Topic Models, Methods, and Medicine

by Jamie Haddock
(Harvey Mudd College, Department of Mathematics)
on June 23, 2022,
Stauffer Lecture

supported by NSF DMS #2211318

matH of Algorithms, Data & Decisions research group @ HARVEY MUDD COLLEGE

`https://ieeexplore.ieee.org/document/9022678` (CAMSAP 2019)
joint with Mengdi Gao[•], Denali Molitor[*], Deanna Needell, Eli Sadovnik[*], Tyler Will[•], Runyu Zhang[•]
collaboration with LymeDisease.org

`https://arxiv.org/abs/2010.11365` (ICASSP 2021)
joint with Deanna Needell, Liza Rebrova, Joshua Vendrow[•]

`https://arxiv.org/abs/2010.07956` (ACSSC 2021)
joint with Miju Ahn, Rachel Grotheer, Lara Kassab[*], Alona Kryshchenko, Kathryn Leonard, Sixian Li[•], R. W. M. A. Madushani, Thomas Merkh[*], Deanna Needell, Elena Sizikova, Chuntian Wang

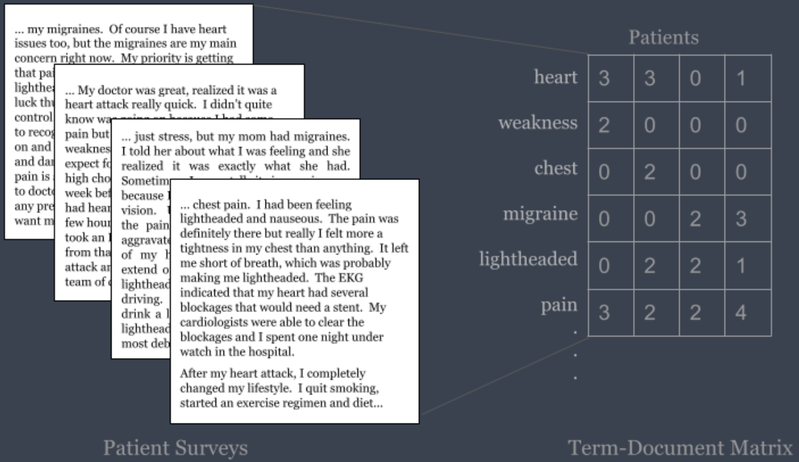
Soon to appear work (ACSSC 2021)
joint with Joshua Vendrow[•], Deanna Needell

`https://arxiv.org/pdf/2109.14820.pdf` (ICASSP 2022)
joint with Joshua Vendrow[•], Deanna Needell

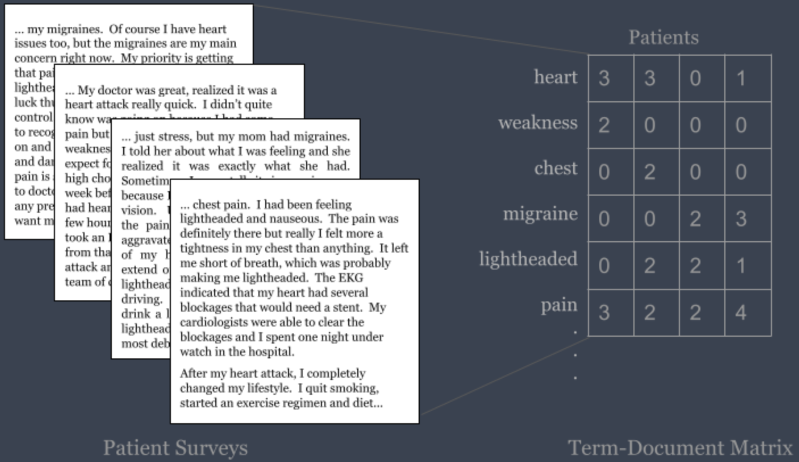
Forthcoming work
joint with Edwin Chau[•], Moisey Alaev[•], Joshua Vendrow[•], Rachel Grotheer, Alona Kryshchenko, Kathryn Leonard, Deanna Needell
collaboration with Harbor-UCLA Medical Center Department of Cardiology

Motivation

» Learn trends in high-dimensional data



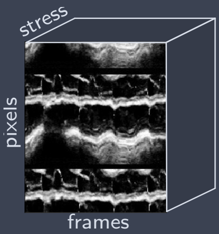
» Learn trends in high-dimensional data



Can we understand symptom trends and shared patient experiences automatically?

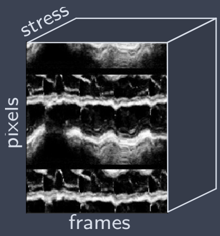
» Learn trends in high-dimensional data

» Learn trends in high-dimensional data



» Learn trends in high-dimensional data

Can we learn cohesive parts and separate noise in medical image studies?



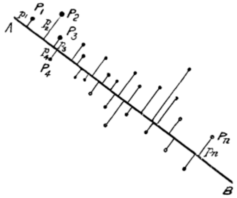
Introduction

» Topic Modeling

- ▷ principal component analysis (PCA)

  [Pearson 1901]

  [Hotelling 1933]



Pearson, K. (1901) *On lines and planes of closest fit to systems of points in space.*

» Topic Modeling

- ▷ principal component analysis (PCA)

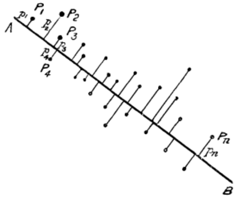
  [Pearson 1901]

  [Hotelling 1933]

- ▷ latent dirichlet allocation (LDA)

  [Pritchard, Stephens, Donnelly 2000]

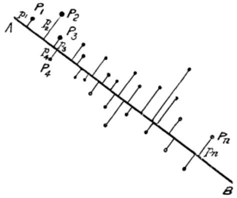
  [Blei, Ng, Jordan 2003]



Pearson, K. (1901) *On lines and planes of closest fit to systems of points in space.*

» Topic Modeling

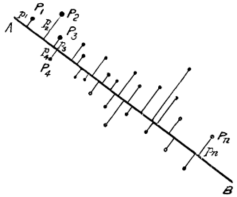
- ▷ principal component analysis (PCA)
  [Pearson 1901]
  [Hotelling 1933]
- ▷ latent dirichlet allocation (LDA)
  [Pritchard, Stephens, Donnelly 2000]
  [Blei, Ng, Jordan 2003]
- ▷ nonnegative matrix factorization (NMF)
  [Paatero, Tapper 1994]
  [Lee, Seung 1999]



Pearson, K. (1901) *On lines and planes of closest fit to systems of points in space.*

» Topic Modeling

- ▷ principal component analysis (PCA)
  [Pearson 1901]
  [Hotelling 1933]
  - ∗ supervised PCA
    [Bair, Hastie, Paul, Tibshirani 2006]

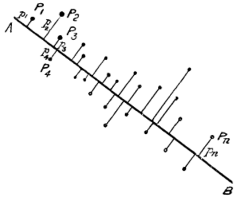


Pearson, K. (1901) *On lines and planes of closest fit to systems of points in space.*

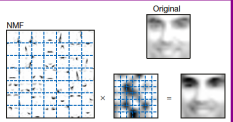
- ▷ latent dirichlet allocation (LDA)
  [Pritchard, Stephens, Donnelly 2000]
  [Blei, Ng, Jordan 2003]
- ▷ nonnegative matrix factorization (NMF)
  [Paatero, Tapper 1994]
  [Lee, Seung 1999]

» Topic Modeling

- ▷ principal component analysis (PCA)
  - [Pearson 1901]
  - [Hotelling 1933]
    - ∗ supervised PCA [Bair, Hastie, Paul, Tibshirani 2006]
- ▷ latent dirichlet allocation (LDA)
  - [Pritchard, Stephens, Donnelly 2000]
  - [Blei, Ng, Jordan 2003]
    - ∗ supervised LDA [Blei, McAuliffe 2008]
- ▷ nonnegative matrix factorization (NMF)
  - [Paatero, Tapper 1994]
  - [Lee, Seung 1999]



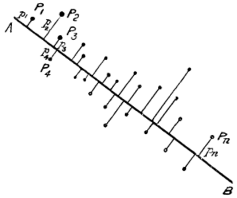
Pearson, K. (1901) *On lines and planes of closest fit to systems of points in space.*



Lee, D., Seung, S. (1999) *Learning the parts of objects by non-negative matrix factorization.*

» Topic Modeling

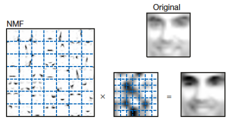
- ▷ principal component analysis (PCA)
  - [Pearson 1901]
  - [Hotelling 1933]
    - ∗ supervised PCA [Bair, Hastie, Paul, Tibshirani 2006]



Pearson, K. (1901) *On lines and planes of closest fit to systems of points in space.*

- ▷ latent dirichlet allocation (LDA)
  - [Pritchard, Stephens, Donnelly 2000]
  - [Blei, Ng, Jordan 2003]
    - ∗ supervised LDA [Blei, McAuliffe 2008]

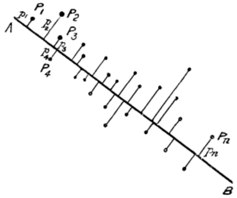
- ▷ nonnegative matrix factorization (NMF)
  - [Paatero, Tapper 1994]
  - [Lee, Seung 1999]
    - ∗ semi-supervised NMF [H. Lee, Yoo, Choi 2009]



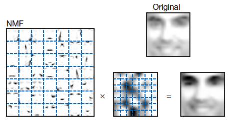
Lee, D., Seung, S. (1999) *Learning the parts of objects by non-negative matrix factorization.*

