Hierarchical and Neural Nonnegative Tensor Decompositions

by Jamie Haddock
(Harvey Mudd College, Department of Mathematics)
on December 2, 2022,
IPAM "Multi-Modal Imaging with Deep Learning and Modeling"

https://ieeexplore.ieee.org/document/9022678 (CAMSAP 2019)
joint with M. Gao•, D. Molitor, E. Sadovnik, T. Will•, R. Zhang•, D. Needell

https://ieeexplore.ieee.org/document/9723126 (ACSSC 2021)
joint with Joshua Vendrow•, Deanna Needell

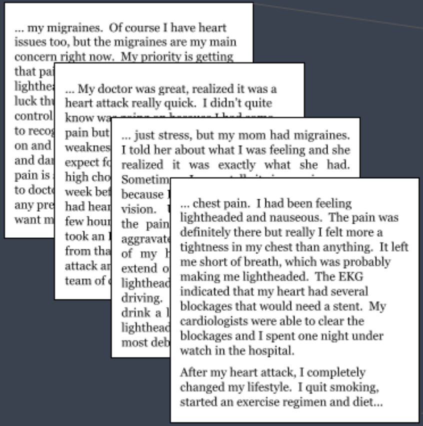
https://ieeexplore.ieee.org/document/9747810 (ICASSP 2022)
joint with Joshua Vendrow•, Deanna Needell

NSF DMS #2211318

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

Motivation

» Learn trends in high-dimensional data



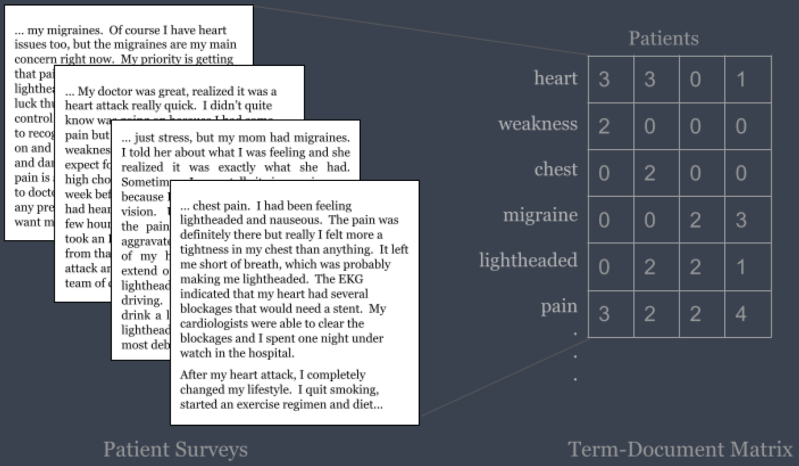
Patient Surveys

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

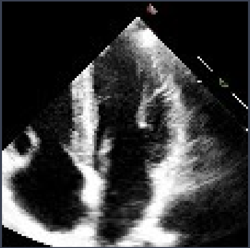
» Learn trends in high-dimensional data

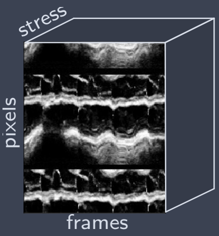


Understand symptom trends and shared patient experiences automatically.

» Learn trends in high-dimensional data

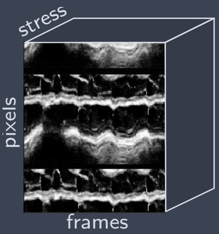
» Learn trends in high-dimensional data





» Learn trends in high-dimensional data

Learn cohesive parts and separate noise in medical image studies.



Can we tell how the resulting parts/topics are related?

Can we tell how the resulting parts/topics are related?

How do we choose the number of topics or parts to learn?

Can we tell how the resulting parts/topics are related?

How do we choose the number of topics or parts to learn?

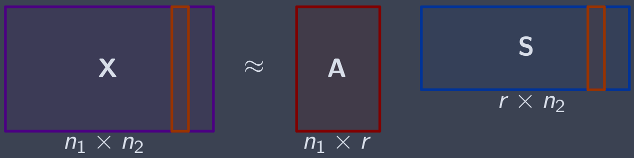
Hierarchical matrix factorization and tensor decomposition topic models!

