Non-negative Matrix Factorization and Machine Learning

Giri Narasimhan

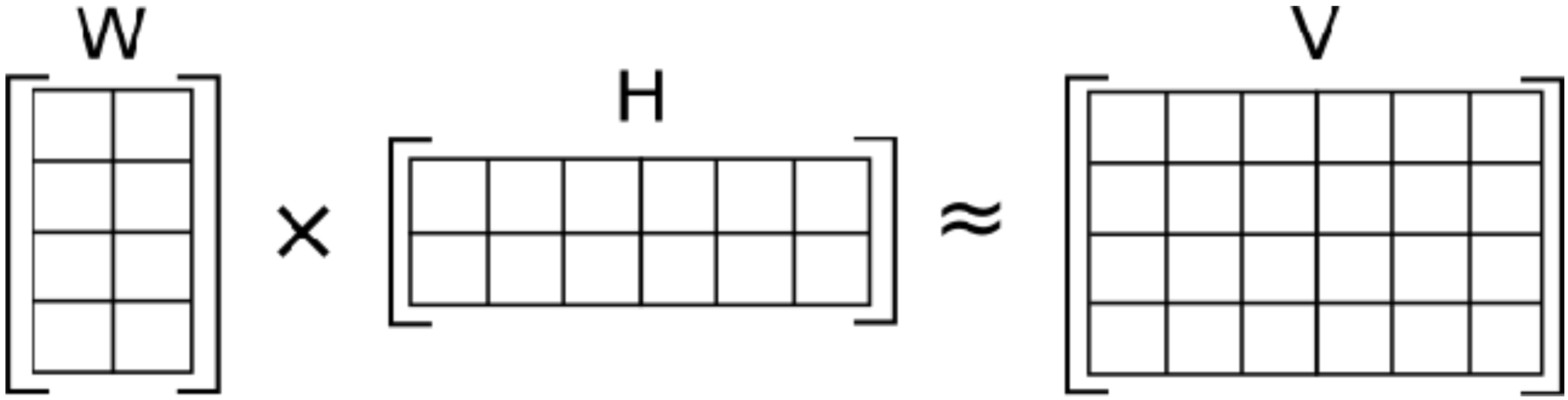
BioRG Lab Meeting

May 2019

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

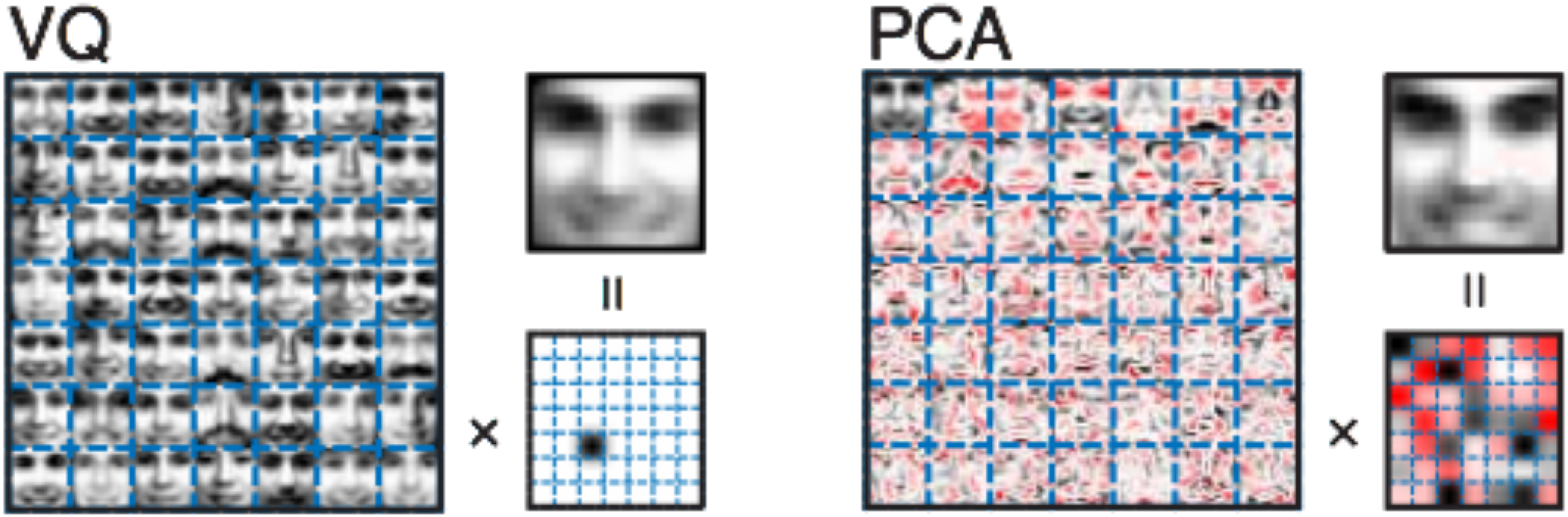
Allows negative weights

Allow only non-negative weights

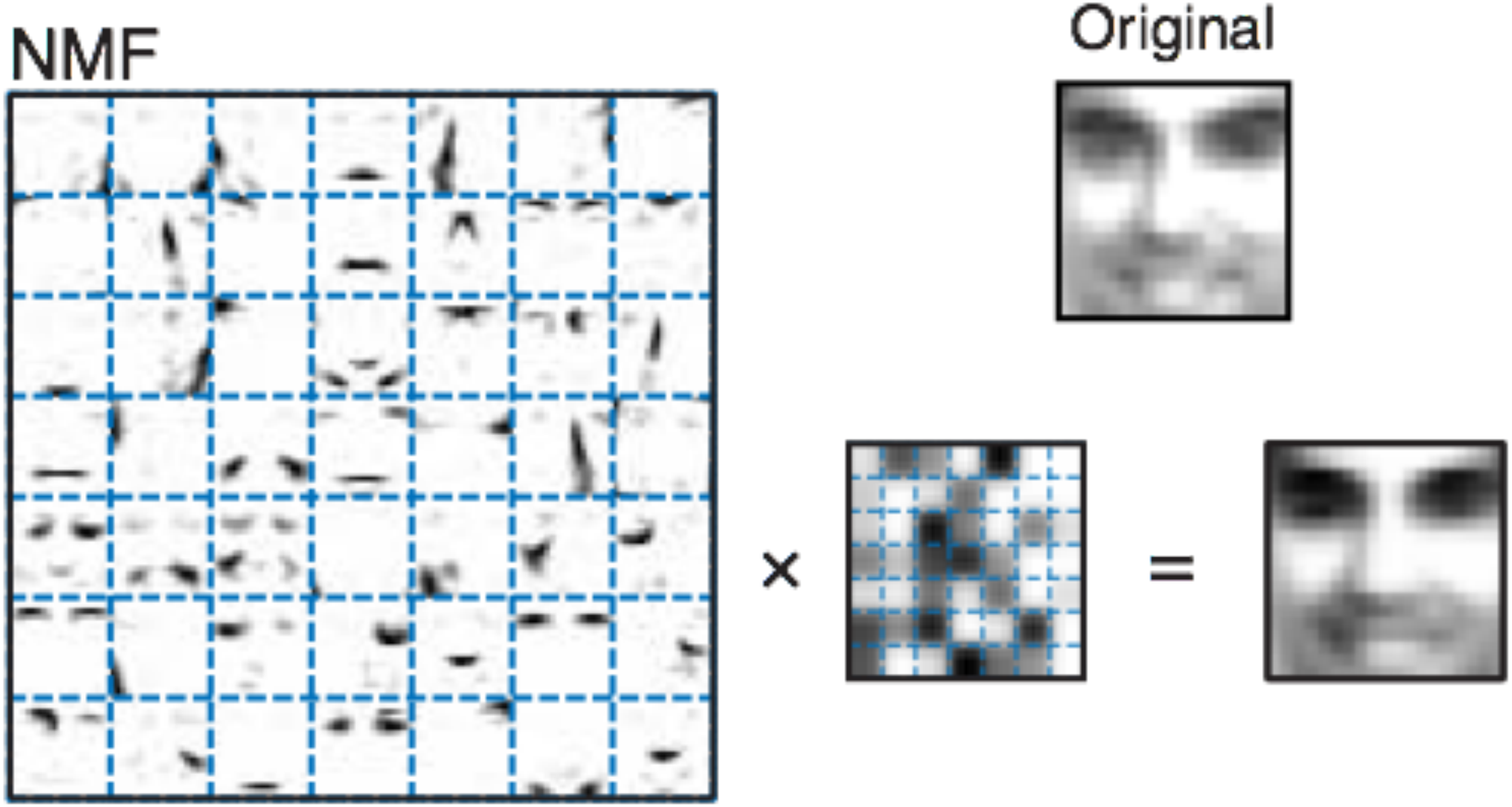
VQ vs PCA vs NMF

- In VQ, each column is a **unary** vector
  - Chooses prototypical faces
  - Each face approximated by nearest prototype
- PCA – **orthonormal** bases
  - Each face is linear combination of basis faces
  - Eigenface – largest variance, but not intuitive
- Only positive weights in NMF
  - Each face is **additive combination** of basis faces

VQ and PCA



NMF



Objective Function & Update Functions

- Minimize:

?