» Topic Modeling

- ▷ principal component analysis (PCA)
  - [Pearson 1901]
  - [Hotelling 1933]
  - ∗ supervised PCA [Bair, Hastie, Paul, Tibshirani 2006]
- ▷ latent dirichlet allocation (LDA)
  - [Pritchard, Stephens, Donnelly 2000]
  - [Blei, Ng, Jordan 2003]
  - ∗ supervised LDA [Blei, McAuliffe 2008]
  - ∗ hierarchical LDA [Blei, Griffiths, Jordan, Tenenbaum 2003]
- ▷ nonnegative matrix factorization (NMF)
  - [Paatero, Tapper 1994]
  - [Lee, Seung 1999]
  - ∗ semi-supervised NMF [H. Lee, Yoo, Choi 2009]



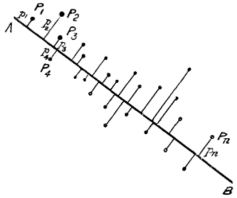
Pearson, K. (1901) *On lines and planes of closest fit to systems of points in space.*



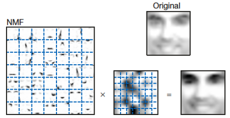
Lee, D., Seung, S. (1999) *Learning the parts of objects by non-negative matrix factorization.*

» Topic Modeling

- principal component analysis (PCA)
  - [Pearson 1901]
  - [Hotelling 1933]
    - supervised PCA [Bair, Hastie, Paul, Tibshirani 2006]
- latent dirichlet allocation (LDA)
  - [Pritchard, Stephens, Donnelly 2000]
  - [Blei, Ng, Jordan 2003]
    - supervised LDA [Blei, McAuliffe 2008]
    - hierarchical LDA [Blei, Griffiths, Jordan, Tenenbaum 2003]
- nonnegative matrix factorization (NMF)
  - [Paatero, Tapper 1994]
  - [Lee, Seung 1999]
    - semi-supervised NMF [H. Lee, Yoo, Choi 2009]
    - hierarchical NMF [Cichocki, Zdunek 2006]



Pearson, K. (1901) *On lines and planes of closest fit to systems of points in space.*

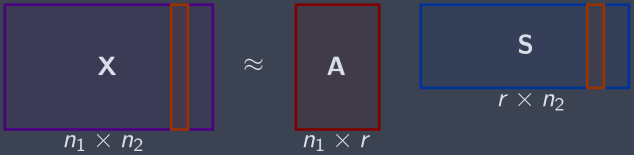


Lee, D., Seung, S. (1999) *Learning the parts of objects by non-negative matrix factorization.*

» Nonnegative Matrix Factorization (NMF)

Model: Given nonnegative data $\mathbf{X}$, compute nonnegative $\mathbf{A}$ and $\mathbf{S}$ of lower rank so that

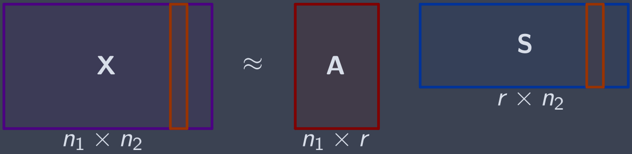
$$\mathbf{X} \approx \mathbf{AS}.$$



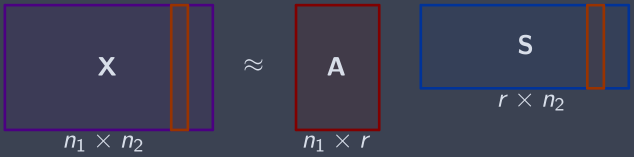
» Nonnegative Matrix Factorization (NMF)

Model: Given nonnegative data $\mathbf{X}$, compute nonnegative $\mathbf{A}$ and $\mathbf{S}$ of lower rank so that

$$\mathbf{X} \approx \mathbf{AS}.$$

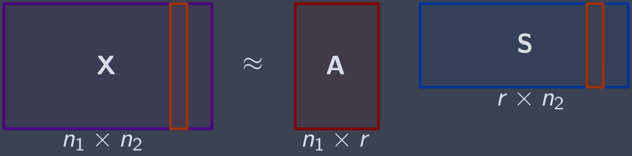


» Nonnegative Matrix Factorization (NMF)



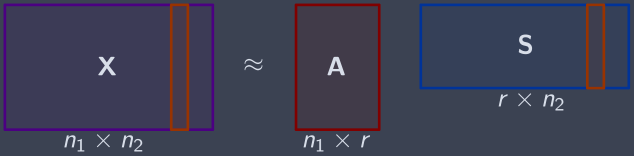
- Popularized by [Lee & Seung 1999]

» Nonnegative Matrix Factorization (NMF)



- ▷ Popularized by [Lee & Seung 1999]
- ▷ Employed for dimensionality-reduction and topic modeling

» Nonnegative Matrix Factorization (NMF)

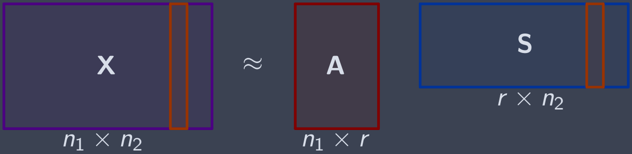


- ▷ Popularized by [Lee & Seung 1999]
- ▷ Employed for dimensionality-reduction and topic modeling
- ▷ Often formulated as

$$\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{n_1\times r},\mathbf{S}\in\mathbb{R}_{\geq 0}^{r\times n_2}} \|\mathbf{X}-\mathbf{AS}\|_F^2 \quad \text{or} \quad \min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{n_1\times r},\mathbf{S}\in\mathbb{R}_{\geq 0}^{r\times n_2}} D(\mathbf{X}\|\mathbf{AS}).^{1}$$

[1] information divergence $D(\mathbf{A}\|\mathbf{B}) = \sum_{i,j}\left(\mathbf{A}_{ij}\log\frac{\mathbf{A}_{ij}}{\mathbf{B}_{ij}} - \mathbf{A}_{ij} + \mathbf{B}_{ij}\right)$

» Nonnegative Matrix Factorization (NMF)



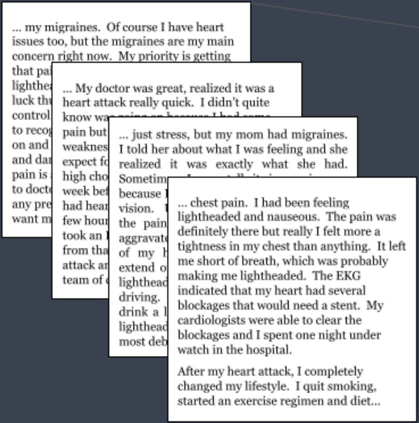
- Popularized by [Lee & Seung 1999]
- Employed for dimensionality-reduction and topic modeling
- Often formulated as

$$\min_{\mathbf{A}\in\mathbb{R}^{n_1\times r}_{\geq 0},\mathbf{S}\in\mathbb{R}^{r\times n_2}_{\geq 0}} \|\mathbf{X}-\mathbf{AS}\|_F^2 \quad \text{or} \quad \min_{\mathbf{A}\in\mathbb{R}^{n_1\times r}_{\geq 0},\mathbf{S}\in\mathbb{R}^{r\times n_2}_{\geq 0}} D(\mathbf{X}\|\mathbf{AS}).^{1}$$

- These formulations are MLE models

[1] information divergence $D(\mathbf{A}\|\mathbf{B}) = \sum_{i,j}\left(\mathbf{A}_{ij}\log\frac{\mathbf{A}_{ij}}{\mathbf{B}_{ij}} - \mathbf{A}_{ij} + \mathbf{B}_{ij}\right)$

» NMF Example



Patient Surveys

	Patients			
heart	3	3	0	1
weakness	2	0	0	0
chest	0	2	0	0
migraine	0	0	2	3
lightheaded	0	2	2	1
pain	3	2	2	4
⋮				

Term-Document Matrix

» NMF Example

Patients

heart	3	3	0	1
weakness	2	0	0	0
chest	0	2	0	0
migraine	0	0	2	3
lightheaded	0	2	2	1
pain	3	3	2	4
⋮				

NMF Example

$$
\begin{array}{r}
 \\ \text{heart} \\ \text{weakness} \\ \text{chest} \\ \text{migraine} \\ \text{lightheaded} \\ \text{pain} \\ \vdots
\end{array}
\begin{array}{c}
\text{Patients} \\
\begin{bmatrix} 3 & 3 & 0 & 1 \\ 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 3 \\ 0 & 2 & 2 & 1 \\ 3 & 3 & 2 & 4 \end{bmatrix} \\
\end{array}
\approx
\begin{bmatrix} 4.3 & 0.3 \\ 1.4 & 0 \\ 1.4 & 0 \\ 0 & 3.6 \\ 1.2 & 1.9 \\ 4.1 & 3.9 \end{bmatrix}
\begin{bmatrix} 0.7 & 0.7 & 0 & 0.1 \\ 0 & 0.1 & 0.6 & 0.8 \end{bmatrix}
$$

» NMF Example

Patients

heart	3	3	0	1
weakness	2	0	0	0
chest	0	2	0	0
migraine	0	0	2	3
lightheaded	0	2	2	1
pain	3	3	2	4
⋮				

≈

Topics

4.3	0.3
1.4	0
1.4	0
0	3.6
1.2	1.9
4.1	3.9

0.7	0.7	0	0.1
0	0.1	0.6	0.8

Lower-dimensional representation

» NMF Example

Patients

heart	3	3	0	1
weakness	2	0	0	0
chest	0	2	0	0
migraine	0	0	2	3
lightheaded	0	2	2	1
pain	3	3	2	4
⋮				

