Introduction to Data Science

GIRI NARASIMHAN, SCIS, FIU

Data as Matrices

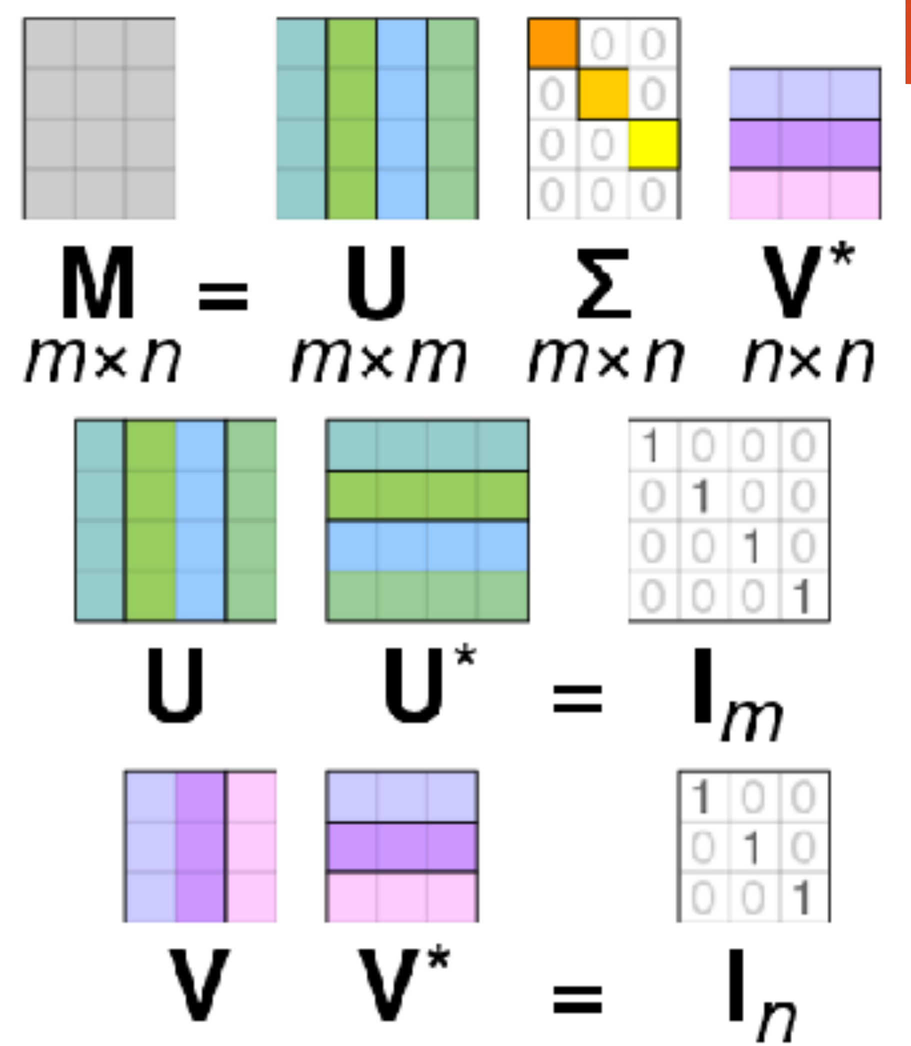
$$\underset{(n\times p)}{\mathbf{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix} \begin{matrix} \leftarrow \text{1st (multivariate) observation} \\ \\ \\ \leftarrow n\text{th (multivariate) observation} \end{matrix}$$