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

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

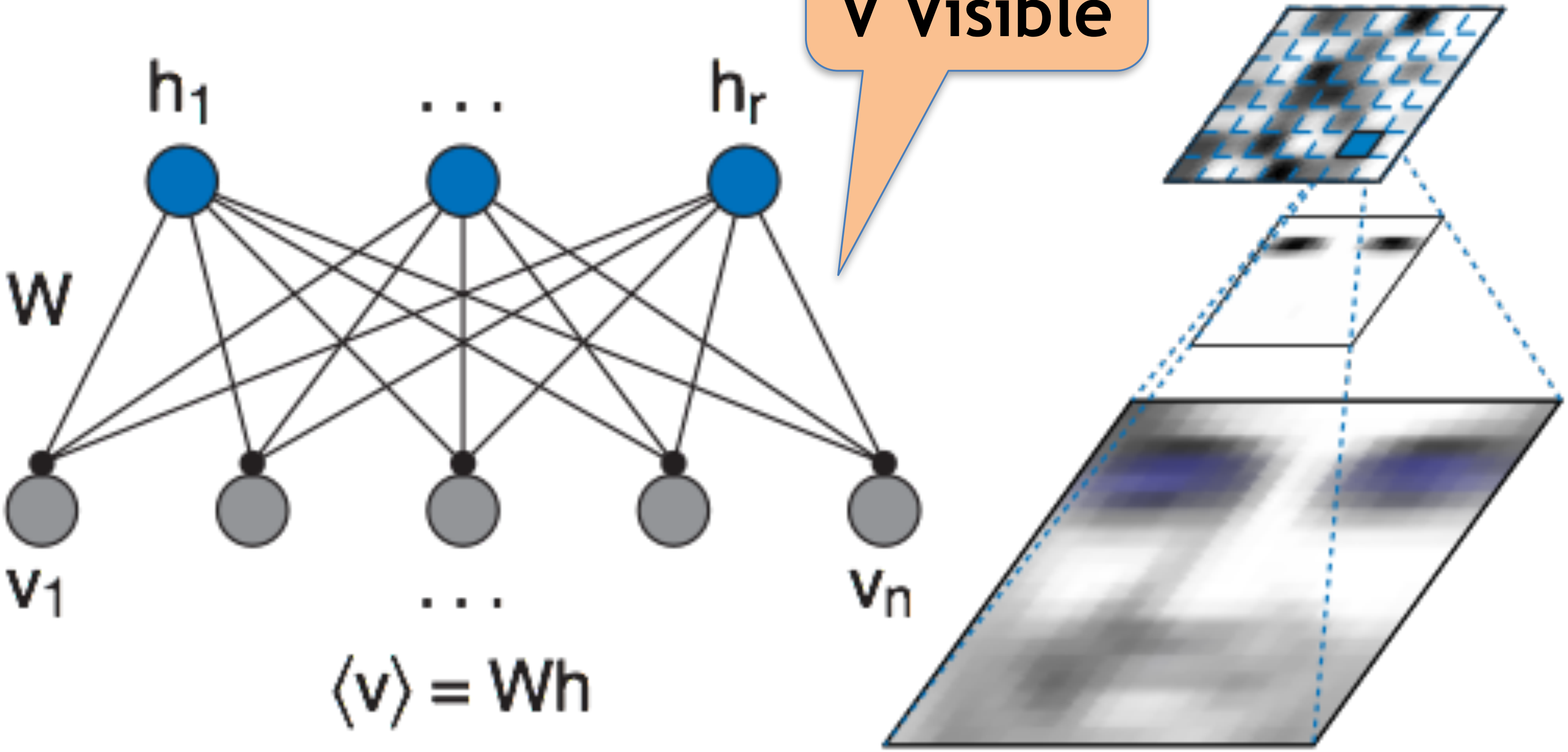
$$H_{a\mu} \leftarrow H_{a\mu} \sum_{i} W_{ia} \frac{V_{i\mu}}{(WH)_{i\mu}}$$

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**

Visualizing the Dependencies



H Hidden
V Visible

Columns of H to Columns of V

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)

NMF Features

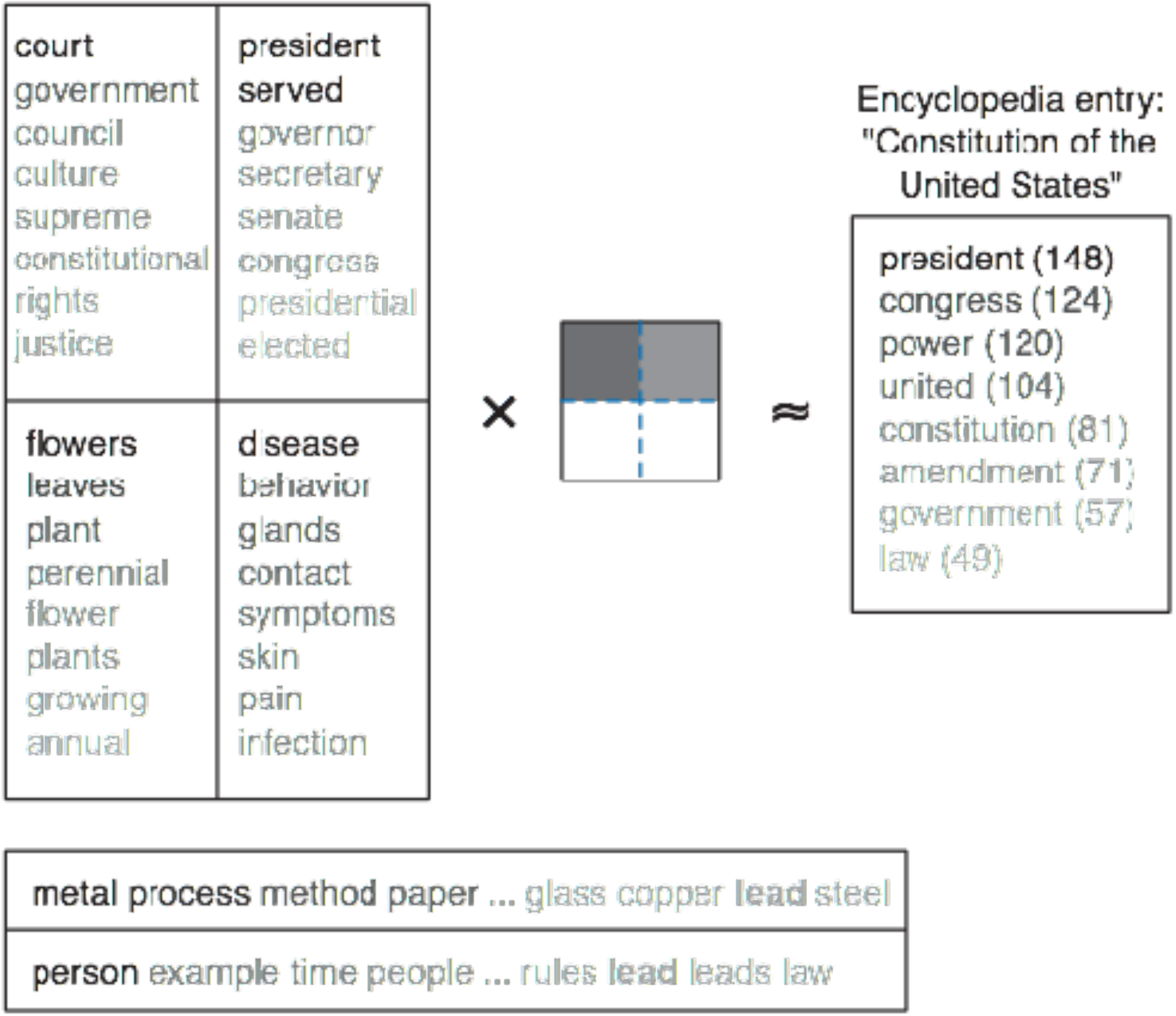
- Models **dependent** hidden variables
  - If independent, use ICA
- NMF is not good for images of objects viewed from very different viewpoints or highly articulated objects
  - Need hierarchical models w/ multiple levels of hidden variables

