Neural Nonnegative Matrix Factorization for Hierarchical Multilayer Topic Modeling

Jamie Haddock

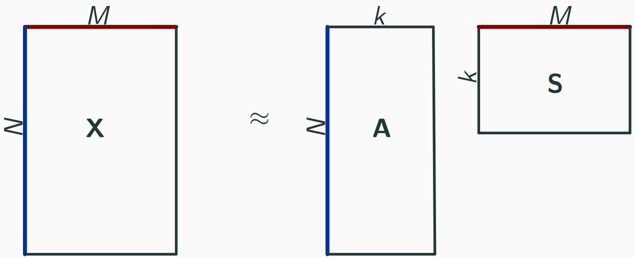
CAMSAP 2019,

December 16, 2019

Computational and Applied Mathematics
UCLA

joint with Mengdi Gao, Denali Molitor, Deanna Needell, Eli Sadovnik, Tyler Will, Runyu Zhang

Nonnegative Matrix Factorization (NMF)

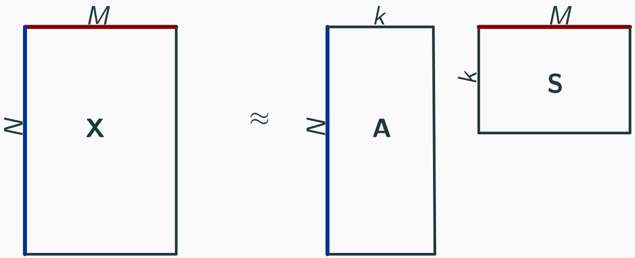


$$\min_{\mathbf{A}\in\mathbb{R}^{N\times k}_{\geq 0},\mathbf{S}\in\mathbb{R}^{k\times M}_{\geq 0}} \|\mathbf{X} - \mathbf{AS}\|_F^2$$

Problem Setup:

Problem Challenges:

Nonnegative Matrix Factorization (NMF)



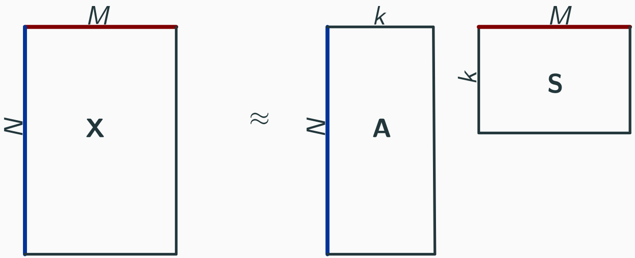
$$\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{N\times k},\mathbf{S}\in\mathbb{R}_{\geq 0}^{k\times M}} \|\mathbf{X}-\mathbf{AS}\|_F^2$$

Problem Setup:

$\mathbf{X} \in \mathbb{R}_{\geq 0}^{N\times M}$: data matrix

Problem Challenges:

Nonnegative Matrix Factorization (NMF)



$$\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{N\times k},\mathbf{S}\in\mathbb{R}_{\geq 0}^{k\times M}} \|\mathbf{X}-\mathbf{AS}\|_F^2$$

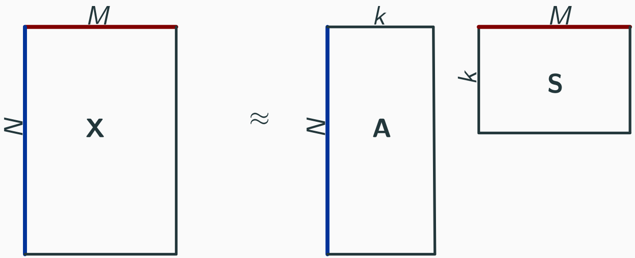
Problem Setup:

$\mathbf{X} \in \mathbb{R}_{\geq 0}^{N\times M}$: data matrix

$\mathbf{A} \in \mathbb{R}_{\geq 0}^{N\times k}$: features matrix

Problem Challenges:

Nonnegative Matrix Factorization (NMF)



$$\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{N\times k},\mathbf{S}\in\mathbb{R}_{\geq 0}^{k\times M}} \|\mathbf{X}-\mathbf{AS}\|_F^2$$

Problem Setup:

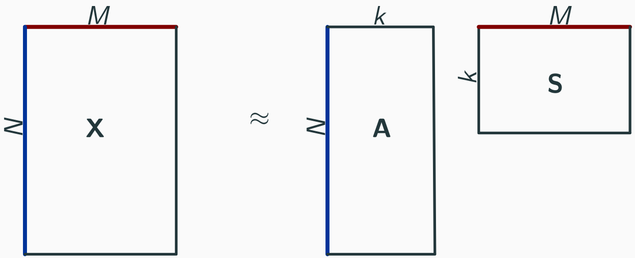
$\mathbf{X} \in \mathbb{R}_{\geq 0}^{N\times M}$: data matrix

$\mathbf{A} \in \mathbb{R}_{\geq 0}^{N\times k}$: features matrix

$\mathbf{S} \in \mathbb{R}_{\geq 0}^{k\times M}$: coefficients matrix

Problem Challenges:

Nonnegative Matrix Factorization (NMF)



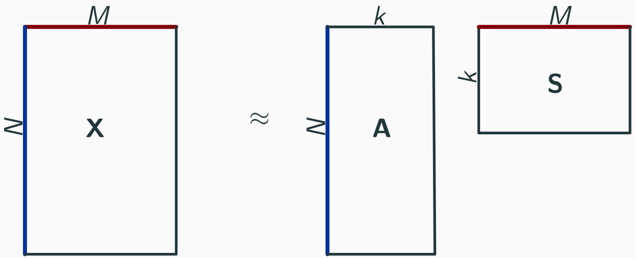
$$\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{N\times k},\mathbf{S}\in\mathbb{R}_{\geq 0}^{k\times M}} \|\mathbf{X}-\mathbf{AS}\|_F^2$$

Problem Setup:

$\mathbf{X} \in \mathbb{R}_{\geq 0}^{N\times M}$: data matrix
$\mathbf{A} \in \mathbb{R}_{\geq 0}^{N\times k}$: features matrix
$\mathbf{S} \in \mathbb{R}_{\geq 0}^{k\times M}$: coefficients matrix
k: user chosen parameter

Problem Challenges:

Nonnegative Matrix Factorization (NMF)



$$\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{N\times k}, \mathbf{S}\in\mathbb{R}_{\geq 0}^{k\times M}} \|\mathbf{X} - \mathbf{AS}\|_F^2$$

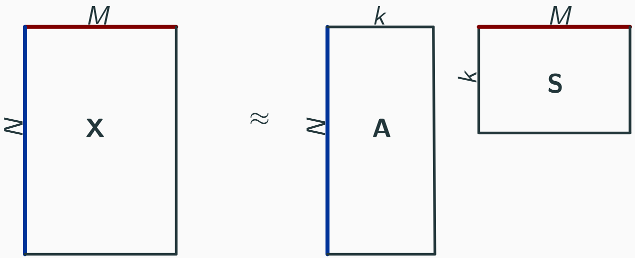
Problem Setup:

$\mathbf{X} \in \mathbb{R}_{\geq 0}^{N\times M}$: data matrix
$\mathbf{A} \in \mathbb{R}_{\geq 0}^{N\times k}$: features matrix
$\mathbf{S} \in \mathbb{R}_{\geq 0}^{k\times M}$: coefficients matrix
k: user chosen parameter

Problem Challenges:

- ▷ nonconvex in **A** and **S**, NP-hard [Vavasis '08]

Nonnegative Matrix Factorization (NMF)



$$\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{N\times k},\mathbf{S}\in\mathbb{R}_{\geq 0}^{k\times M}} \|\mathbf{X}-\mathbf{A}\mathbf{S}\|_F^2$$

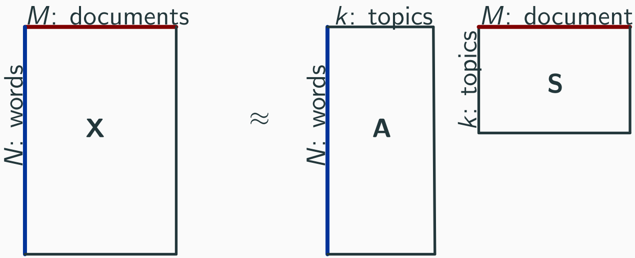
Problem Setup:

$\mathbf{X} \in \mathbb{R}_{\geq 0}^{N\times M}$: data matrix
$\mathbf{A} \in \mathbb{R}_{\geq 0}^{N\times k}$: features matrix
$\mathbf{S} \in \mathbb{R}_{\geq 0}^{k\times M}$: coefficients matrix
k: user chosen parameter

Problem Challenges:

- ▷ nonconvex in **A** and **S**, NP-hard [Vavasis '08]
- ▷ interpretability of factors dependent upon k

Nonnegative Matrix Factorization (NMF)



$$\min_{\mathbf{A}\in\mathbb{R}^{N\times k}_{\geq 0}, \mathbf{S}\in\mathbb{R}^{k\times M}_{\geq 0}} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2$$

Problem Setup:

$\mathbf{X} \in \mathbb{R}^{N\times M}_{\geq 0}$: data matrix
$\mathbf{A} \in \mathbb{R}^{N\times k}_{\geq 0}$: features matrix
$\mathbf{S} \in \mathbb{R}^{k\times M}_{\geq 0}$: coefficients matrix
k: user chosen parameter

Problem Challenges:

- nonconvex in $\mathbf{A}$ and $\mathbf{S}$, NP-hard [Vavasis '08]
- interpretability of factors dependent upon k

NMF

Applications:

Methods:

NMF

Applications:

- low-rank approximation

Methods:

NMF

Applications:

- low-rank approximation
- clustering

Methods:

NMF

Applications:

- low-rank approximation
- clustering
- topic modeling

Methods:

NMF

Applications:

- low-rank approximation
- clustering
- topic modeling
- feature extraction

Methods:

NMF

Applications:

- low-rank approximation
- clustering
- topic modeling
- feature extraction

Methods:

- multiplicative updates

NMF

Applications:

- ▷ low-rank approximation
- ▷ clustering
- ▷ topic modeling
- ▷ feature extraction

Methods:

- ▷ multiplicative updates
- ▷ alternating nonnegative least squares

NMF

Applications:

- low-rank approximation
- clustering
- topic modeling
- feature extraction

Methods:

- multiplicative updates
- alternating nonnegative least squares
- many others

(Semi)supervised NMF

Goal: Incorporate known label information into problem.

$$\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{N\times k},\mathbf{S}\in\mathbb{R}_{\geq 0}^{k\times M},\mathbf{B}\in\mathbb{R}_{\geq 0}^{P\times k}} \|\mathbf{W}\odot(\mathbf{X}-\mathbf{AS})\|_F^2+\lambda\|\mathbf{L}\odot(\mathbf{Y}-\mathbf{BS})\|_F^2$$

Problem Setup:

Problem Advantages:

(Semi)supervised NMF

Goal: Incorporate known label information into problem.

$$\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{N\times k},\mathbf{S}\in\mathbb{R}_{\geq 0}^{k\times M},\mathbf{B}\in\mathbb{R}_{\geq 0}^{P\times k}} \|\mathbf{W}\odot(\mathbf{X}-\mathbf{AS})\|_F^2+\lambda\|\mathbf{L}\odot(\mathbf{Y}-\mathbf{BS})\|_F^2$$

Problem Setup:

$\mathbf{Y}\in\{0,1\}_{\geq 0}^{P\times M}$: label matrix

Problem Advantages:

(Semi)supervised NMF

Goal: Incorporate known label information into problem.

$$\min_{\mathbf{A}\in\mathbb{R}^{N\times k}_{\geq 0},\mathbf{S}\in\mathbb{R}^{k\times M}_{\geq 0},\mathbf{B}\in\mathbb{R}^{P\times k}_{\geq 0}} \|\mathbf{W}\odot(\mathbf{X}-\mathbf{AS})\|_F^2+\lambda\|\mathbf{L}\odot(\mathbf{Y}-\mathbf{BS})\|_F^2$$

Problem Setup:

$\mathbf{Y}\in\{0,1\}^{P\times M}_{\geq 0}$: label matrix

P: number of classes

Problem Advantages:

(Semi)supervised NMF

Goal: Incorporate known label information into problem.

$$\min_{\mathbf{A}\in\mathbb{R}^{N\times k}_{\geq 0},\mathbf{S}\in\mathbb{R}^{k\times M}_{\geq 0},\mathbf{B}\in\mathbb{R}^{P\times k}_{\geq 0}} \|\mathbf{W}\odot(\mathbf{X}-\mathbf{AS})\|_F^2+\lambda\|\mathbf{L}\odot(\mathbf{Y}-\mathbf{BS})\|_F^2$$

Problem Setup:

$\mathbf{Y} \in \{0,1\}^{P\times M}_{\geq 0}$: label matrix

P: number of classes

$\mathbf{W} \in \{0,1\}^{N\times M}_{\geq 0}$: data indicator

Problem Advantages:

(Semi)supervised NMF

Goal: Incorporate known label information into problem.

$$\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{N\times k},\mathbf{S}\in\mathbb{R}_{\geq 0}^{k\times M},\mathbf{B}\in\mathbb{R}_{\geq 0}^{P\times k}} \|\mathbf{W}\odot(\mathbf{X}-\mathbf{AS})\|_F^2+\lambda\|\mathbf{L}\odot(\mathbf{Y}-\mathbf{BS})\|_F^2$$

Problem Setup:

$\mathbf{Y}\in\{0,1\}_{\geq 0}^{P\times M}$: label matrix
P: number of classes
$\mathbf{W}\in\{0,1\}_{\geq 0}^{N\times M}$: data indicator
$\mathbf{L}\in\{0,1\}_{\geq 0}^{P\times M}$: label indicator

Problem Advantages:

(Semi)supervised NMF

Goal: Incorporate known label information into problem.

$$\min_{\mathbf{A}\in\mathbb{R}_{\geq 0}^{N\times k},\mathbf{S}\in\mathbb{R}_{\geq 0}^{k\times M},\mathbf{B}\in\mathbb{R}_{\geq 0}^{P\times k}} \|\mathbf{W}\odot(\mathbf{X}-\mathbf{AS})\|_F^2+\lambda\|\mathbf{L}\odot(\mathbf{Y}-\mathbf{BS})\|_F^2$$

Problem Setup:

$\mathbf{Y}\in\{0,1\}_{\geq 0}^{P\times M}$: label matrix
P: number of classes
$\mathbf{W}\in\{0,1\}_{\geq 0}^{N\times M}$: data indicator
$\mathbf{L}\in\{0,1\}_{\geq 0}^{P\times M}$: label indicator
λ: user defined hyperparameter