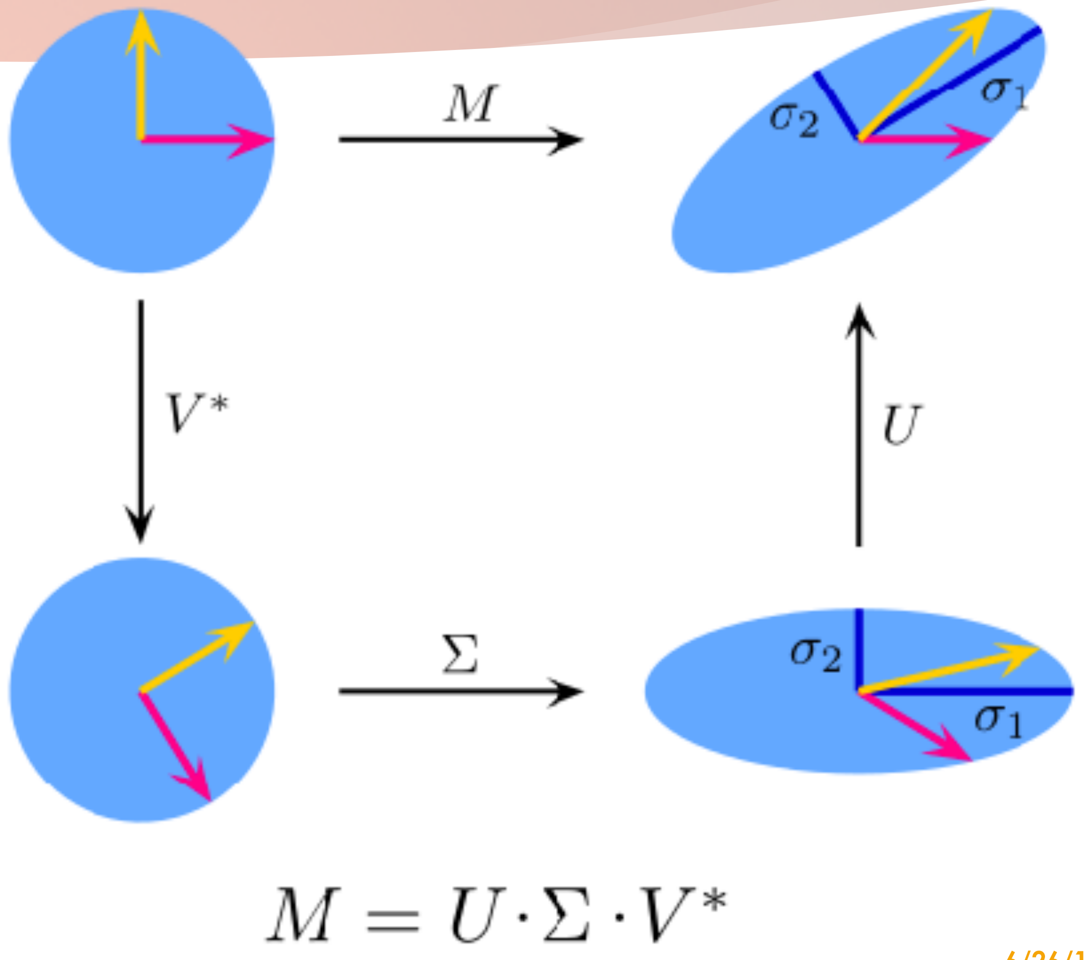
Singular Value Decomposition

- U and V are unitary
  - $UU^* = I$
- diagonal weight matrix

$$\underset{(m\times k)}{\mathbf{A}} = \underset{(m\times m)}{\mathbf{U}}\ \underset{(m\times k)}{\mathbf{\Lambda}}\ \underset{(k\times k)}{\mathbf{V}'}$$

SVD: Rotation-Scaling-Rotation





SVD

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \end{bmatrix}$$

SVD

$$\mathbf{U} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{V}^* = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{bmatrix}$$

The Unitary Matrices

$$\mathbf{U}\mathbf{U}^* = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{I}_4$$

$$\mathbf{V}\mathbf{V}^* = \begin{bmatrix} 0 & 0 & \sqrt{0.2} & 0 & -\sqrt{0.8} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \sqrt{0.8} & 0 & \sqrt{0.2} \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{I}_5$$

SVD Approximations

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \end{bmatrix}$$

Approximations

- M' = [0 0 0 0 0]
  [0 0 3 0 0]
  [0 0 0 0 0]
  [0 2 0 0 0]

Approximations

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \end{bmatrix}$$

- M’ = [**0** 0 0 0 **0**]
  [0 0 3 0 0]
  [0 0 0 0 0]
  [0 2 0 0 0]