≈

Topics

Heart attack	Chronic migraine
4.3	0.3
1.4	0
1.4	0
0	3.6
1.2	1.9
4.1	3.9

0.7	0.7	0	0.1
0	0.1	0.6	0.8

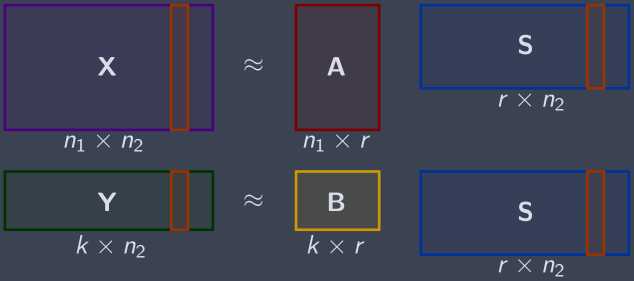
Lower-dimensional representation

Supervised Models

» Semi-supervised NMF (SSNMF)

Model: Jointly factorize nonnegative data $\mathbf{X}$ and supervision information $\mathbf{Y}$ so that

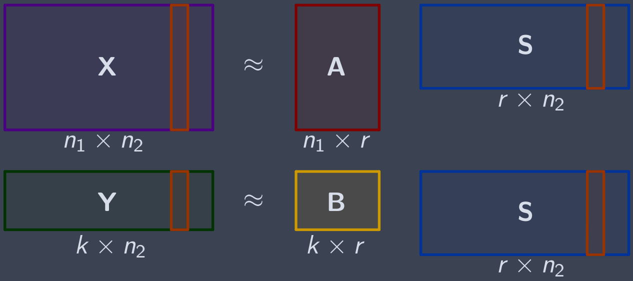
$$\mathbf{X} \approx \mathbf{AS} \text{ and } \mathbf{Y} \approx \mathbf{BS}.$$



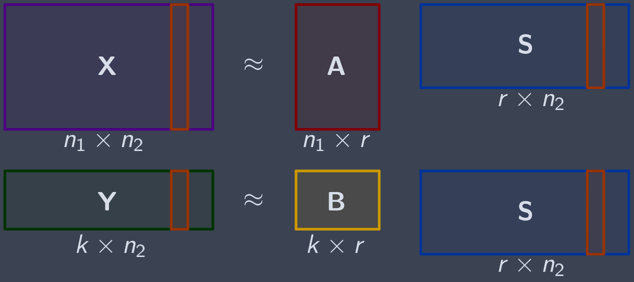
» Semi-supervised NMF (SSNMF)

Model: Jointly factorize nonnegative data $\mathbf{X}$ and supervision information $\mathbf{Y}$ so that

$$\mathbf{X} \approx \mathbf{AS} \text{ and } \mathbf{Y} \approx \mathbf{BS}.$$



» Semi-supervised NMF (SSNMF)



- ▷ Formulated in [Lee, Yoo & Choi 2009] as optimization problem

$$\min_{\mathbf{A}\in\mathbb{R}^{n_1\times r}_{\geq 0},\mathbf{B}\in\mathbb{R}^{k\times r}_{\geq 0},\mathbf{S}\in\mathbb{R}^{r\times n_2}_{\geq 0}} \|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda\|\mathbf{Y}-\mathbf{BS}\|_F^2.$$

SSNMF example

	Patients			
heart	3	3	0	1
weakness	2	0	0	0
chest	0	2	0	0
migraine	0	0	2	3
lightheaded	0	2	2	1
pain	3	3	2	4
⋮				

Classes	1	1	0	1
	0	0	1	1

» SSNMF example

$$
\begin{array}{r}
\\
\text{heart}\\
\text{weakness}\\
\text{chest}\\
\text{migraine}\\
\text{lightheaded}\\
\text{pain}\\
\vdots
\end{array}
\overset{\text{Patients}}{\begin{bmatrix} 3 & 3 & 0 & 1\\ 2 & 0 & 0 & 0\\ 0 & 2 & 0 & 0\\ 0 & 0 & 2 & 3\\ 0 & 2 & 2 & 1\\ 3 & 3 & 2 & 4 \end{bmatrix}}
\approx
\overset{\text{Topics}}{\begin{bmatrix} 4.3 & 0.1\\ 1.4 & 0\\ 1.4 & 0\\ 0 & 3.6\\ 1.2 & 1.9\\ 4.2 & 3.8 \end{bmatrix}}
\begin{bmatrix} 0.7 & 0.7 & 0 & 0.2\\ 0 & 0.1 & 0.6 & 0.8 \end{bmatrix}
$$

Lower-dimensional representation

$$
\text{Classes}\ \begin{bmatrix} 1 & 1 & 0 & 1\\ 0 & 0 & 1 & 1 \end{bmatrix}
\approx
\begin{bmatrix} 1.5 & 0.6\\ 0 & 1.4 \end{bmatrix}
\begin{bmatrix} 0.7 & 0.7 & 0 & 0.2\\ 0 & 0.1 & 0.6 & 0.8 \end{bmatrix}
$$

» SSNMF example

Patients | Topics

$$
\begin{array}{r}
\text{heart} \\ \text{weakness} \\ \text{chest} \\ \text{migraine} \\ \text{lightheaded} \\ \text{pain} \\ \vdots
\end{array}
\begin{bmatrix}
3 & 3 & 0 & 1 \\
2 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 \\
0 & 0 & 2 & 3 \\
0 & 2 & 2 & 1 \\
3 & 3 & 2 & 4
\end{bmatrix}
\approx
\begin{bmatrix}
4.3 & 0.1 \\
1.4 & 0 \\
1.4 & 0 \\
0 & 3.6 \\
1.2 & 1.9 \\
4.2 & 3.8
\end{bmatrix}
\begin{bmatrix}
0.7 & 0.7 & 0 & 0.2 \\
0 & 0.1 & 0.6 & 0.8
\end{bmatrix}
$$

Lower-dimensional representation

$$
\text{Classes}
\begin{bmatrix}
1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1
\end{bmatrix}
\approx
\begin{bmatrix}
1.5 & 0.6 \\
0 & 1.4
\end{bmatrix}
\begin{bmatrix}
0.7 & 0.7 & 0 & 0.2 \\
0 & 0.1 & 0.6 & 0.8
\end{bmatrix}
$$

yields insight into relationship between symptom expression and diagnoses, insight into common co-occuring diagnoses, and a predictive model

» NMF and SSNMF Models

NMF formulations

» NMF and SSNMF Models

NMF formulations

- [Lee & Seung 1999] $\|\cdot\|_F$-NMF:
  $\operatorname{argmin}_{\mathbf{A},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2$

» NMF and SSNMF Models

NMF formulations

- [Lee & Seung 1999] $\|\cdot\|_F$-NMF: $\operatorname{argmin}_{\mathbf{A},\mathbf{S}\geq 0}\|\mathbf{X} - \mathbf{AS}\|_F^2$
- [Lee & Seung 2001] $D(\cdot\|\cdot)$-NMF: $\operatorname{argmin}_{\mathbf{A},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})$

» NMF and SSNMF Models

NMF formulations

- [Lee & Seung 1999] $\|\cdot\|_F$-NMF: $\operatorname{argmin}_{\mathbf{A},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2$
- [Lee & Seung 2001] $D(\cdot\|\cdot)$-NMF: $\operatorname{argmin}_{\mathbf{A},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})$

* multiplicative updates [Lee & Seung 2001]

» NMF and SSNMF Models

NMF formulations

- [Lee & Seung 1999] $\|\cdot\|_F$-NMF: $\text{argmin}_{\mathbf{A},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2$
- [Lee & Seung 2001] $D(\cdot\|\cdot)$-NMF: $\text{argmin}_{\mathbf{A},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})$

* multiplicative updates [Lee & Seung 2001]
* MLE [Favaro & Soatto 2007]

» NMF and SSNMF Models

NMF formulations

- [Lee & Seung 1999] $\|\cdot\|_F$-NMF: $\operatorname{argmin}_{\mathbf{A},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2$
- [Lee & Seung 2001] $D(\cdot\|\cdot)$-NMF: $\operatorname{argmin}_{\mathbf{A},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})$

SSNMF formulations

- [H. Lee, Yoo & Choi 2009] $(\|\cdot\|_F, \|\cdot\|_F)$-SSNMF: $\operatorname{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda\|\mathbf{Y}-\mathbf{BS}\|_F^2$

* multiplicative updates [Lee & Seung 2001]
* MLE [Favaro & Soatto 2007]

» NMF and SSNMF Models

NMF formulations

- ▷ [Lee & Seung 1999] $\|\cdot\|_F$-NMF: $\operatorname{argmin}_{\mathbf{A},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2$
- ▷ [Lee & Seung 2001] $D(\cdot\|\cdot)$-NMF: $\operatorname{argmin}_{\mathbf{A},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})$

SSNMF formulations

- ▷ [H. Lee, Yoo & Choi 2009] $(\|\cdot\|_F, \|\cdot\|_F)$-SSNMF: $\operatorname{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda\|\mathbf{Y}-\mathbf{BS}\|_F^2$
- ▷ [H., Kassab, Li, et. al. 2020] $(\|\cdot\|_F, D(\cdot\|\cdot))$-SSNMF: $\operatorname{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda D(\mathbf{Y}\|\mathbf{BS})$

* multiplicative updates [Lee & Seung 2001]
* MLE [Favaro & Soatto 2007]

» NMF and SSNMF Models

NMF formulations

- ▷ [Lee & Seung 1999] $\|\cdot\|_F$-NMF: $\text{argmin}_{\mathbf{A},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2$
- ▷ [Lee & Seung 2001] $D(\cdot\|\cdot)$-NMF: $\text{argmin}_{\mathbf{A},\mathbf{S}\geq 0}\, D(\mathbf{X}\|\mathbf{AS})$