Problem Advantages:

(Semi)supervised NMF

Goal: Incorporate known label information into problem.

$$\min_{\mathbf{A}\in\mathbb{R}^{N\times k}_{\geq 0},\mathbf{S}\in\mathbb{R}^{k\times M}_{\geq 0},\mathbf{B}\in\mathbb{R}^{P\times k}_{\geq 0}} \|\mathbf{W}\odot(\mathbf{X}-\mathbf{AS})\|_F^2 + \lambda\|\mathbf{L}\odot(\mathbf{Y}-\mathbf{BS})\|_F^2$$

Problem Setup:

$\mathbf{Y}\in\{0,1\}^{P\times M}_{\geq 0}$: label matrix
P: number of classes
$\mathbf{W}\in\{0,1\}^{N\times M}_{\geq 0}$: data indicator
$\mathbf{L}\in\{0,1\}^{P\times M}_{\geq 0}$: label indicator
λ: user defined hyperparameter

Problem Advantages:

- ▷ use of label information

(Semi)supervised NMF

Goal: Incorporate known label information into problem.

$$\min_{\mathbf{A}\in\mathbb{R}^{N\times k}_{\geq 0},\mathbf{S}\in\mathbb{R}^{k\times M}_{\geq 0},\mathbf{B}\in\mathbb{R}^{P\times k}_{\geq 0}} \|\mathbf{W}\odot(\mathbf{X}-\mathbf{AS})\|_F^2+\lambda\|\mathbf{L}\odot(\mathbf{Y}-\mathbf{BS})\|_F^2$$

Problem Setup:

$\mathbf{Y}\in\{0,1\}^{P\times M}_{\geq 0}$: label matrix
P: number of classes
$\mathbf{W}\in\{0,1\}^{N\times M}_{\geq 0}$: data indicator
$\mathbf{L}\in\{0,1\}^{P\times M}_{\geq 0}$: label indicator
λ: user defined hyperparameter

Problem Advantages:

- use of label information
- can extend multiplicative updates method to SSNMF

Hierarchical NMF

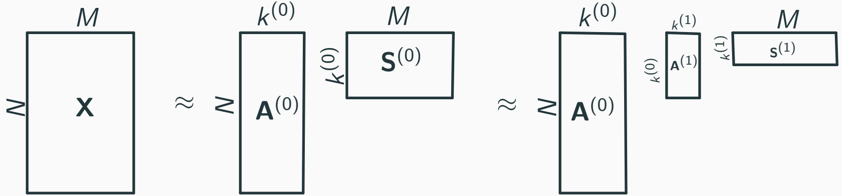
Goal: Discover hierarchical topic structure within $\mathbf{X}$.

Problem Setup:

Problem Challenges:

Hierarchical NMF

Goal: Discover hierarchical topic structure within $\mathbf{X}$.



$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{S}^{(0)}$

$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{A}^{(1)}\mathbf{S}^{(1)}$

$\vdots$

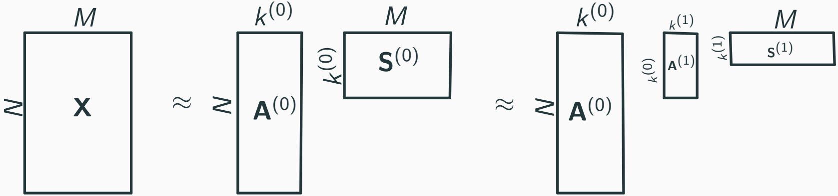
$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{A}^{(1)}\dots\mathbf{A}^{(\mathcal{L})}\mathbf{S}^{(\mathcal{L})}$

Problem Setup:

Problem Challenges:

Hierarchical NMF

Goal: Discover hierarchical topic structure within $\mathbf{X}$.



$$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{S}^{(0)}$$

$$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{A}^{(1)}\mathbf{S}^{(1)}$$

$$\vdots$$

$$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{A}^{(1)}\dots\mathbf{A}^{(\mathcal{L})}\mathbf{S}^{(\mathcal{L})}$$

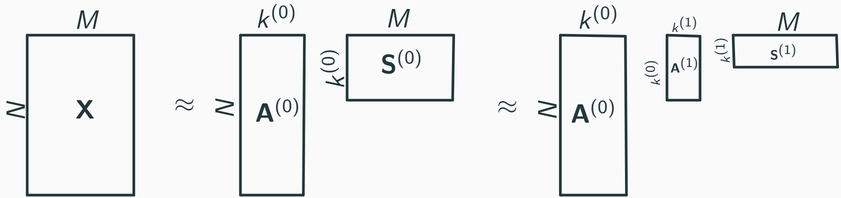
Problem Setup:

- $k^{(0)}, k^{(1)}, \dots, k^{(\mathcal{L})}$: user defined parameters

Problem Challenges:

Hierarchical NMF

Goal: Discover hierarchical topic structure within $\mathbf{X}$.