Approximations

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \end{bmatrix}$$

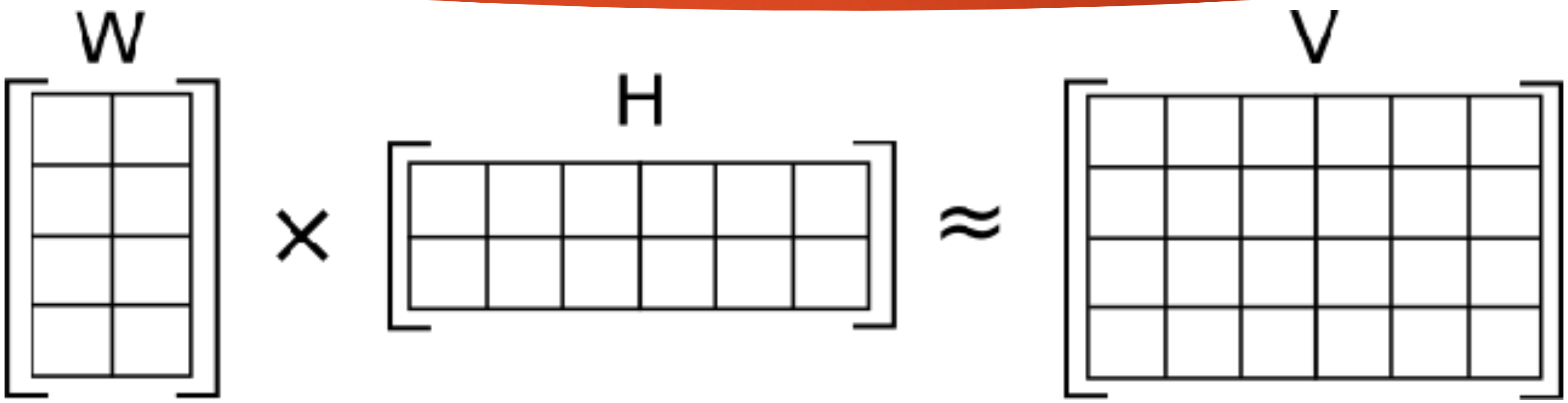
- M'' = [**1** 0 0 0 **2**]
  [0 0 3 0 0]
  [0 0 0 0 0]
  [0 2 0 0 0]

Non-negative matrix factorization

Facial Recognition Problem

- Work of Lee and Seung (1999)
- Database of 2429 faces (19 X 19 pixels)
- Want to learn **eigenfaces**
  - Basis faces for all faces
  - All faces are linear combinations of basis faces