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



court
government
council
culture
supreme
constitutional
rights
justice

president
served
governor
secretary
senate
congress
presidential
elected

flowers
leaves
plant
perennial
flower
plants
growing
annual

disease
behavior
glands
contact
symptoms
skin
pain
infection

×

≈

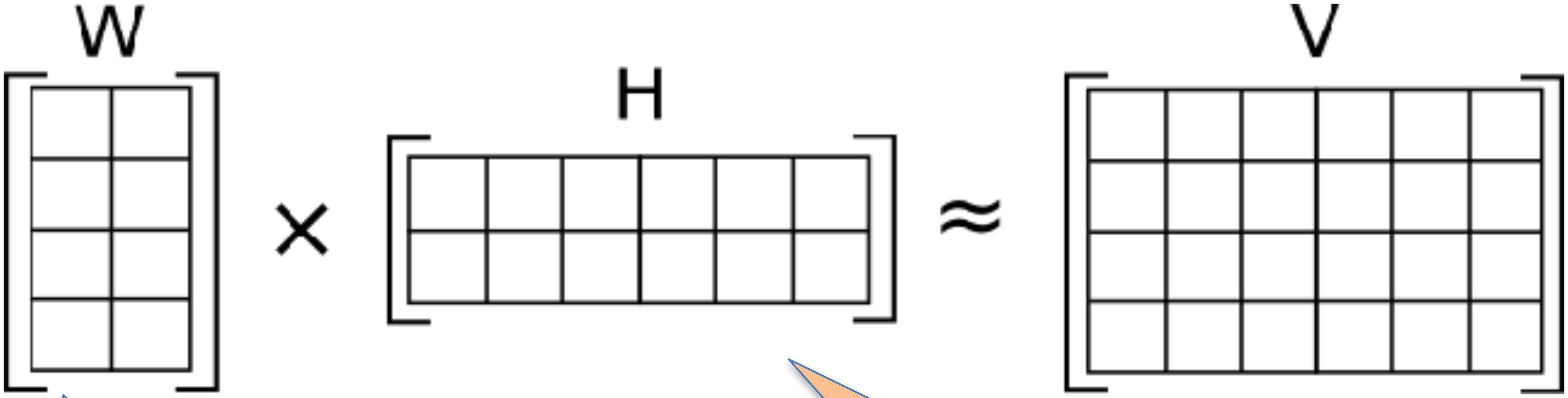
Encyclopedia entry:
"Constitution of the United States"

president (148)
congress (124)
power (120)
united (104)
constitution (81)
amendment (71)
government (57)
law (49)