$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{S}^{(0)}$

$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{A}^{(1)}\mathbf{S}^{(1)}$

$\vdots$

$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{A}^{(1)}\dots\mathbf{A}^{(\mathcal{L})}\mathbf{S}^{(\mathcal{L})}$

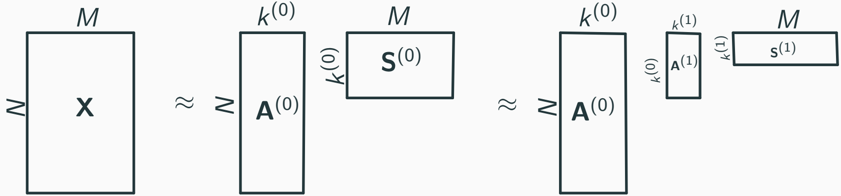
Problem Setup:

- $k^{(0)}, k^{(1)}, \dots, k^{(\mathcal{L})}$: user defined parameters
- $k^{(\ell)}$: supertopics collecting $k^{(\ell-1)}$ subtopics

Problem Challenges:

Hierarchical NMF

Goal: Discover hierarchical topic structure within $\mathbf{X}$.



$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{S}^{(0)}$

$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{A}^{(1)}\mathbf{S}^{(1)}$

$\vdots$

$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{A}^{(1)} \dots \mathbf{A}^{(\mathcal{L})}\mathbf{S}^{(\mathcal{L})}$

Problem Setup:

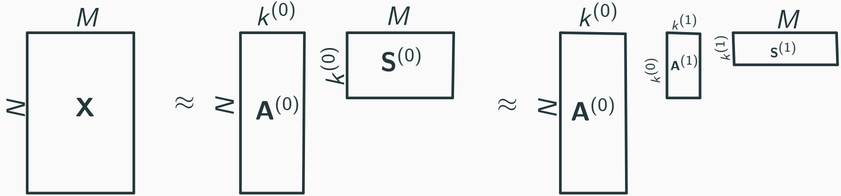
- $k^{(0)}, k^{(1)}, \dots, k^{(\mathcal{L})}$: user defined parameters
- $k^{(\ell)}$: supertopics collecting $k^{(\ell-1)}$ subtopics

Problem Challenges:

- $\{k^{(i)}\}$ must be chosen

Hierarchical NMF

Goal: Discover hierarchical topic structure within $\mathbf{X}$.



$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{S}^{(0)}$

$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{A}^{(1)}\mathbf{S}^{(1)}$

$\vdots$

$\mathbf{X} \approx \mathbf{A}^{(0)}\mathbf{A}^{(1)} \dots \mathbf{A}^{(\mathcal{L})}\mathbf{S}^{(\mathcal{L})}$

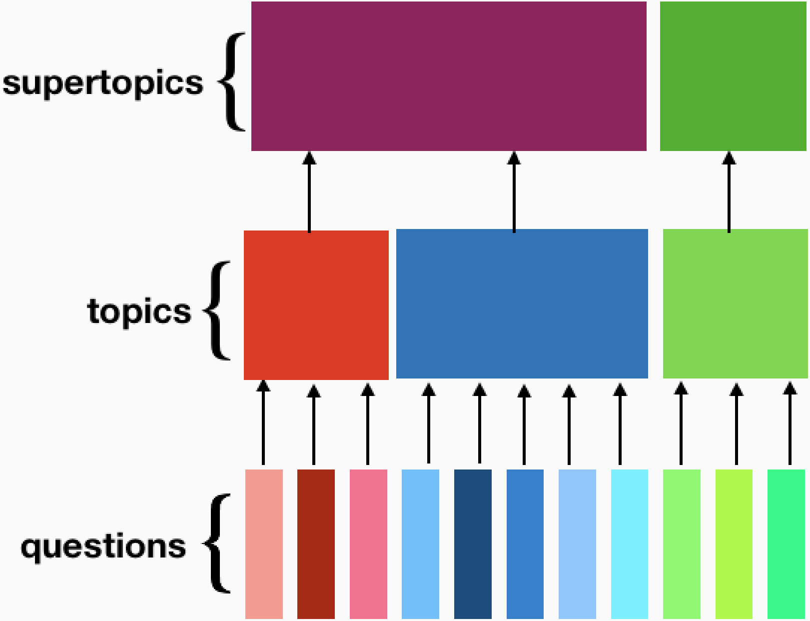
Problem Setup:

- $k^{(0)}, k^{(1)}, \dots, k^{(\mathcal{L})}$: user defined parameters
- $k^{(\ell)}$: supertopics collecting $k^{(\ell-1)}$ subtopics

Problem Challenges:

- $\{k^{(i)}\}$ must be chosen
- error propagates through layers

Hierarchical NMF



Deep NMF

Goal: Exploit similarities between neural networks and hierarchical NMF.