Matrix Factorization Methods

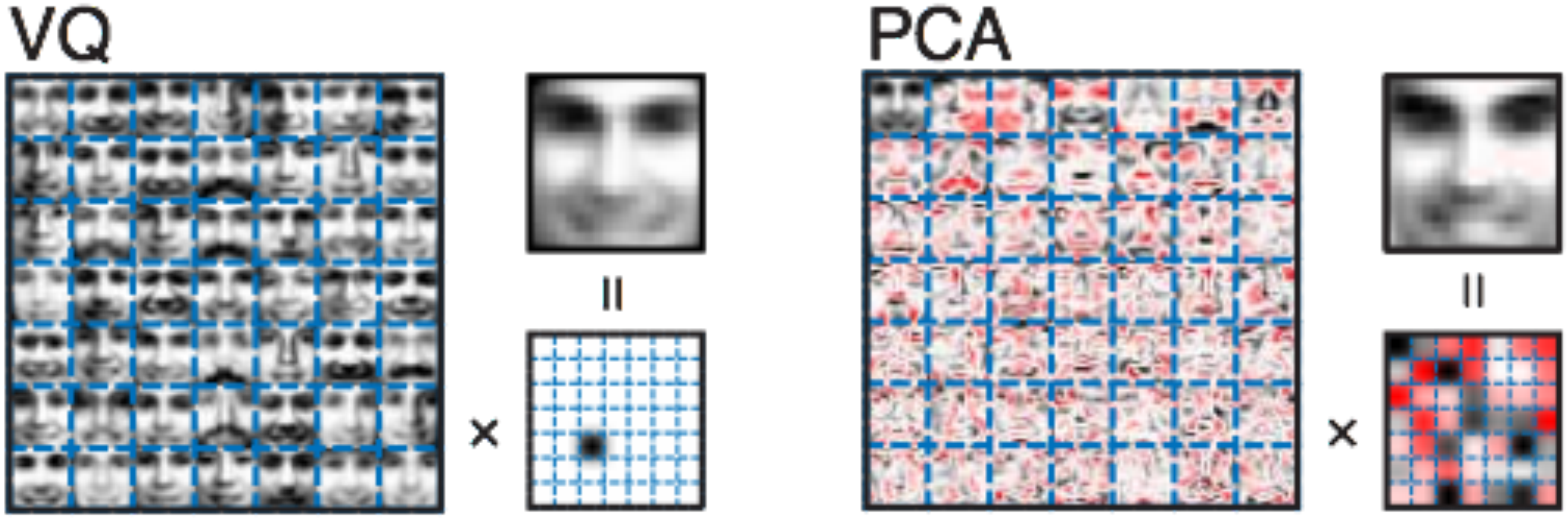


- Matrix Factorization Techniques
  - PCA
  - VQ
  - NMF

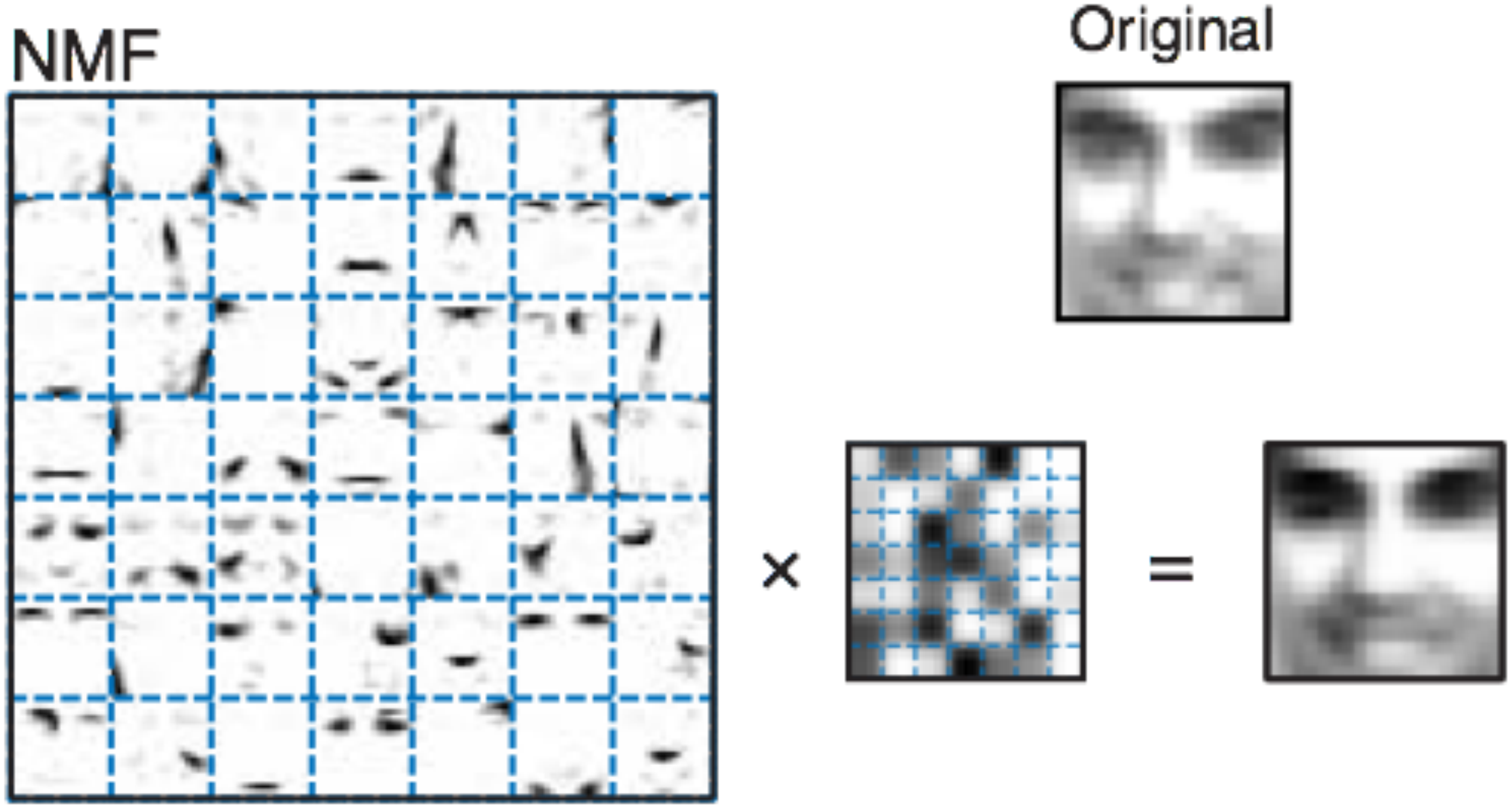
PCA, VQ: Allows negative weights

NMF: Allow only non-negative weights

VQ and PCA



NMF



Objective Function & Update Functions

- Minimize: ?

$$W_{ia} \leftarrow W_{ia} \sum_{\mu} \frac{V_{i\mu}}{(WH)_{i\mu}} H_{a\mu}$$

$$W_{ia} \leftarrow \frac{W_{ia}}{\sum_{j} W_{ja}}$$

$$H_{a\mu} \leftarrow H_{a\mu} \sum_{i} W_{ia} \frac{V_{i\mu}}{(WH)_{i\mu}}$$

Works well for large DB

Sparsity