SSNMF formulations

- ▷ [H. Lee, Yoo & Choi 2009] $(\|\cdot\|_F, \|\cdot\|_F)$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda\|\mathbf{Y}-\mathbf{BS}\|_F^2$
- ▷ [H., Kassab, Li, et. al. 2020] $(\|\cdot\|_F, D(\cdot\|\cdot))$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda D(\mathbf{Y}\|\mathbf{BS})$
- ▷ [H., Kassab, Li, et. al. 2020] $(D(\cdot\|\cdot), \|\cdot\|_F)$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\, D(\mathbf{X}\|\mathbf{AS}) + \lambda\|\mathbf{Y}-\mathbf{BS}\|_F^2$

* ∗ multiplicative updates [Lee & Seung 2001]
* ∗ MLE [Favaro & Soatto 2007]

» NMF and SSNMF Models

NMF formulations

- [Lee & Seung 1999] $\|\cdot\|_F$-NMF: $\text{argmin}_{\mathbf{A},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2$
- [Lee & Seung 2001] $D(\cdot\|\cdot)$-NMF: $\text{argmin}_{\mathbf{A},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})$

SSNMF formulations

- [H. Lee, Yoo & Choi 2009] $(\|\cdot\|_F, \|\cdot\|_F)$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda\|\mathbf{Y}-\mathbf{BS}\|_F^2$
- [H., Kassab, Li, et. al. 2020] $(\|\cdot\|_F, D(\cdot\|\cdot))$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda D(\mathbf{Y}\|\mathbf{BS})$
- [H., Kassab, Li, et. al. 2020] $(D(\cdot\|\cdot), \|\cdot\|_F)$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})+\lambda\|\mathbf{Y}-\mathbf{BS}\|_F^2$
- [H., Kassab, Li, et. al. 2020] $(D(\cdot\|\cdot), D(\cdot\|\cdot))$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})+\lambda D(\mathbf{Y}\|\mathbf{BS})$

* multiplicative updates [Lee & Seung 2001]
* MLE [Favaro & Soatto 2007]

» NMF and SSNMF Models

NMF formulations

- [Lee & Seung 1999] $\|\cdot\|_F$-NMF: $\text{argmin}_{\mathbf{A},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2$
- [Lee & Seung 2001] $D(\cdot\|\cdot)$-NMF: $\text{argmin}_{\mathbf{A},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})$

* multiplicative updates [Lee & Seung 2001]
* MLE [Favaro & Soatto 2007]

SSNMF formulations

- [H. Lee, Yoo & Choi 2009] $(\|\cdot\|_F, \|\cdot\|_F)$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda\|\mathbf{Y}-\mathbf{BS}\|_F^2$
- [H., Kassab, Li, et. al. 2020] $(\|\cdot\|_F, D(\cdot\|\cdot))$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda D(\mathbf{Y}\|\mathbf{BS})$
- [H., Kassab, Li, et. al. 2020] $(D(\cdot\|\cdot), \|\cdot\|_F)$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})+\lambda\|\mathbf{Y}-\mathbf{BS}\|_F^2$
- [H., Kassab, Li, et. al. 2020] $(D(\cdot\|\cdot), D(\cdot\|\cdot))$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})+\lambda D(\mathbf{Y}\|\mathbf{BS})$

* $(\|\cdot\|_F, \|\cdot\|_F)$-SSNMF multiplicative updates [H. Lee, Yoo & Choi 2009]

» NMF and SSNMF Models

NMF formulations

- [Lee & Seung 1999] $\|\cdot\|_F$-NMF: $\text{argmin}_{\mathbf{A},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2$
- [Lee & Seung 2001] $D(\cdot\|\cdot)$-NMF: $\text{argmin}_{\mathbf{A},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS})$

* multiplicative updates [Lee & Seung 2001]
* MLE [Favaro & Soatto 2007]

SSNMF formulations

- [H. Lee, Yoo & Choi 2009] $(\|\cdot\|_F, \|\cdot\|_F)$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda\|\mathbf{Y}-\mathbf{BS}\|_F^2$
- [H., Kassab, Li, et. al. 2020] $(\|\cdot\|_F, D(\cdot\|\cdot))$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0}\|\mathbf{X}-\mathbf{AS}\|_F^2+\lambda D(\mathbf{Y}\|\mathbf{BS})$
- [H., Kassab, Li, et. al. 2020] $(D(\cdot\|\cdot), \|\cdot\|_F)$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS}) + \lambda\|\mathbf{Y}-\mathbf{BS}\|_F^2$
- [H., Kassab, Li, et. al. 2020] $(D(\cdot\|\cdot), D(\cdot\|\cdot))$-SSNMF: $\text{argmin}_{\mathbf{A},\mathbf{B},\mathbf{S}\geq 0} D(\mathbf{X}\|\mathbf{AS}) + \lambda D(\mathbf{Y}\|\mathbf{BS})$

* $(\|\cdot\|_F, \|\cdot\|_F)$-SSNMF multiplicative updates [H. Lee, Yoo & Choi 2009]
* multiplicative updates, MLE [H., Kassab, Li, et. al. 2020]

» Guided NMF

Idea: Instead of supervising factorization with class label matrix, supervise with model or seed topics.

[H., Needell, Rebrova, Vendrow 2021]
Related work: [Jagarlamudi, Jagadeesh, Daume, Udupa 2012]

» Guided NMF

Idea: Instead of supervising factorization with class label matrix, supervise with model or seed topics.

Model: Jointly factorize nonnegative data $\mathbf{X} \in \mathbb{R}_{\geq 0}^{n_1 \times n_2}$ and seed matrix $\mathbf{Y} = [\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \ldots, \boldsymbol{v}^{(c)}] \in \mathbb{R}_{\geq 0}^{n_1 \times c}$ so that

$$\mathbf{X} \approx \mathbf{AS} \text{ and } \mathbf{Y} \approx \mathbf{AB}.$$

[H., Needell, Rebrova, Vendrow 2021]
Related work: [Jagarlamudi, Jagadeesh, Daume, Udupa 2012]

» Guided NMF

Idea: Instead of supervising factorization with class label matrix, supervise with model or seed topics.

Model: Jointly factorize nonnegative data $\mathbf{X} \in \mathbb{R}_{\geq 0}^{n_1 \times n_2}$ and seed matrix $\mathbf{Y} = [\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \ldots, \boldsymbol{v}^{(c)}] \in \mathbb{R}_{\geq 0}^{n_1 \times c}$ so that

$$\mathbf{X} \approx \mathbf{AS} \text{ and } \mathbf{Y} \approx \mathbf{AB}.$$

- formulated as transposed SSNMF model, i.e.,

$$\min_{\mathbf{A},\mathbf{S},\mathbf{B} \geq 0} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \lambda \|\mathbf{Y} - \mathbf{AB}\|_F^2$$

[H., Needell, Rebrova, Vendrow 2021]
Related work: [Jagarlamudi, Jagadeesh, Daume, Udupa 2012]

» Guided NMF

Idea: Instead of supervising factorization with class label matrix, supervise with model or seed topics.

Model: Jointly factorize nonnegative data $\mathbf{X} \in \mathbb{R}_{\geq 0}^{n_1 \times n_2}$ and seed matrix $\mathbf{Y} = [\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \ldots, \boldsymbol{v}^{(c)}] \in \mathbb{R}_{\geq 0}^{n_1 \times c}$ so that

$$\mathbf{X} \approx \mathbf{AS} \text{ and } \mathbf{Y} \approx \mathbf{AB}.$$

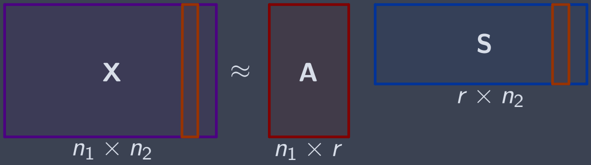
- ▷ formulated as transposed SSNMF model, i.e.,

$$\min_{\mathbf{A},\mathbf{S},\mathbf{B} \geq 0} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \lambda \|\mathbf{Y} - \mathbf{AB}\|_F^2$$

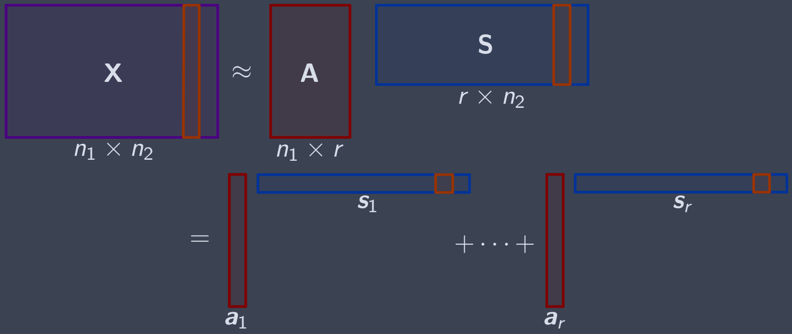
- ▷ `allows for use of expert guidance`

[H., Needell, Rebrova, Vendrow 2021]
Related work: [Jagarlamudi, Jagadeesh, Daume, Udupa 2012]

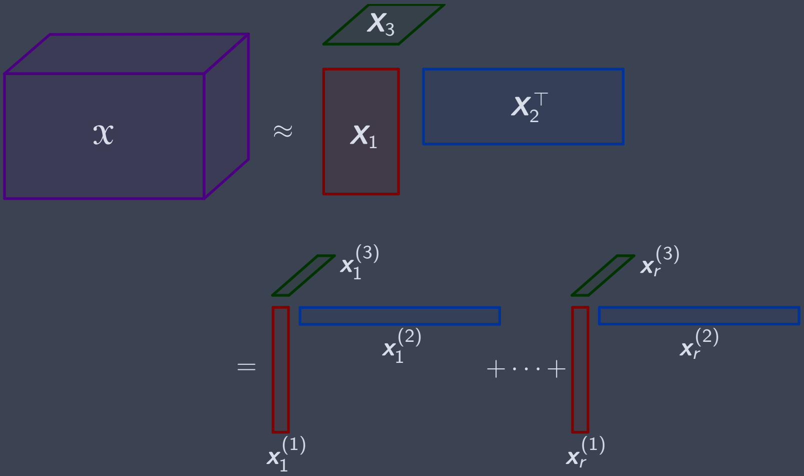
» What about tensor data?



