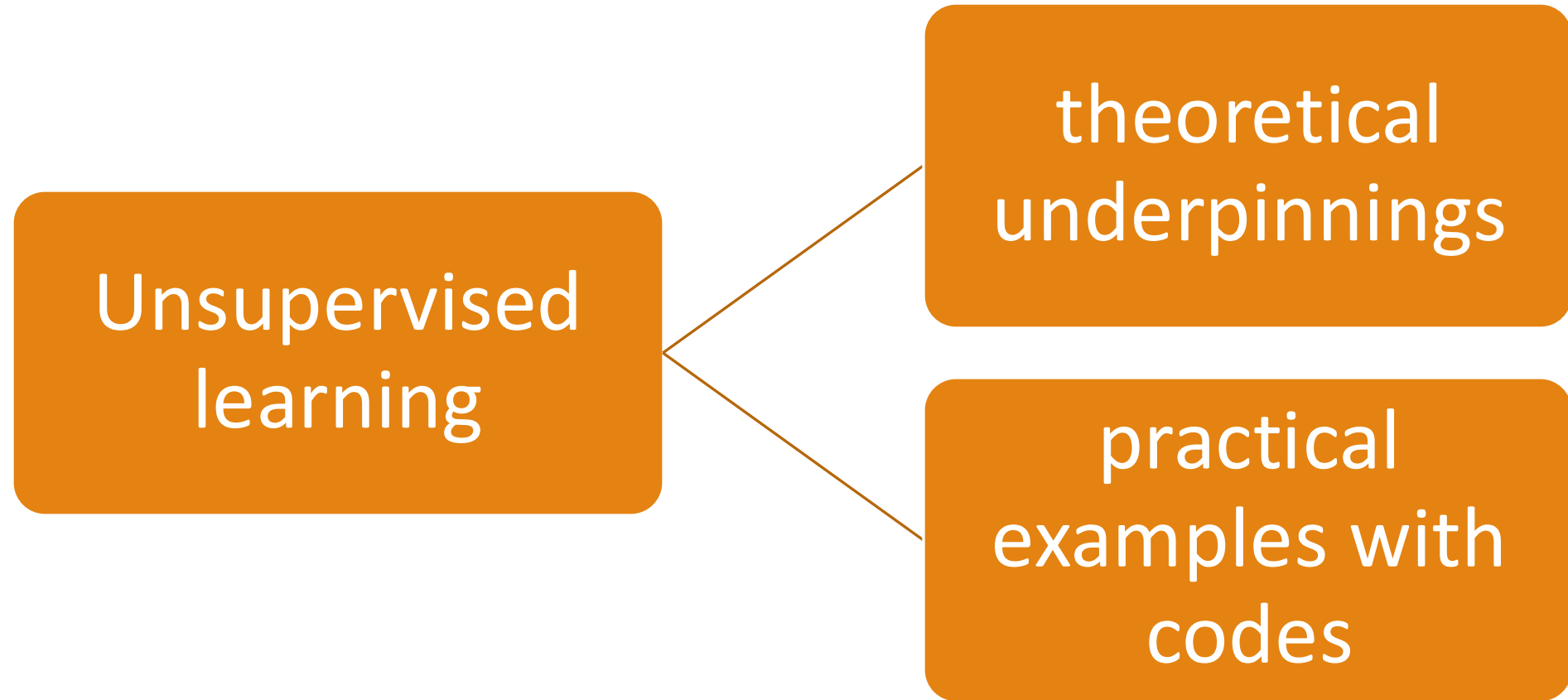
Goals of this course

▪The module is aimed at providing you knowledge about unsupervised learning methods with a special focus on economic problems.

▪Materials include both theoretical underpinnings of machine learning algorithms and their practical applications.

▪After completing this module you are going to be able to apply unsupervised machine learning algorithms in business, social and research problems.

▪A good data scientist is not a quick coder. A MA in DS&BA should be an expert, understanding deeply what is happening in the black box of codes – of course with practical skills in analyses.

▪Data science is like art – you cannot apply in blind any method to any data and be happy, that there is some result.

▪A research university is not a technical school – we are to give you scientific background for data analytics and quantitative methods.

▪Class, except technical skills, is to show you the scientific perspective of analysis. Remember that progress in data science somehow stems from researchers, who perfectly understand the existing methods and add new components and solutions.

Course outline

1. Clustering (classes 1-5)

2. Dimension reduction (classes 6-9)

3. Association rules (basket analysis) (classes 10-13)

Course outline cont.

Unsupervised learning

theoretical underpinnings

practical examples with codes

Rules of assessment

3 individual assignments - projects (100%)

Individual presentation – comment on a paper (obligatory but not evaluated)

Attendance

- Recommended & expected, and obligatory

- You have to be present during the classes, when you are about to give a presentation (see the next slides)

- Your presentation takes place in the group that you are assigned to in the USOS

Projects

- Individual works that should be innovative, show deep understanding and be comprehensive.
    - 3 projects requires (one on clustering, one on dimension reduction and one on association rules; you can merge 2 topics in one project if you wish)
    - For each you may get between 0 and 100 pts

- Ideas
    - R-bloggers style or similar explanation of how to do the things – condition: this cannot exist before
    - Wikipedia note on the given method (in English) – condition: this cannot exist before
    - Typical analytical paper for any dataset with reach analytics – posted on RPubs or similar
    - A research paper which analyses the theoretical features of the methods discussed – posted on RPubs or similar
    - Replication on own data of any good paper on clustering – posted on RPubs or similar

- Papers published online can be a value-added to your professional portfolio – possibly to be included in your CV!
    - Rpubs/nbviewer/GitHub/etc.

- Deadlines for sending/publishing:
    - 01.02.2025 [first attempt]
    - 01.03.2025 [retake session, only if you fail in the first attempt]

Presentation/comment on a paper

- At the beginning of the class

- 10 min max

- See the "Dates of presentations" to check, when you are going to give a presentation
  - Random assignment

- The same file contains information about the paper that you have to comment
  - E.g. "Class 1.4.1" listed next to your student album no means that you will present during the 4th classes on the 1st module (clustering) and you have to refer to the "Class 1.4.1 [...]" article available in the "Clustering papers" folder

- Follow the "Tips on how to discuss a paper"

Homeworks

- Facultative (not obligatory)

- They they may increase your grade by max. 0.5 (e.g. from 4 to 4+) (10% in the grade scale – see the next slide)

- The goal of homework is, in general, to give you challenges to practice the topics

Grade scale (applied to single projects) – the final grade is the average of your points collected from the projects

Range (% of points)	grade	meaning	meaning
<0;50)	2	Negative evaluation	serious mistakes, no submission, very narrow scope
<50;60)	3	Acceptable	basic application of the USL tools (way less than in class)
<60;70)	3+	Almost good	simple application of the USL tools (less than in class)
<70;80)	4	Good	fine application of the USL tools
<80;90)	4+	Almost very good	proper application of the USL tools (interesting topic/data)
<90;95)	5	Very good	very good application of tools introduced during the classes
<95;100>	5!	Excellent!	extra things, beyond the classes, were used

Few words about me

About me

- Contact details
  - jlewkowicz@wne.uw.edu.pl
  - jaceklewkowicz.com

- Experience
  - academia
  - central bank

Consultations

- online

- please email me first and then we will schedule an appointment

Useful links

Useful links

- https://rpubs.com/williamsurles/310847
  - USL in R

- http://www.mit.edu/~9.54/fall14/slides/Class13.pdf
  - External lecture on USL

- https://lgatto.github.io/IntroMachineLearningWithR/unsupervised-learning.html
  - Quick starter on USL

Thank you!