- NMF basis and encodings are **sparse** & contain large number of vanishing coefficients
  - Not true for VQ and PCA
- Basis images are **non-global**

Related Work

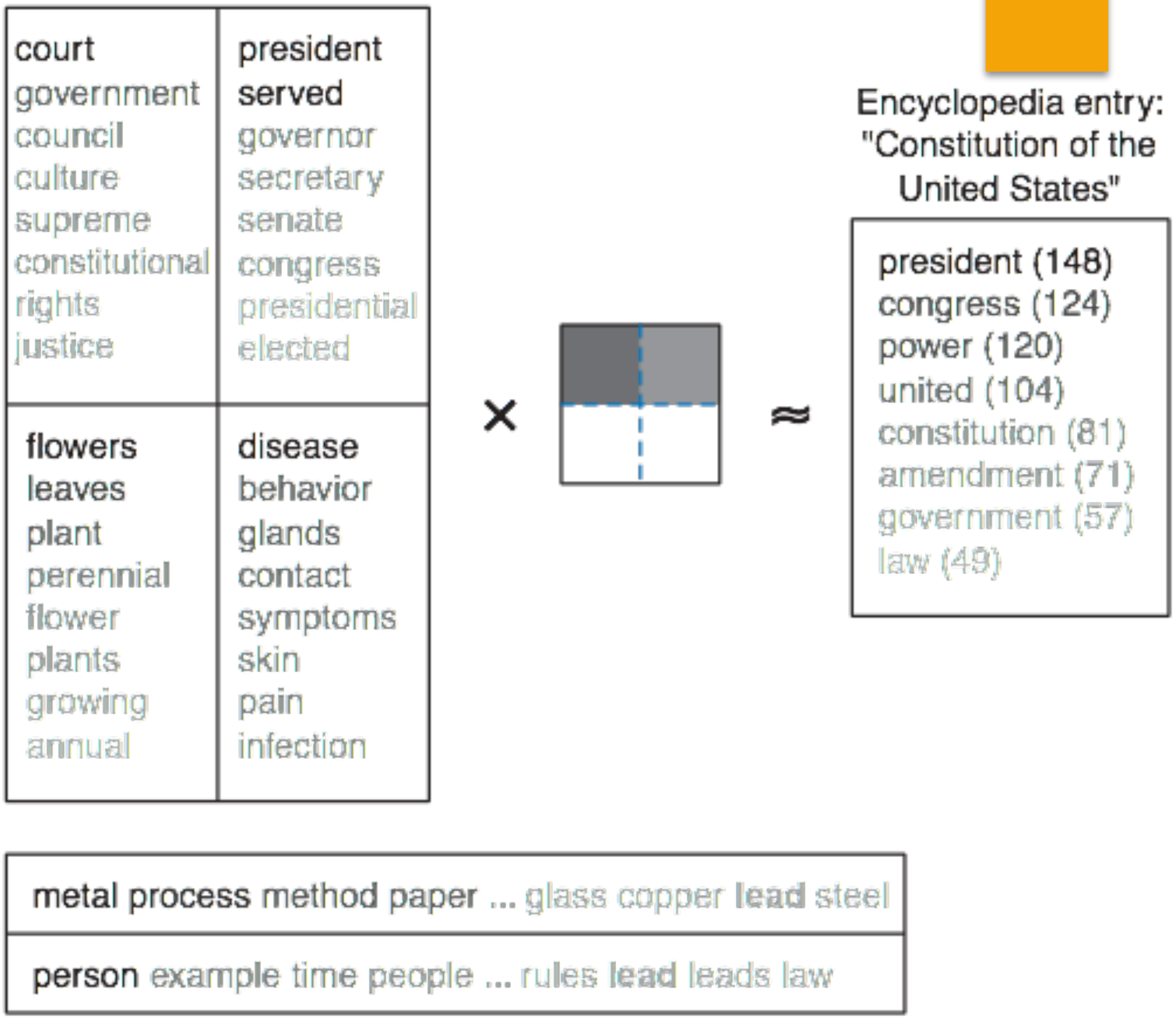
- **Nonnegative Rank** (Gregory and Pullman, '83)
  - Survey applns (Cohen & Rothblum, '93)
- **Approx. Factorization** (Paatero & Tapper, '94)
- **Images** (Lee & Seung '99, *Nature*, **401** (6755))
- **Text Mining**: **pLSI** (Hofmann, SIGIR '99)
- **Latent Dirichlet allocation** (**LDA**) (Blei, Ng, Jordan, JMLR '03)
- **Algorithms** (Lee & Seung NIPS '00)