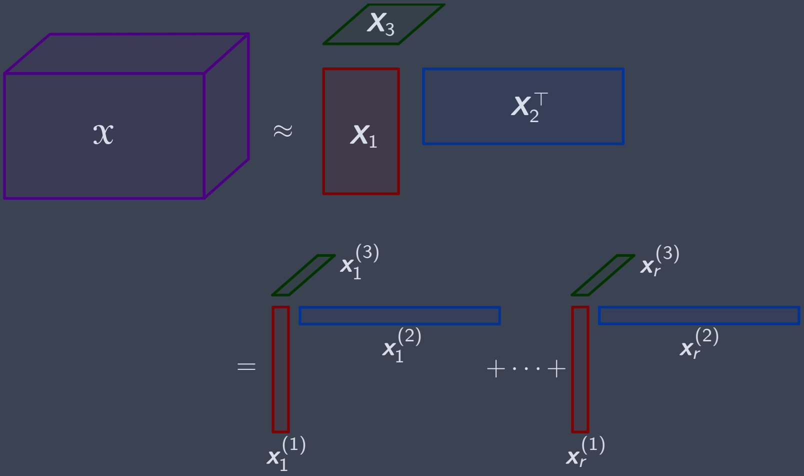
» What about tensor data?



» What about tensor data?



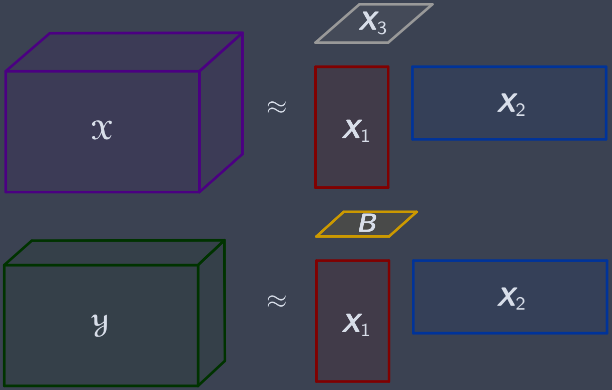
» What about tensor data?



nonnegative CANDECOMP/PARAFAC (CP) decomposition (NCPD)

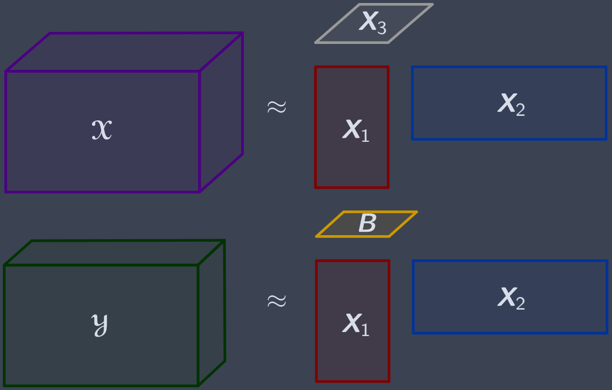
[Carroll, Chang 1970] [Harshman 1970]

» Semi-supervised NCPD



forthcoming work with Chau, Alaev, Vendrow, Grotheer, Kryshchenko, Leonard, Needell
Related work: [Verma, Liu, Wang, Zhu 2017], [Cao, Lu, Wei, Philip, Leow 2016], [Lock, Li 2018]

» Semi-supervised NCPD

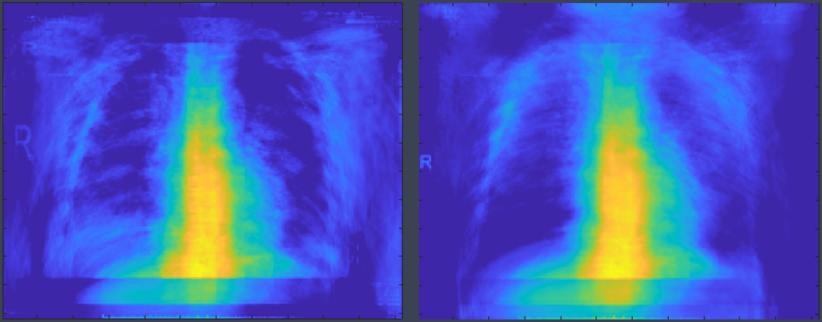


flexible to many forms of supervision

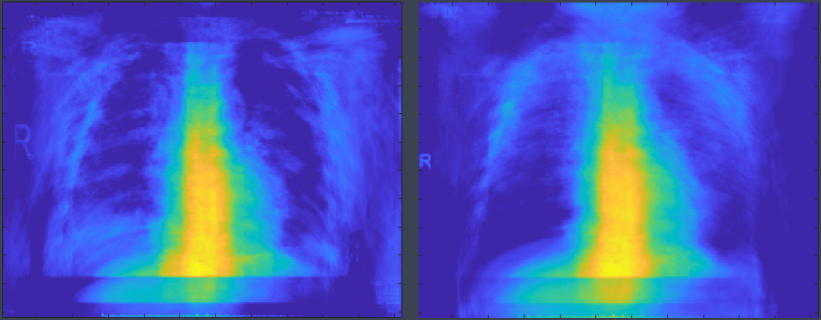
---

forthcoming work with Chau, Alaev, Vendrow, Grotheer, Kryshchenko, Leonard, Needell
Related work: [Verma, Liu, Wang, Zhu 2017], [Cao, Lu, Wei, Philip, Leow 2016], [Lock, Li 2018]

» Application: COVIDx archive



» Application: COVIDx archive



Left: ‘viral pneumonia’ topic; Right: ‘Normal’ topic.

Hierarchical Models

» Hierarchical NMF

Model: Sequentially factorize

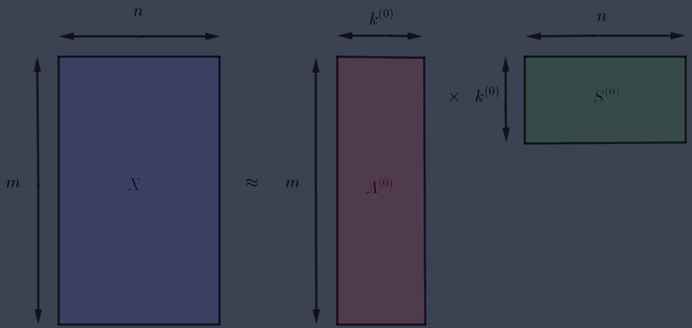
$$X \approx A^{(0)}S^{(0)}, S^{(0)} \approx A^{(1)}S^{(1)}, S^{(1)} \approx A^{(2)}S^{(2)}, \ldots, S^{(\mathcal{L}-1)} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}.$$

[Cichocki, Zdunek 2006]

» Hierarchical NMF

Model: Sequentially factorize

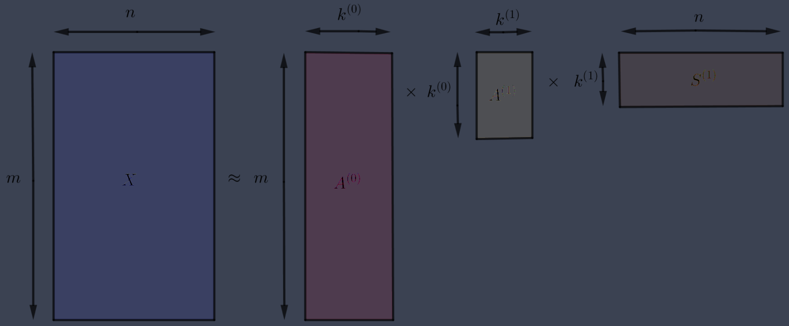
$$X \approx A^{(0)}S^{(0)}, S^{(0)} \approx A^{(1)}S^{(1)}, S^{(1)} \approx A^{(2)}S^{(2)}, \ldots, S^{(\mathcal{L}-1)} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}.$$



» Hierarchical NMF

Model: Sequentially factorize

$$X \approx A^{(0)} S^{(0)}, S^{(0)} \approx A^{(1)} S^{(1)}, S^{(1)} \approx A^{(2)} S^{(2)}, \ldots, S^{(\mathcal{L}-1)} \approx A^{(\mathcal{L})} S^{(\mathcal{L})}.$$

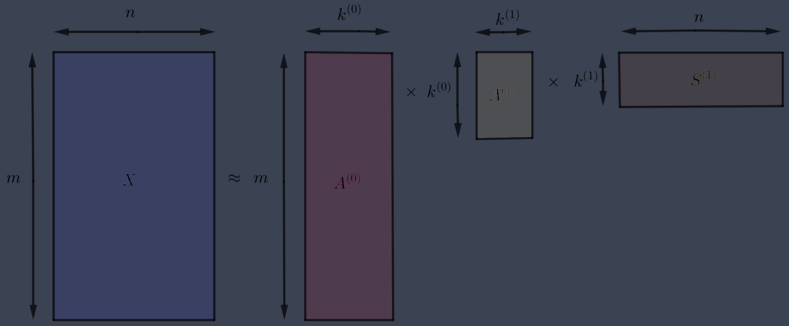


[Cichocki, Zdunek 2006]