metal process method paper ... glass copper **lead** steel

person example time people ... rules **lead** leads law

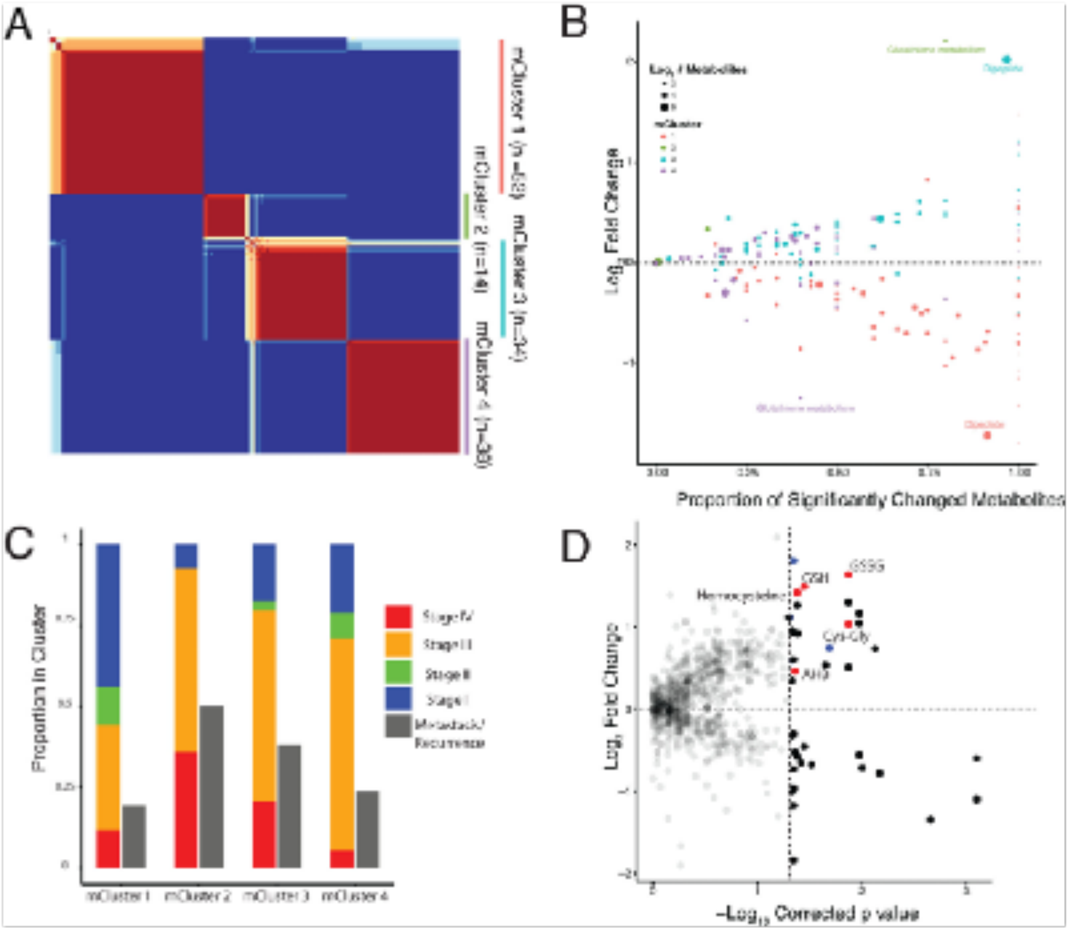
NMF Applications & Interpretation



Columns describe docs in terms of topics

Rows describe topics in terms of observed words

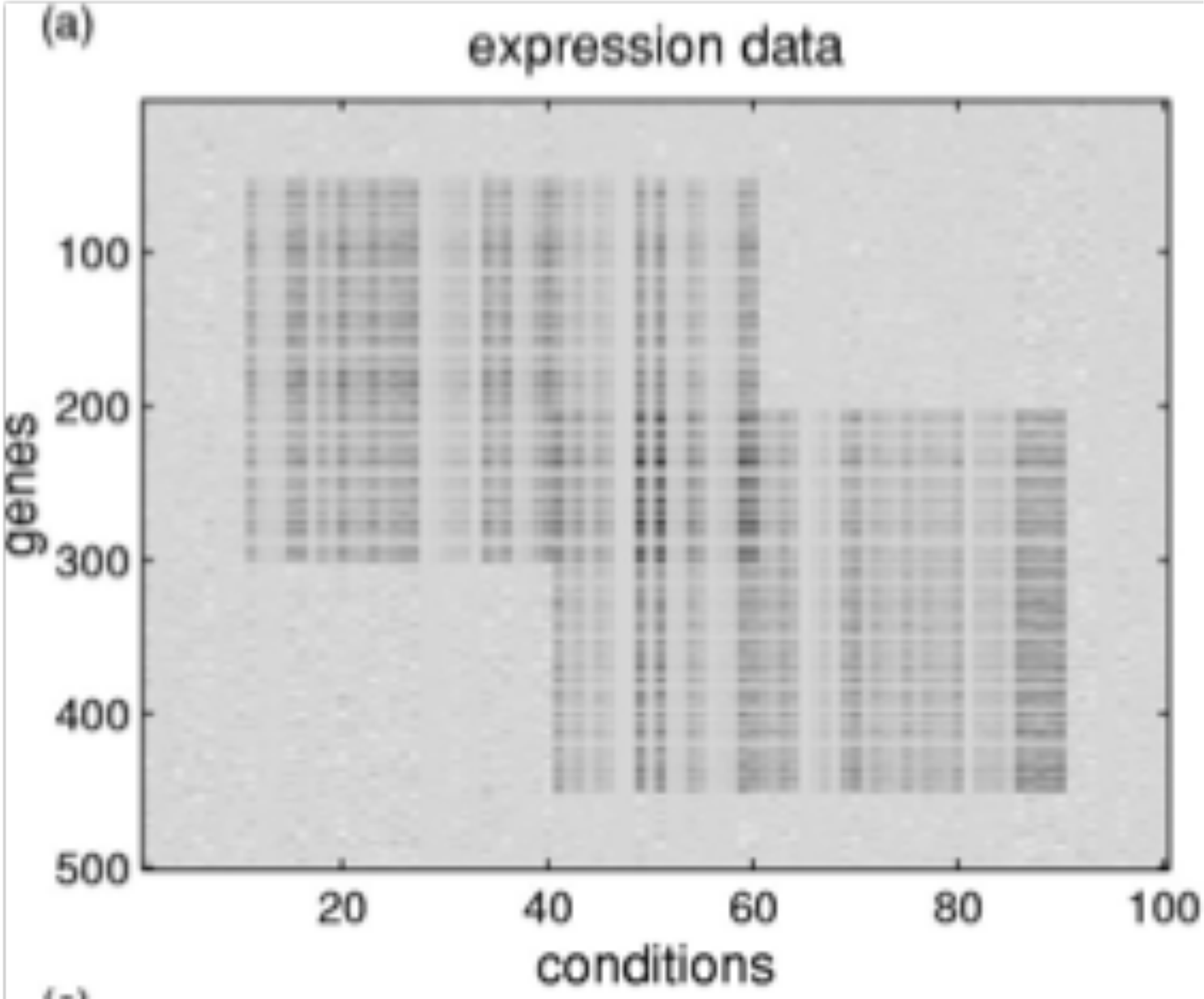
Renal Cell Carcinoma: Hakimi *et al.*, *Cancer Cell*. 2016 29(1): 104