Deep NMF

Goal: Exploit similarities between neural networks and hierarchical NMF.

- ▷ [Flenner, Hunter '18]
  - introduces nonlinear pooling operator after each layer
  - introduces multiplicative updates method meant to backpropagate

Deep NMF

Goal: Exploit similarities between neural networks and hierarchical NMF.

▷ [Flenner, Hunter '18]

- introduces nonlinear pooling operator after each layer
- introduces multiplicative updates method meant to backpropagate

▷ [Trigeorgis, Bousmalis, Zafeiriou, Schuller '16]

- relaxes some of nonnegativity constraints in hNMF

Deep NMF

Goal: Exploit similarities between neural networks and hierarchical NMF.

▷ [Flenner, Hunter '18]
  - introduces nonlinear pooling operator after each layer
  - introduces multiplicative updates method meant to backpropagate

▷ [Trigeorgis, Bousmalis, Zafeiriou, Schuller '16]
  - relaxes some of nonnegativity constraints in hNMF

▷ [Le Roux, Hershey, Weninger '15]
  - introduces NMF backpropagation algorithm with "unfolding" (no hierarchy)

Deep NMF

Goal: Exploit similarities between neural networks and hierarchical NMF.

▷ [Flenner, Hunter '18]

- introduces nonlinear pooling operator after each layer
- introduces multiplicative updates method meant to backpropagate

▷ [Trigeorgis, Bousmalis, Zafeiriou, Schuller '16]

- relaxes some of nonnegativity constraints in hNMF

▷ [Le Roux, Hershey, Weninger '15]

- introduces NMF backpropagation algorithm with "unfolding" (no hierarchy)

▷ [Sun, Nasrabadi, Tran '17]

- similar method lacking nonnegativity constraints

Our method: Neural NMF

Goal: Develop true backpropagation algorithm for hNMF model.

Our method: Neural NMF

Goal: Develop true backpropagation algorithm for hNMF model.

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

Our method: Neural NMF

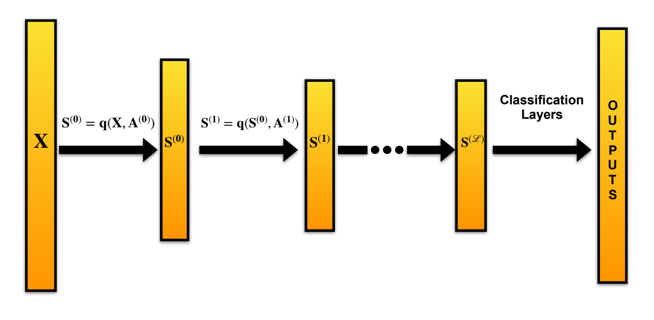
Goal: Develop true backpropagation algorithm for hNMF model.

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

Our method: Neural NMF

Goal: Develop true backpropagation algorithm for hNMF model.

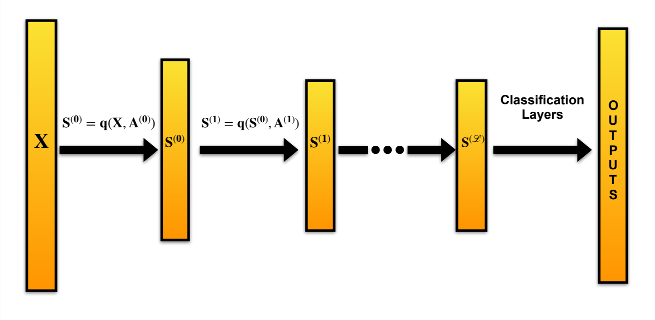
- Regard the **A** matrices as independent variables, determine the **S** matrices from the **A** matrices.
- Define $q(\mathbf{X}, \mathbf{A}) := \operatorname{argmin}_{S \geq 0} \|\mathbf{X} - \mathbf{AS}\|_F^2$.



Our method: Neural NMF

Goal: Develop true backpropagation algorithm for hNMF model.

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

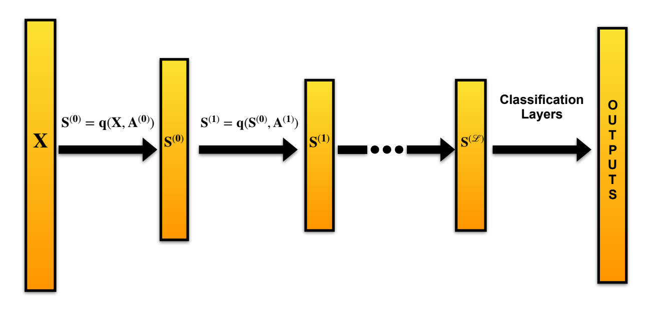


- ▷ Pin the values of **S** to those of **A** by recursively setting $\mathbf{S}^{(\ell)} := q(\mathbf{S}^{(\ell-1)}, \mathbf{A}^{(\ell)})$.

Our method: Neural NMF

Goal: Develop true backpropagation algorithm for hNMF model.