» Hierarchical NMF

Model: Sequentially factorize

$$X \approx A^{(0)}S^{(0)}, S^{(0)} \approx A^{(1)}S^{(1)}, S^{(1)} \approx A^{(2)}S^{(2)}, \ldots, S^{(\mathcal{L}-1)} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}.$$



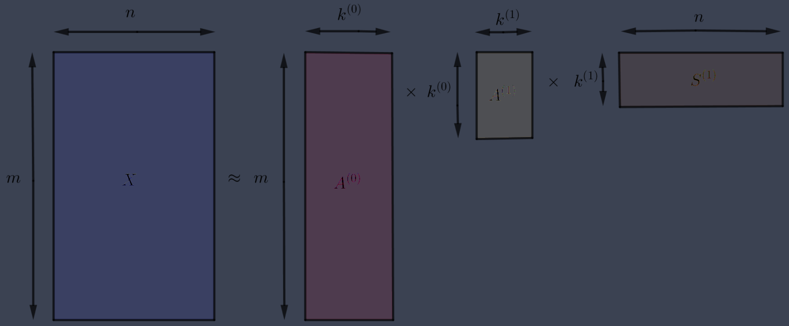
- ▷ $k^{(\ell)}$: supertopics collecting $k^{(\ell-1)}$ subtopics

[Cichocki, Zdunek 2006]

» Hierarchical NMF

Model: Sequentially factorize

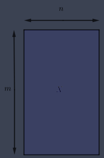
$$X \approx A^{(0)}S^{(0)}, S^{(0)} \approx A^{(1)}S^{(1)}, S^{(1)} \approx A^{(2)}S^{(2)}, ..., S^{(\mathcal{L}-1)} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}.$$



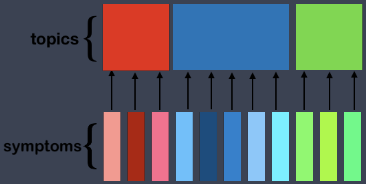
- ▷ $k^{(\ell)}$: supertopics collecting $k^{(\ell-1)}$ subtopics
- ▷ error propagates through layers

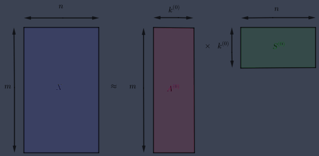
[Cichocki, Zdunek 2006]

» Hierarchical NMF

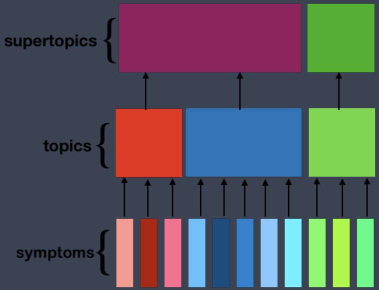


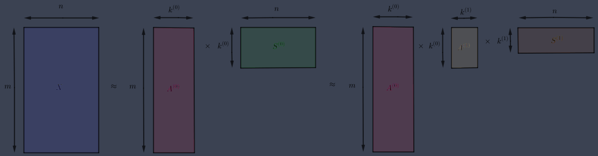
Hierarchical NMF



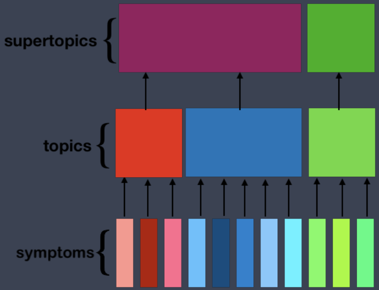


» Hierarchical NMF

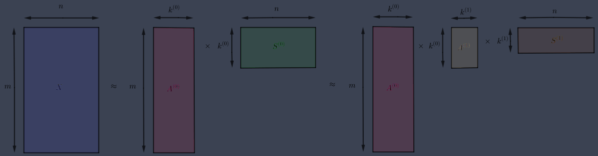




» Hierarchical NMF



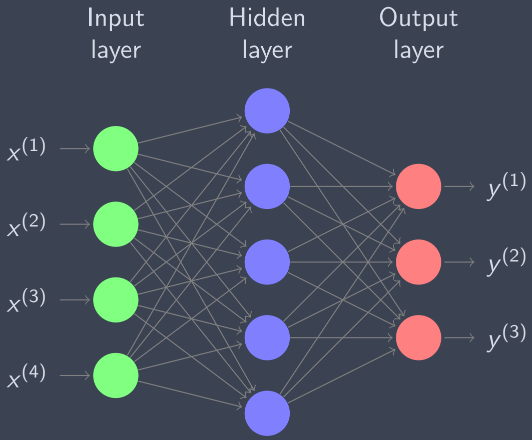
- ▷ hNMF can be implemented in a `feed-forward neural network` structure



» Feed-forward Neural Networks

Goal: Identify weights $W^{(1)}, W^{(2)}, \dots, W^{(L)}$ to minimize model error

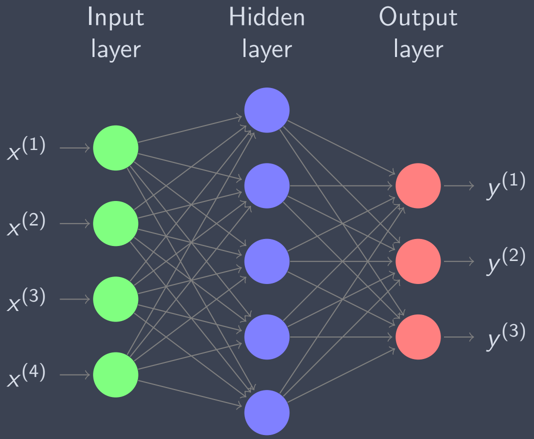
$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} f(\boldsymbol{y}(\boldsymbol{x}_n, \{W^{(i)}\}), \boldsymbol{x}_n, \boldsymbol{t}_n).$$



» Feed-forward Neural Networks

Goal: Identify weights $W^{(1)}, W^{(2)}, ..., W^{(L)}$ to minimize model error

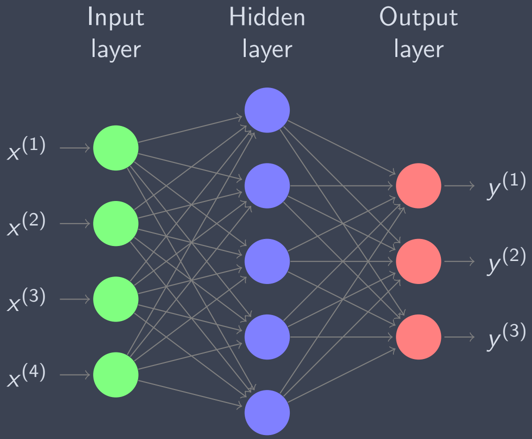
$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} \|\boldsymbol{y}(\boldsymbol{x}_n, \{W^{(i)}\}) - \boldsymbol{t}_n\|_2^2.$$



» Feed-forward Neural Networks

Goal: Identify weights $W^{(1)}, W^{(2)}, ..., W^{(L)}$ to minimize model error

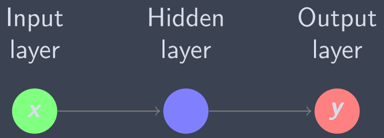
$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} f(\boldsymbol{y}(\boldsymbol{x}_n, \{W^{(i)}\}), \boldsymbol{x}_n, \boldsymbol{t}_n).$$



» Feed-forward Neural Networks

Goal: Identify weights $W^{(1)}, W^{(2)}, \dots, W^{(L)}$ to minimize model error

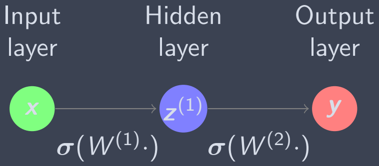
$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} f(\boldsymbol{y}(\boldsymbol{x}_n, \{W^{(i)}\}), \boldsymbol{x}_n, \boldsymbol{t}_n).$$



» Feed-forward Neural Networks

Goal: Identify weights $W^{(1)}, W^{(2)}, \ldots, W^{(L)}$ to minimize model error

$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} f(\boldsymbol{y}(\boldsymbol{x}_n, \{W^{(i)}\}), \boldsymbol{x}_n, \boldsymbol{t}_n).$$



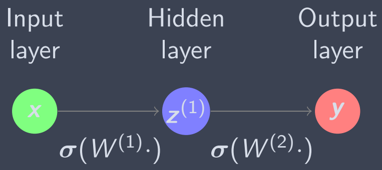
Training:

- ▷ forward propagation:
  $\boldsymbol{z}^{(1)} = \sigma(W^{(1)}\boldsymbol{x})$,
  $\boldsymbol{z}^{(2)} = \sigma(W^{(2)}\boldsymbol{z}_1)$,
  $\ldots$,
  $\boldsymbol{y} = \sigma(W^{(L)}\boldsymbol{z}^{(L-1)})$

» Feed-forward Neural Networks

Goal: Identify weights $W^{(1)}, W^{(2)}, \ldots, W^{(L)}$ to minimize model error

$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} f(\mathbf{y}(\mathbf{x}_n, \{W^{(i)}\}), \mathbf{x}_n, \mathbf{t}_n).$$

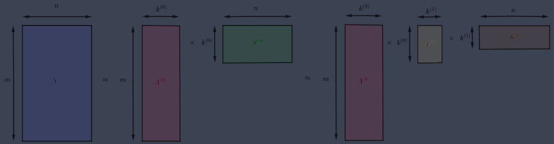


Training:

- ▷ forward propagation: $\mathbf{z}^{(1)} = \sigma(W^{(1)}\mathbf{x})$, $\mathbf{z}^{(2)} = \sigma(W^{(2)}\mathbf{z}_1)$, ..., $\mathbf{y} = \sigma(W^{(L)}\mathbf{z}^{(L-1)})$
- ▷ back propagation: update $\{W^{(i)}\}$ with $\nabla E(\{W^{(i)}\})$

» Neural NMF

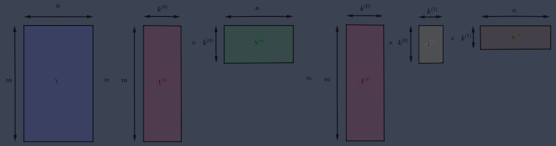
Goal: Develop forward and back propagation algorithms for hNMF.