Renal Cell Carcinoma: Hakimi *et al.*, *Cancer Cell*. 2016 29(1): 104

- **Unsupervised clustering of ccRCC based on metabolite signatures** (A) NMF clustering of metabolomics data. (B) Metabolites significantly increased or decreased in each cluster, relative to all other tumors. X-axis indicates, for a given cluster of patient samples, the proportion of metabolites in a pathway that are significantly changed (both increased and decreased) in a cluster. Y-axis plots the average log2 fold change of these metabolites. (C) Clinical stages and the eventual metastasis of each individual metabolic cluster. (D) Comparison of metabolite abundances in tumors developing metastases versus those not developing metastases.

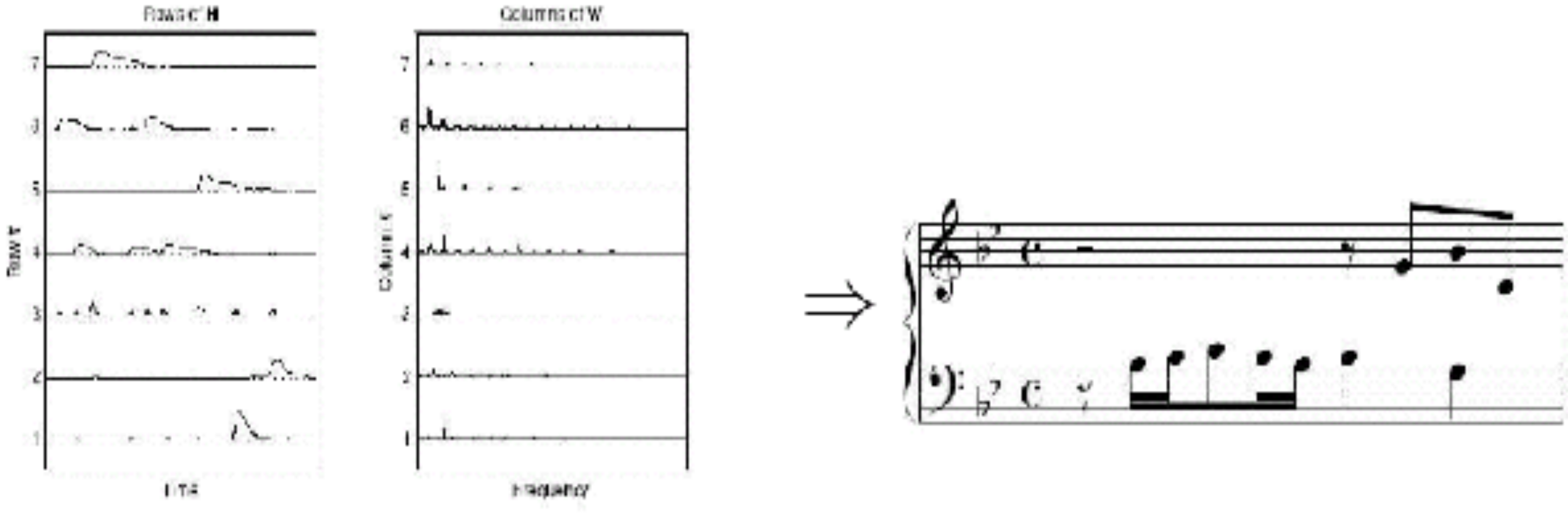
Finding Modules from expr data



Bergman, Ihmels, Barkai, 2003

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)