Introduction

» Nonnegative Matrix Factorization (NMF)

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

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

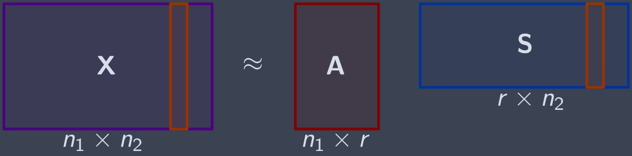


Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.

» Nonnegative Matrix Factorization (NMF)

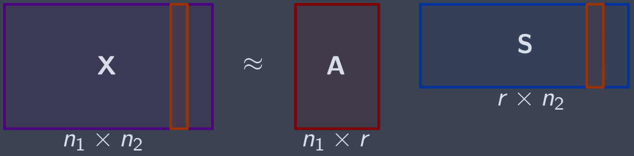
Model: Given nonnegative data $\mathbf{X}$, compute nonnegative $\mathbf{A}$ and $\mathbf{S}$ of lower rank so that

$$\mathbf{X} \approx \mathbf{A}\mathbf{S}.$$



Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.

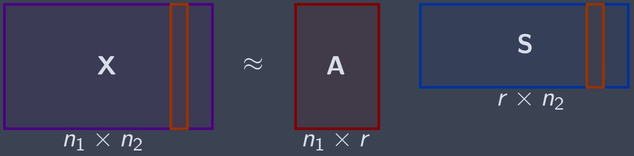
» Nonnegative Matrix Factorization (NMF)



▷ Employed for dimensionality-reduction and topic modeling

Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.

» Nonnegative Matrix Factorization (NMF)

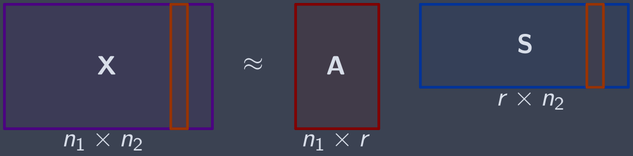


- ▷ Employed for dimensionality-reduction and topic modeling
- ▷ Often formulated as

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

[1] Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.

» Nonnegative Matrix Factorization (NMF)



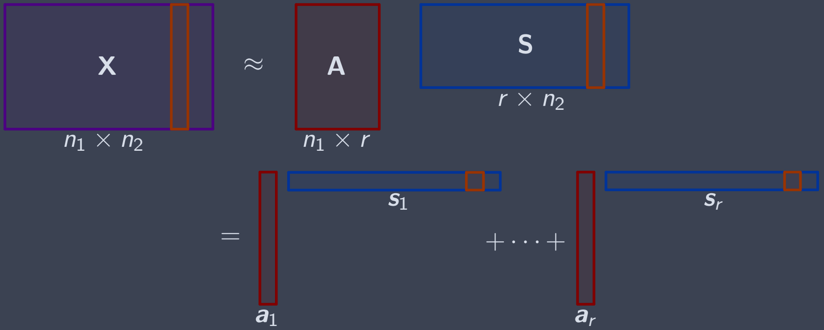
- Employed for dimensionality-reduction and topic modeling
- Often formulated as

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

- non-convex optimization problems

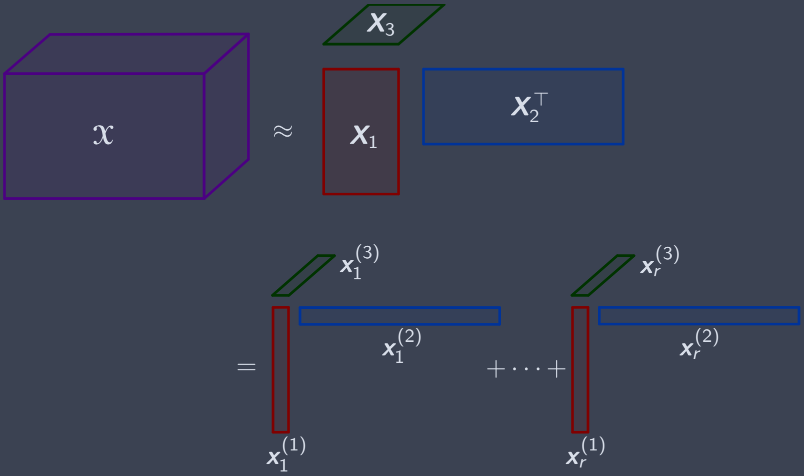
Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.