Applications

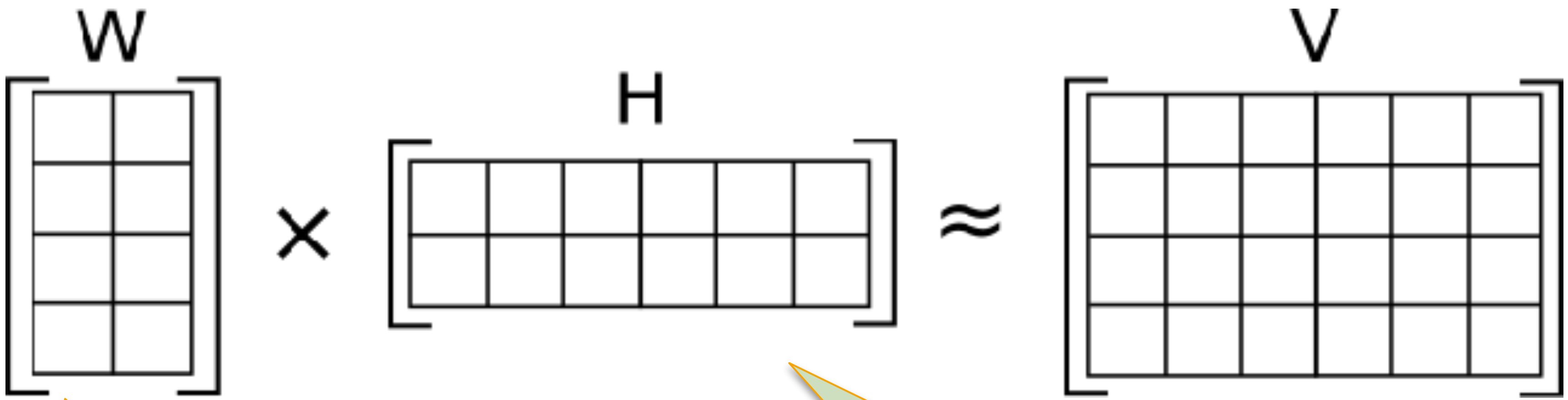
Clustering

- When $V = HH^T$ (and HSH^T), we get
  - K-means and Laplacian-based spectral clustering (and their weighted versions)
- When V represents bipartite graphs
  - Simultaneous row & column clustering
- C. Ding, X. He, H.D. Simon (2005). "On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering". Proc. SIAM Int'l Conf. Data Mining, pp. 606-10. '05