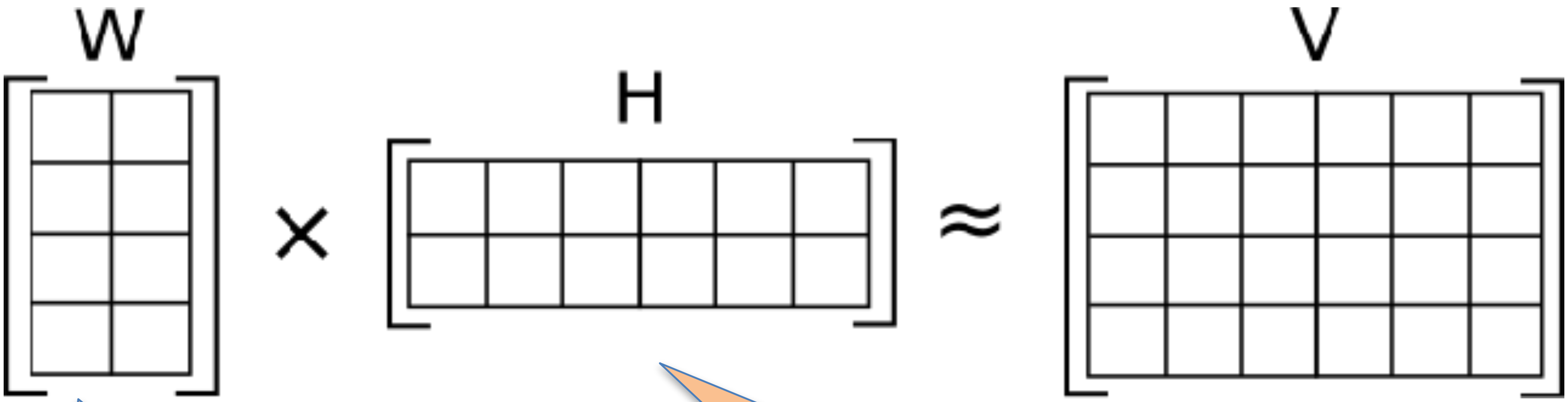
Modeling Latent Factors

- 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

Gene Expression data

- Latent factors could be
  - Env. variables acting on specific sets of genes
  - Clusters of genes acting together where each cluster can be thought of as being associated with a function or a pathway

NMF Applications & Interpretation



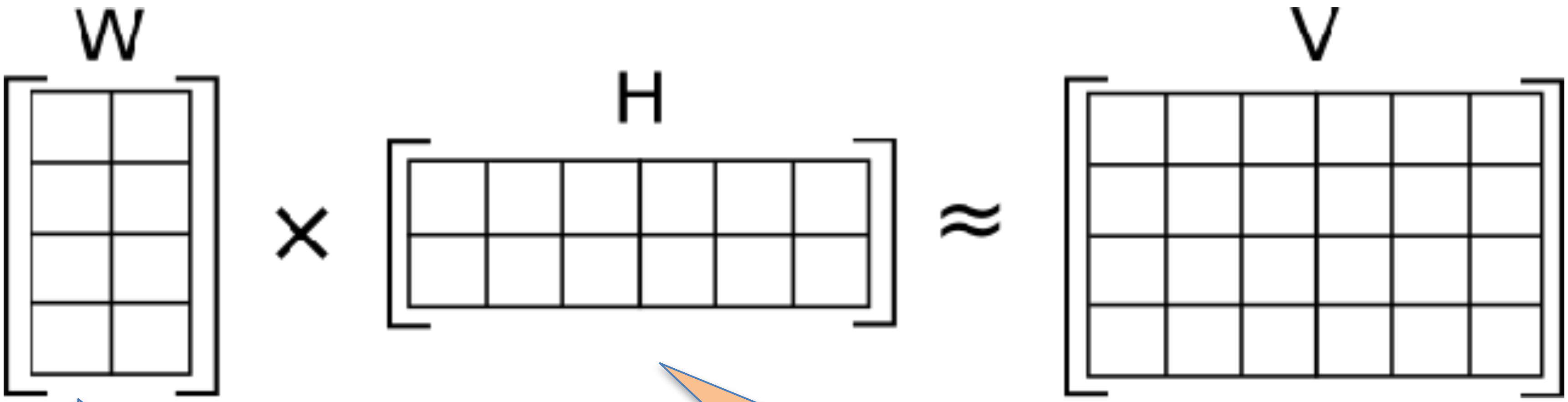
Columns describe samples in terms of functional group of genes

Rows describe groups of genes in terms of gene expression

Microbiome Data

- Each sample is composed of communities
- Each community is composed of microbes, possibly associated with a function
- Latent factors are communities
- Factorization will tell you the composition of the communities

