Lab 4 - Advanced Machine Learning

Matrix factorization

Pierre HOUDOUIN

CentraleSupélec

20 Octobre 2023

General information

- **Assignment** : Alone or in pairs, you will code the algorithms you learnt in Ścikit-learn formalism'
- **Due** : Each assignment has to be sent at most 7 days after the lab session at pierre.houdouin@centralesupelec.fr
- **Grading** : There are 5 lab sessions, each lab session is worth 4 points, the average will count for half of your final grade
- **Questions** : If you have questions, comments or feedbacks about the lab session, feel free to contact me by email

Lesson recap

Pros and cons

Each observed vector $x \in \mathbb{R}^p$ is embedded as :

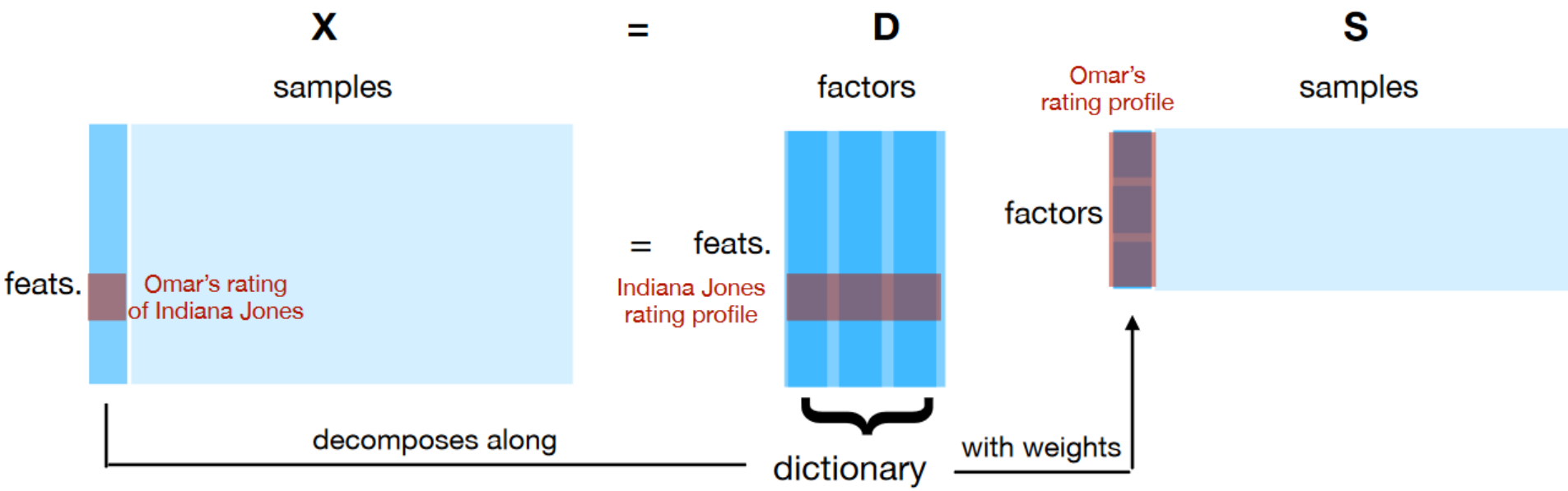
- a source vector $h \in \mathbb{R}^r$
- via a linear map W

Pros and cons

The linear map can be interpreted:

- The columns form a dictionary for x
- x is the result of a weighted (by H) sum of the elements of this dictionary
- Each row provides a feature embedding

Lesson recap



Application	Recommender	Vision	Video
features	movie ratings	pixels	a timeseries
factor	movie genre rating	image template	timeseries template
sample	user	image	recording

Algorithms

Loss

$$\mathcal{L}(W, H) = \frac{1}{2}||X - WH||_F^2 + \frac{\mu}{2}||H||_F^2 + \lambda||H||_1 + \frac{\nu}{2}||W||_F^2$$

Update at each iteration

$$W \leftarrow W \circ \frac{XH^T}{W(HH^T + \nu I_r)}$$

$$H \leftarrow H \circ \frac{W^T X - \lambda \mathbf{1}_{r \times n}}{(W^T W + \mu I_r)H}$$

Assignment plan

- **Part 1** : Implement your own NMF with multiplicative updates
- **Part 2** : Application to vision
- **Part 3** : Application to text