[Gao, H., Molitor, Needell, Sadovnik, Will, Zhang 2019]
Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Neural NMF

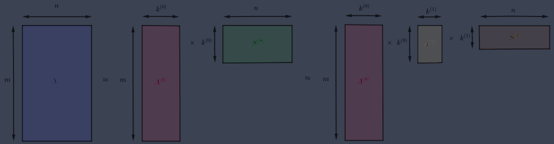
Goal: Develop forward and back propagation algorithms for hNMF.



- ▷ Regard the A matrices as independent variables, determine the S matrices from the A matrices.

» Neural NMF

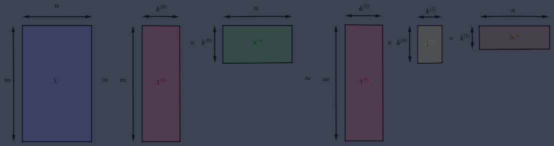
Goal: Develop forward and back propagation algorithms for hNMF.



- ▷ Regard the A matrices as independent variables, determine the S matrices from the A matrices.
- ▷ Define $q(X, A) := \text{argmin}_{S \geq 0} \|X - AS\|_F^2$ (least-squares).

» Neural NMF

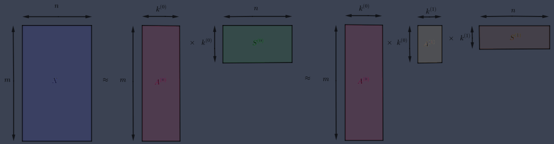
Goal: Develop forward and back propagation algorithms for hNMF.



- Regard the A matrices as independent variables, determine the S matrices from the A matrices.
- Define $q(X, A) := \operatorname{argmin}_{S \geq 0} \|X - AS\|_F^2$ (least-squares).
- Pin the values of S to those of A by recursively setting $S^{(\ell)} := q(S^{(\ell-1)}, A^{(\ell)})$.

» Neural NMF

Goal: Develop forward and back propagation algorithms for hNMF.

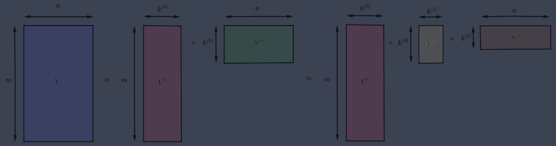


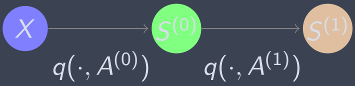
- Regard the A matrices as independent variables, determine the S matrices from the A matrices.
- Define $q(X, A) := \text{argmin}_{S \geq 0} \|X - AS\|_F^2$ (least-squares).
- Pin the values of S to those of A by recursively setting $S^{(\ell)} := q(S^{(\ell-1)}, A^{(\ell)})$.

$$X \xrightarrow{q(\cdot, A^{(0)})} S^{(0)} \xrightarrow{q(\cdot, A^{(1)})} S^{(1)}$$

» Neural NMF

Goal: Develop forward and back propagation algorithms for hNMF.

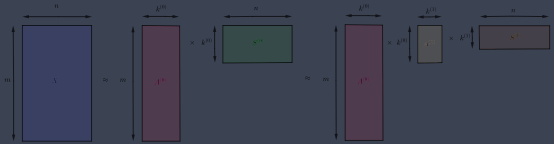




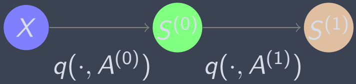
[Gao, H., Molitor, Needell, Sadovnik, Will, Zhang 2019]
Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Neural NMF

Goal: Develop forward and back propagation algorithms for hNMF.



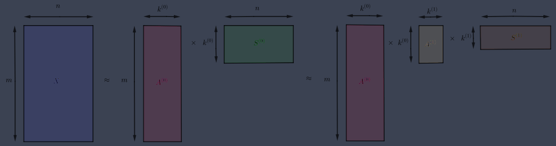
Training:

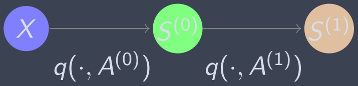


[Gao, H., Molitor, Needell, Sadovnik, Will, Zhang 2019]
Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Neural NMF

Goal: Develop forward and back propagation algorithms for hNMF.



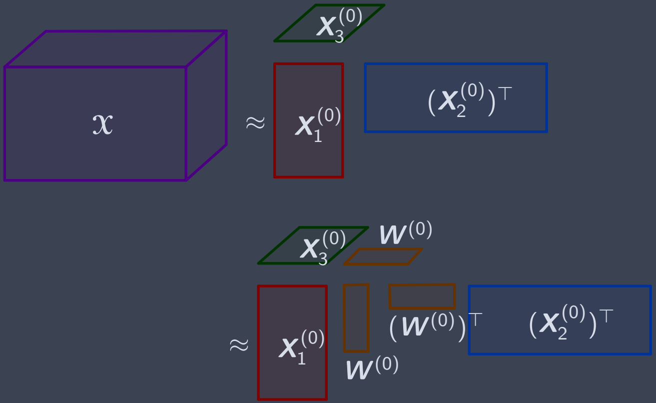


Training:

- forward propagation:
  $S^{(0)} = q(X, A^{(0)})$,
  $S^{(1)} = q(S^{(0)}, A^{(1)})$, ...,
  $S^{(L)} = q(S^{(L-1)}, A^{(L)})$
- back propagation: update $\{A^{(i)}\}$ with $\nabla E(\{A^{(i)}\})$

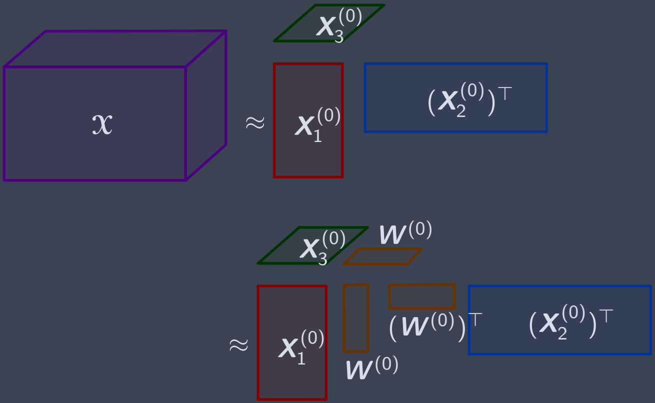
[Gao, H., Molitor, Needell, Sadovnik, Will, Zhang 2019]
Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» What about tensor data?



[Vendrow, H., Needell 2021]

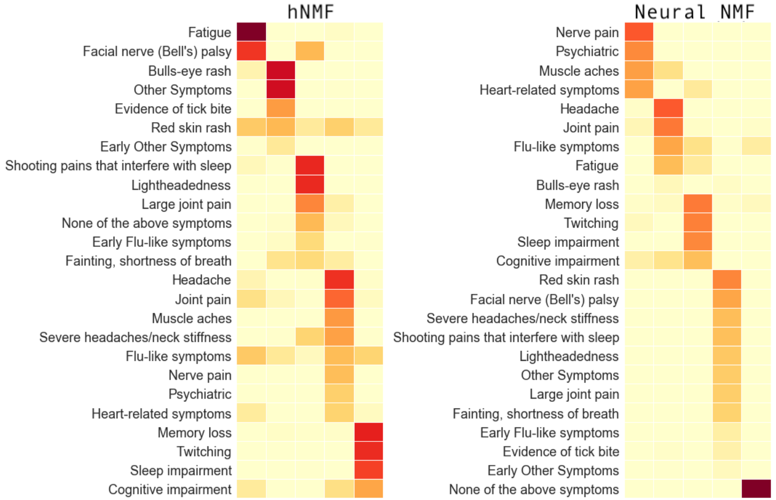
» What about tensor data?