Grolier encyclopedia – 30991 articles, vocabulary 15276 words



NMF Applications & Interpretation

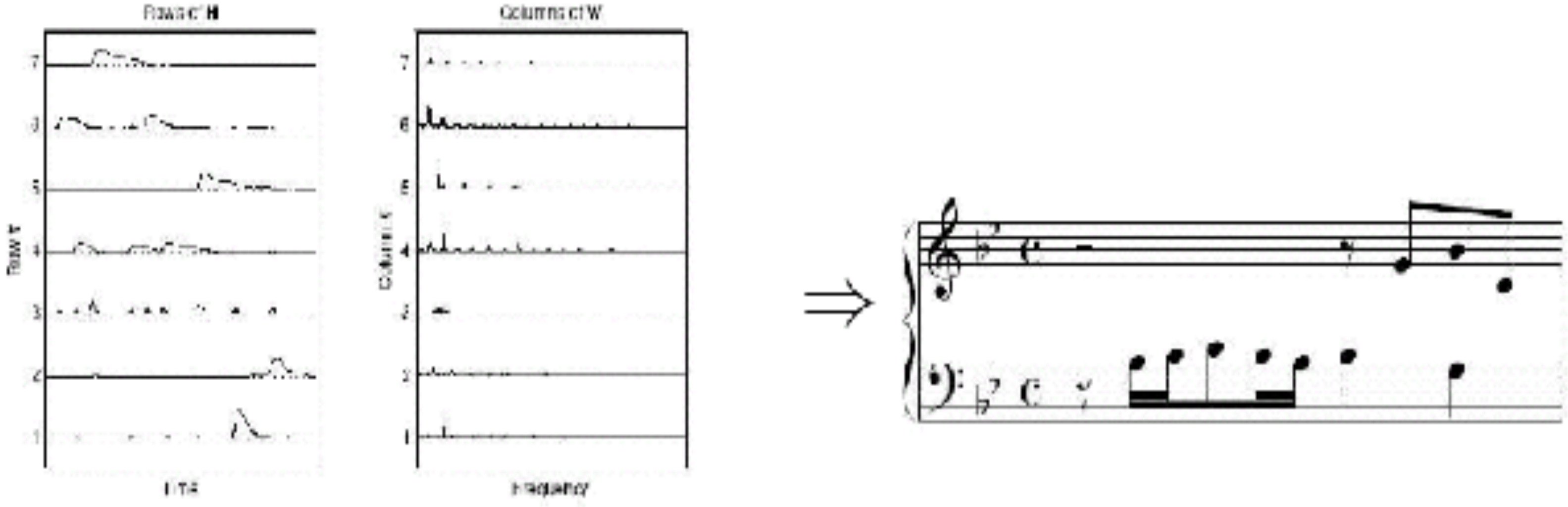


Columns describe docs in terms of topics

Rows describe topics in terms of observed words

Audio Signal Processing

- Smaragdis and Brown ’03; Smaragdis, Raj, Shashanka, NIPS ‘06



$$X = W\, H$$

Recommender Systems

- Social recommendation in social network service (Ma, Yang, Lyu, King, ICIKM '08.)
- Content-based image tagging in image processing (Ning, Cheung, Guoping, Xiangyang, *IEEE Trans PAMI*, '11),
- QoS prediction in service computing (Wu et al. *IEEE TrSMCS* '13; Zheng, et al., *IEEE TrSC* '13)
- Video re-indexing (Weng et al., *ACM Trans. MCCA*, '12)
- Mobile-user tracking in wireless sensor networks (Pan, et al., *IEEE TPAMI*, '12)

Modeling Latent Factors

- Assume rows of V represent observations or samples and columns represent features
  - V = WH
  - Rows of W represent samples and columns of H represent features
  - Columns of W and rows of H represent latent variables or hidden factors

Supervised NMF

- NMF is an **unsupervised** process
- **Supervised** NMF using co-occurrence info has been studied by Cai, Y., Gu, H., & Kenney, T. (2017). Learning Microbial Community Structures with Supervised and Unsupervised Non-negative Matrix Factorization. *Microbiome*, 5(1), 110.

Machine Learning

Machine Learning

- Unsupervised Learning
  - Clustering
  - PCA
- Supervised Learning
  - SVM
  - DT
  - kNN
  - NN

Data-driven Machine Learning