NMF Applications & Interpretation



Columns describe samples in terms of functional communities

Rows describe communities in terms of observed taxa

Multiomics Datasets

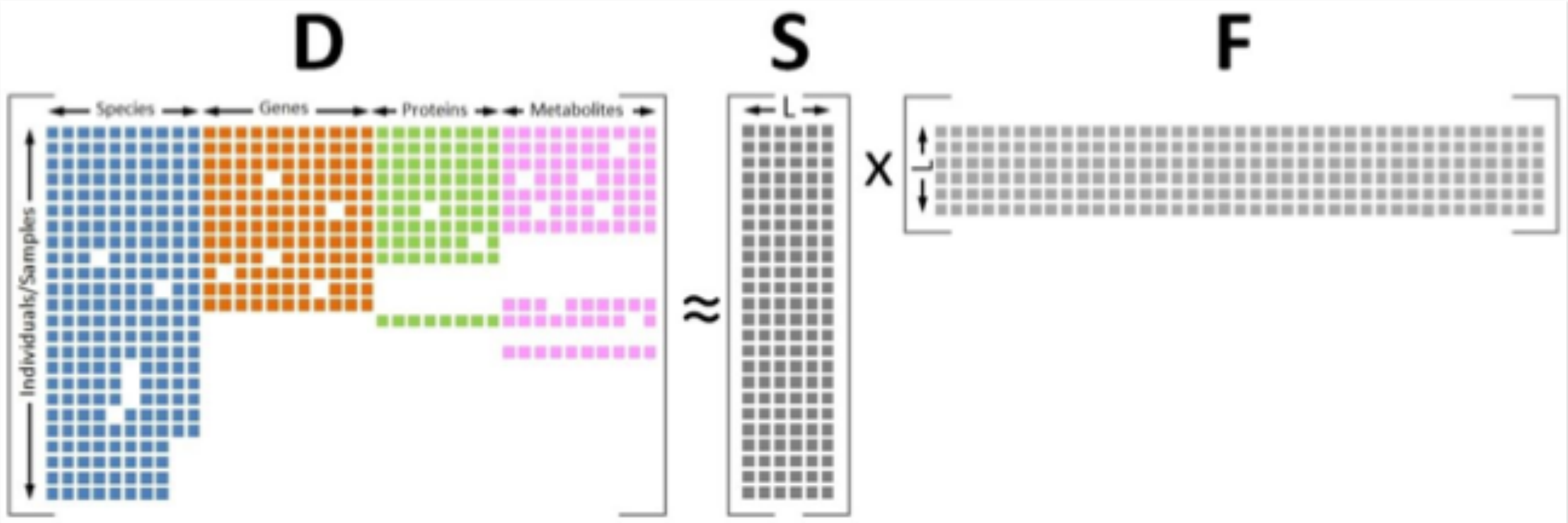


Figure 5. Matrix factorization of multi-omic microbiome data with missing assays and measurements. The multi-omic data matrix D of size N×M, is decomposed into two smaller matrices, S and F, such that D≈SF.

Non-linear relationships

- Matrix multiplication in NMF reflects linear relationships between the measured quantities and latent factors
- What if the relationships were **non-linear**?
  - Matrices could be modified, or
  - Use NNs or even Deep NNs
    - replace dot product of i^{th} row of S and j^{th} column of F with NN (input: concat. of vectors; output: answer)

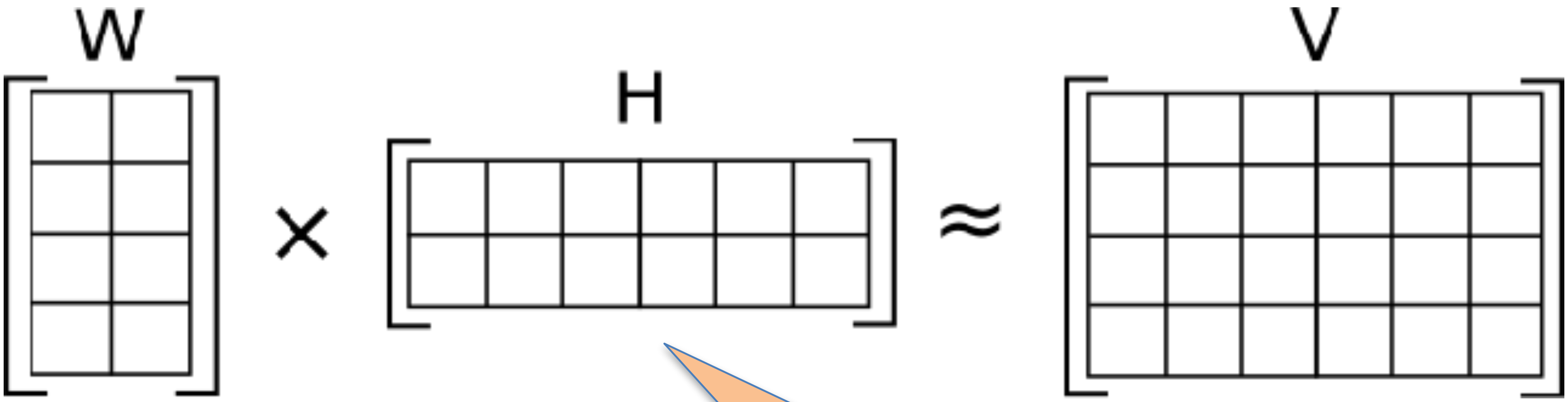
Questions to pursue …

- Matrix of correlation values, i.e., CoNets?
- Matrix of conditional probabilities in BN?
- Matrix of conditional probabilities in DBN?
- Can NMF be adapted for online ML?
    - Periodically estimate hyperparameters
- How to connect to disease?
- How to model growth in microbiome?
- How to compute NMF in the cloud?
- How to model forensics problem using NMF?

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.

NMF Applications & Interpretation



Samples in terms of functional communities

Communities in terms of observed taxa

Secondary factorization: W = Y X P
[Balsamico]
Latent factors: contribution to/from