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



- ▷ Pin the values of **S** to those of **A** by recursively setting $\mathbf{S}^{(\ell)} := q(\mathbf{S}^{(\ell-1)}, \mathbf{A}^{(\ell)})$.
- ▷ Can we compute derivatives and backpropagate?

Neural NMF Backpropagation

Neural NMF Backpropagation

- Differentiate q function and apply chain rule.

Neural NMF Backpropagation

- Differentiate q function and apply chain rule.
- Flexible to cost function (e.g., supervision).

Neural NMF Backpropagation

- Differentiate q function and apply chain rule.
- Flexible to cost function (e.g., supervision).
- Backpropagate and update all $\mathbf{A}$ matrices simultaneously via GD or SGD.

Neural NMF

Method 1 Neural NMF

Require: data matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$, number of layers $\mathcal{L}$, step size γ, cost function C, initial matrices $\mathbf{A}^{(i)}$ for $i = 0, ..., \mathcal{L}$

procedure $\textsc{ForwardPropagation}(\mathbf{A}^{(0)}, \ldots, \mathbf{A}^{(\mathcal{L})})$

  for $i := 0 ... \mathcal{L}$ **do**

    $\mathbf{S}^{(i)} \leftarrow q(\mathbf{A}^{(i)}, \mathbf{S}^{(i-1)})$

ForwardPropagation$(\mathbf{A}^{(0)}, \ldots, \mathbf{A}^{(\mathcal{L})})$

while not converged **do**

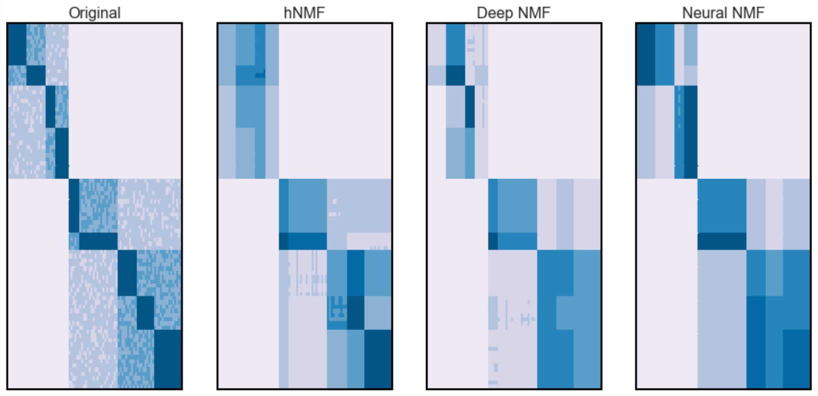
  for $i := 0 ... \mathcal{L}$ **do**

    $\mathbf{A}^{(i)} \leftarrow \mathbf{A}^{(i)} - \gamma * \frac{\partial C}{\partial \mathbf{A}^{(i)}}$ ▷ Gradient descent

    $\mathbf{A}^{(i)} \leftarrow \mathbf{A}^{(i)}_{+}$ ▷ Project onto positive orthant

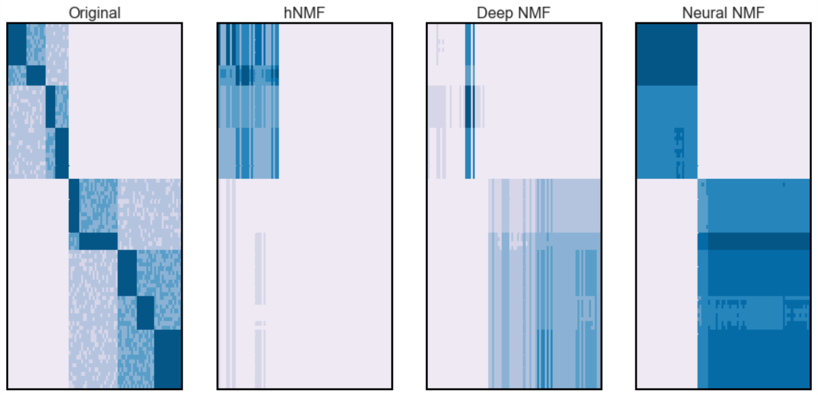
  ForwardPropagation$(\mathbf{A}^{(0)}, \ldots, \mathbf{A}^{(\mathcal{L})})$

Experimental Results



- unsupervised reconstruction with two-layer structure ($k^{(0)} = 9, k^{(1)} = 4$)