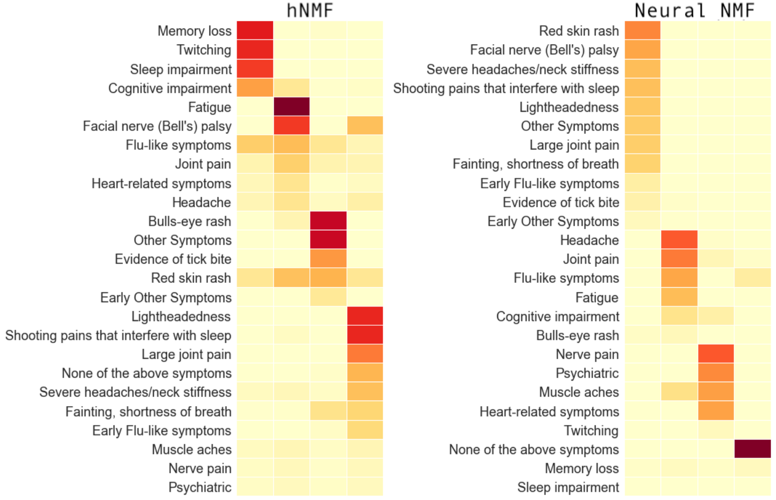
Neural NCPD model can again be formulated and trained in neural network framework

[Vendrow, H., Needell 2021]

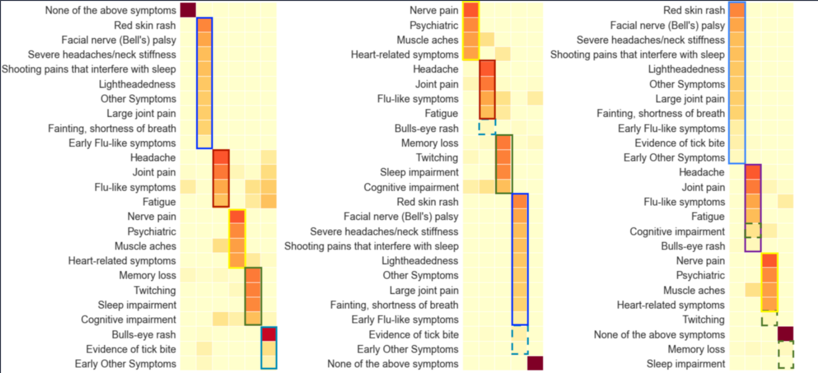
» Application: MyLymeData



» Application: MyLymeData

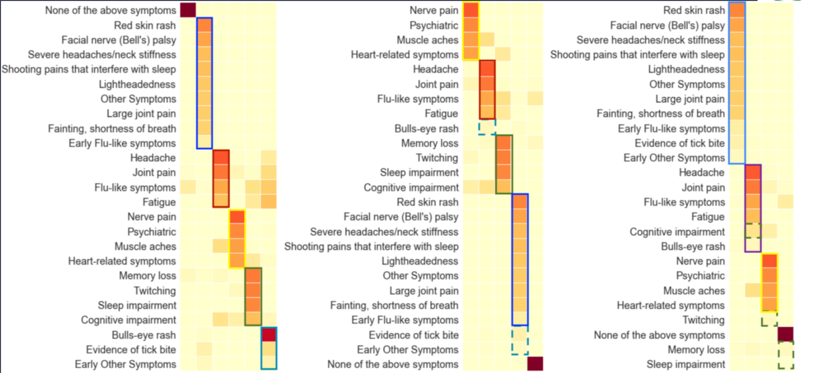


» Application: MyLymeData



$k^{(0)} = 6$ $k^{(1)} = 5$ $k^{(2)} = 4$

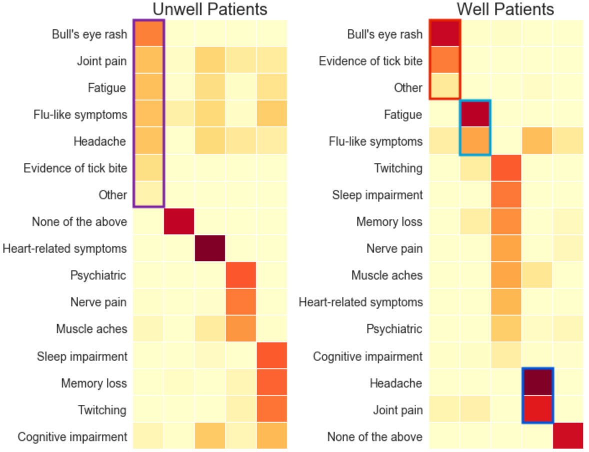
» Application: MyLymeData



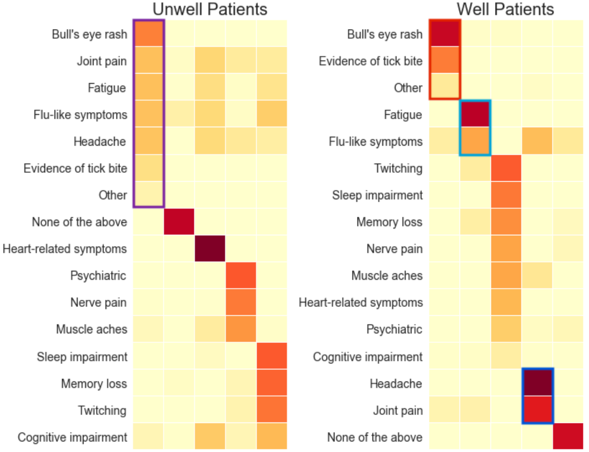
$k^{(0)} = 6$ $k^{(1)} = 5$ $k^{(2)} = 4$

bulls-eye rash (diagnosing symptoms) topic does not seem to persist for smaller number of topics

» Application: MyLymeData



» Application: MyLymeData

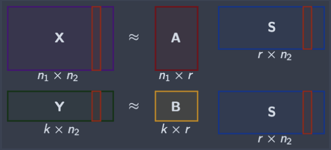


unwell and well patients have very different presentation of bulls-eye rash symptom in topics

Conclusions

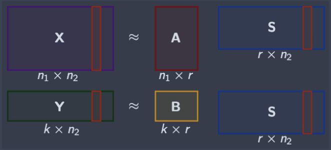
Conclusions

- SSNMF and SSNCPD models are maximum likelihood estimators



» Conclusions

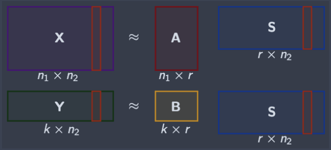
- ▷ SSNMF and SSNCPD models are maximum likelihood estimators



- ▷ they can be trained by `multiplicative updates`

» Conclusions

- SSNMF and SSNCPD models are maximum likelihood estimators



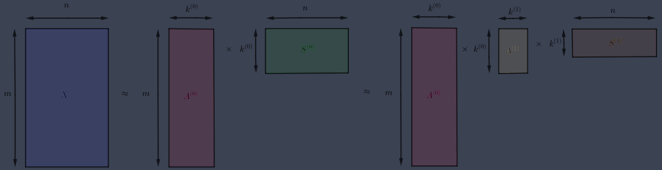
- they can be trained by `multiplicative updates`
- allow for use of side supervision information and expert guidance

» Conclusions

- ▷ SSNMF and SSNCPD models are maximum likelihood estimators

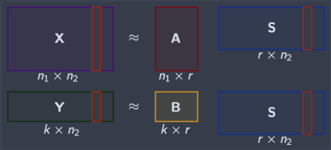
- ▷ they can be trained by `multiplicative updates`
- ▷ allow for use of side supervision information and expert guidance

- ▷ hNMF model can be implemented as a feed-forward neural network



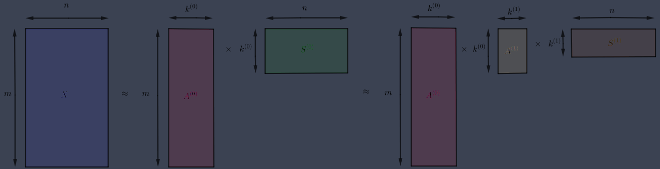
» Conclusions

- ▷ SSNMF and SSNCPD models are maximum likelihood estimators



- ▷ they can be trained by `multiplicative updates`
- ▷ allow for use of side supervision information and expert guidance

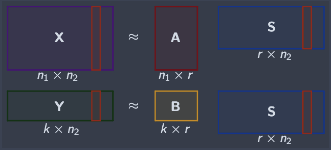
- ▷ hNMF model can be implemented as a feed-forward neural network



- ▷ `Neural NMF` and `Neural NCPD` can decrease error propagation

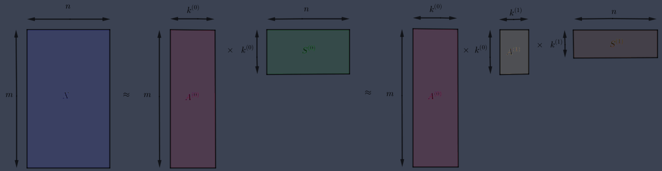
Conclusions

- SSNMF and SSNCPD models are maximum likelihood estimators



- they can be trained by `multiplicative updates`
- allow for use of side supervision information and expert guidance

- hNMF model can be implemented as a feed-forward neural network



- `Neural NMF` and `Neural NCPD` can decrease error propagation
- elucidate hierarchical relationships between learned topics and decrease dependence upon hyperparameters

» Thanks for listening! Questions?

[1] M. Gao, J. Haddock, D. Molitor, D. Needell, E. Sadovnik, T. Will, and R. Zhang. Neural nonnegative matrix factorization for hierarchical multilayer topic modeling. In Proc. Interational Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2019.

[2] J. Haddock, L. Kassab, S. Li, A. Kryshchenko, R. Grotheer, E. Sizikova, C. Wang, T. Merkh, R. W. M. A. Madushani, M. Ahn, D. Needell, and K. Leonard. Semi-supervised nonnegative matrix factorization models for document classification. In Asilomar Conf. on Signals, Systems, Computers (ACSSC), 2021.

[3] J. Vendrow, J. Haddock, and D. Needell. Neural nonnegative CP decomposition for hierarchical tensor analysis. In Asilomar Conf. on Signals, Systems, Computers (ACSSC), 2021.

[4] J. Vendrow, J. Haddock, E. Rebrova, and D. Needell. On a guided nonnegative matrix factorization. In Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2021.

[5] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755):788–791, 1999.

[6] Hyekyoung Lee, Jiho Yoo, and Seungjin Choi. Semi-supervised nonnegative matrix factorization. IEEE Signal Processing Letters, 17(1):4–7, 2009.

[7] Andrzej Cichocki and Rafal Zdunek. Multilayer nonnegative matrix factorisation. Electronics Letters, 42(16):947–948, 2006.

[8] Paolo Favaro and Stefano Soatto. 3-d shape estimation and image restoration: Exploiting defocus and motion-blur. Springer Science & Business Media, 2007.