Introduction to Data Science

GIRI NARASIMHAN, SCIS, FIU
Data as Matrices

\[
\mathbf{X}^{(n \times p)} = \begin{bmatrix}
  x_{11} & x_{12} & \cdots & x_{1p} \\
  x_{21} & x_{22} & \cdots & x_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{n1} & x_{n2} & \cdots & x_{np}
\end{bmatrix} = \begin{bmatrix}
  \mathbf{x}_1' \\
  \mathbf{x}_2' \\
  \vdots \\
  \mathbf{x}_n'
\end{bmatrix}
\]

\[
\leftarrow 1st \text{ (multivariate) observation}
\]

\[
\leftarrow n\text{th} \text{ (multivariate) observation}
\]
Singular Value Decomposition

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

$$A = U \Lambda V'$$

$(m \times k) \quad (m \times m)(m \times k)(k \times k)$
SVD: Rotation-Scaling-Rotation

\[ M = \begin{bmatrix} U & \Sigma & V^* \end{bmatrix} \]

\[ U U^* = I_m \]

\[ V V^* = I_n \]
SVD

\[ M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \end{bmatrix} \]
SVD

\[
U = \begin{bmatrix}
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & -1 \\
1 & 0 & 0 & 0
\end{bmatrix}
\]

\[
\Sigma = \begin{bmatrix}
2 & 0 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 & 0 \\
0 & 0 & \sqrt{5} & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

\[
V^* = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
\sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\
0 & 0 & 0 & 1 & 0 \\
-\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2}
\end{bmatrix}
\]
The Unitary Matrices

\[ UU^* = \begin{bmatrix}
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & -1 \\
1 & 0 & 0 & 0
\end{bmatrix} \cdot \begin{bmatrix}
0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & -1 & 0
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} = I_4 \]

\[ VV^* = \begin{bmatrix}
0 & 0 & \sqrt{0.2} & 0 & -\sqrt{0.8} \\
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & \sqrt{0.8} & 0 & \sqrt{0.2}
\end{bmatrix} \cdot \begin{bmatrix}
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
\sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\
0 & 0 & 0 & 1 & 0 \\
-\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2}
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix} = I_5 \]
SVD Approximations

\[
M = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & 3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0
\end{bmatrix}
\]
Approximations

\[ \mathbf{M}' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \]
Approximations

\[
M = \begin{bmatrix}
1 & 0 & 0 & 0 & 2 \\
0 & 0 & 3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0
\end{bmatrix}
\]

\[
M' = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0
\end{bmatrix}
\]
Approximations

\[ M'' = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \]
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  Allows negative weights
  - VQ
  - NMF  Allow only non-negative weights
VQ and PCA
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
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

Grolier encyclopedia – 30991 articles, vocabulary 15276 words

<table>
<thead>
<tr>
<th>court</th>
<th>president</th>
</tr>
</thead>
<tbody>
<tr>
<td>government</td>
<td>served</td>
</tr>
<tr>
<td>council</td>
<td>governor</td>
</tr>
<tr>
<td>culture</td>
<td>secretary</td>
</tr>
<tr>
<td>supreme</td>
<td>senate</td>
</tr>
<tr>
<td>constitutional</td>
<td>congress</td>
</tr>
<tr>
<td>rights</td>
<td>presidential</td>
</tr>
<tr>
<td>justice</td>
<td>elected</td>
</tr>
</tbody>
</table>

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
Audio Signal Processing

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

\[ X = WH \]
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*)
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
Machine Learning
Machine Learning

- Unsupervised Learning
  - Clustering
  - PCA

- Supervised Learning
  - SVM
  - DT
  - kNN
  - NN

Data-driven Machine Learning