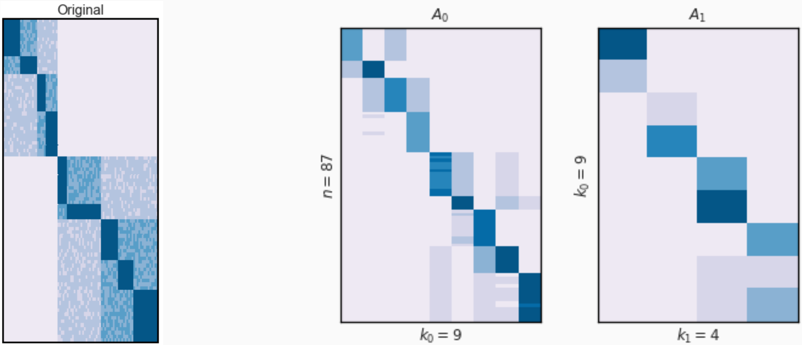
Experimental Results



- semisupervised reconstruction (40% labels) with three-layer structure ($k^{(0)} = 9, k^{(1)} = 4, k^{(2)} = 2$)

Experimental Results

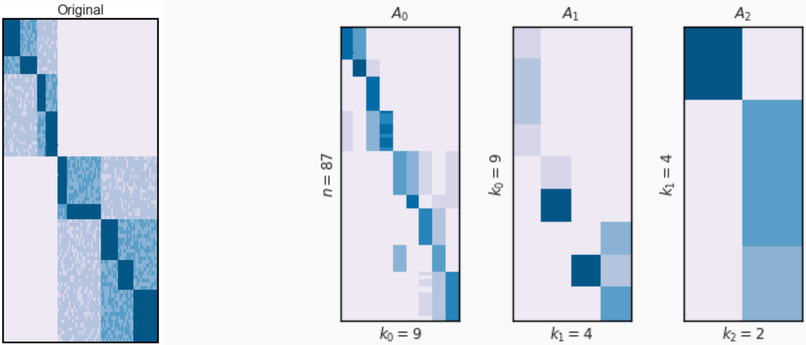
Note that despite reconstruction error increasing as layers increase (since the final rank decreases), the topic structure can be resolved from the intermediate factorizations.



- ▷ unsupervised reconstruction with two-layer structure ($k^{(0)} = 9, k^{(1)} = 4$)

Experimental Results

Note that despite reconstruction error increasing as layers increase (since the final rank decreases), the topic structure can be resolved from the intermediate factorizations.



- ▷ semisupervised reconstruction (40% labels) with three-layer structure ($k^{(0)} = 9, k^{(1)} = 4, k^{(2)} = 2$)

Experimental Results

Table 1: Reconstruction error / classification accuracy

	Layers	Hier. NMF	Deep NMF	Neural NMF
Unsuper.	1	0.053	0.031	0.029
	2	0.399	0.414	**0.310**
	3	0.860	0.838	**0.492**
Semisuper.	1	0.049 / 0.933	0.031 / 0.947	0.042 / **1**
	2	0.374 / 0.926	0.394 / 0.911	**0.305** / **1**
	3	0.676 / 0.930	0.733 / 0.930	**0.496** / **0.990**
Supervised	1	0.052 / 0.960	0.042 / 0.962	0.042 / **1**
	2	0.311 / 0.984	0.310 / 0.984	0.307 / 1
	3	0.495 / 1	0.494 / 1	0.498 / 1

Conclusions and Future Work

Conclusions and Future Work

- presented a novel method for multilayer NMF that incorporates the backpropagation technique from deep learning to minimize error accumulation

Conclusions and Future Work

- presented a novel method for multilayer NMF that incorporates the backpropagation technique from deep learning to minimize error accumulation
- exhibited preliminary tests on toy datasets showing the proposed method outperforms existing multilayer NMF algorithms.

Conclusions and Future Work

- presented a novel method for multilayer NMF that incorporates the backpropagation technique from deep learning to minimize error accumulation
- exhibited preliminary tests on toy datasets showing the proposed method outperforms existing multilayer NMF algorithms.

- compare our method and others on various datasets to find precise regimes in which we offer improvement

Conclusions and Future Work

- presented a novel method for multilayer NMF that incorporates the backpropagation technique from deep learning to minimize error accumulation
- exhibited preliminary tests on toy datasets showing the proposed method outperforms existing multilayer NMF algorithms.

- compare our method and others on various datasets to find precise regimes in which we offer improvement
- extend to method for hierarchical nonnegative tensor factorization

Thanks for listening!

Questions?

[1] Jennifer Flenner and Blake Hunter. **A deep non-negative matrix factorization neural network, 2018.** Unpublished.

[2] Jonathan Le Roux, John R Hershey, and Felix Weninger. **Deep nmf for speech separation.** In Int. Conf. Acoust. Spee., pages 66–70. IEEE, 2015.

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

[4] H. Lee, J. Yoo, and S. Choi. **Semi-supervised nonnegative matrix factorization.** IEEE Signal Proc. Let., 17(1):4–7, Jan 2010.

[5] Xiaoxia Sun, Nasser M. Nasrabadi, and Trac D. Tran. **Supervised multilayer sparse coding networks for image classification.** CoRR, abs/1701.08349, 2017.

[6] George Trigeorgis, Konstantinos Bousmalis, Stefanos Zafeiriou, and Björn W Schuller. **A deep matrix factorization method for learning attribute representations.** IEEE T. Pattern Anal., 39(3):417–429, 2016.