» Nonnegative Matrix Factorization (NMF)



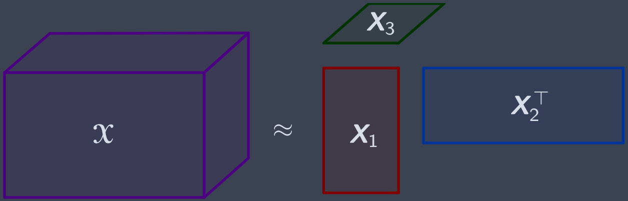
Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.

» Nonnegative CANDECOMP/PARAFAC (CP) decomposition (NCPD)



Carroll, J. Douglas, and Jih-Jie Chang. "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition." Psychometrika 35.3 (1970): 283-319.
Harshman, Richard A. "Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis." (1970): 1-84.

» Nonnegative CANDECOMP/PARAFAC (CP) decomposition (NCPD)



- ▷ formulated as $\min_{X_i \geq 0} \|\mathcal{X} - [\![X_1, X_2, \cdots, X_k]\!]\|_F^2$ where

$$[\![X_1, X_2, \cdots, X_k]\!] \equiv \sum_{j=1}^{r} x_j^{(1)} \otimes x_j^{(2)} \otimes \cdots \otimes x_j^{(k)}$$

Carroll, J. Douglas, and Jih-Jie Chang. "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition." Psychometrika 35.3 (1970): 283-319.
Harshman, Richard A. "Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis." (1970): 1-84.

» Hierarchical NMF

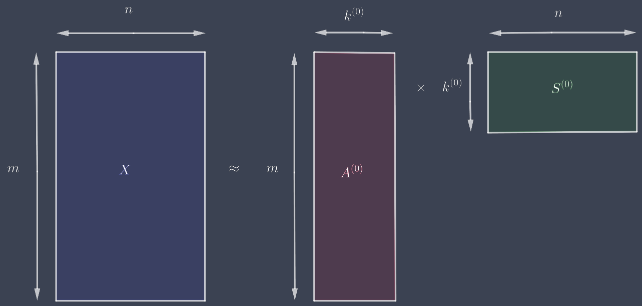
Model: Sequentially factorize

Cichocki, Andrzej, and Rafal Zdunek. "Multilayer nonnegative matrix factorisation." ELECTRONICS LETTERS-IEE 42.16 (2006): 947.

» Hierarchical NMF

Model: Sequentially factorize

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

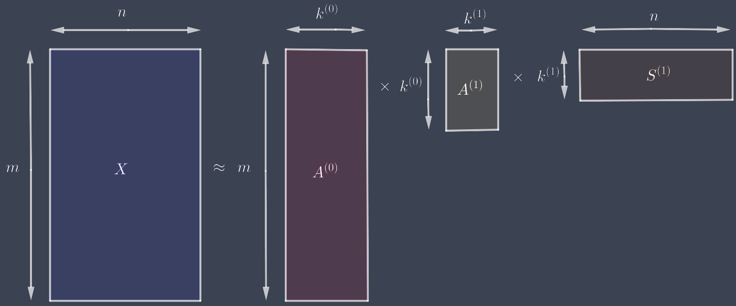


Cichocki, Andrzej, and Rafal Zdunek. "Multilayer nonnegative matrix factorisation." ELECTRONICS LETTERS-IEE 42.16 (2006): 947.

» Hierarchical NMF

Model: Sequentially factorize

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

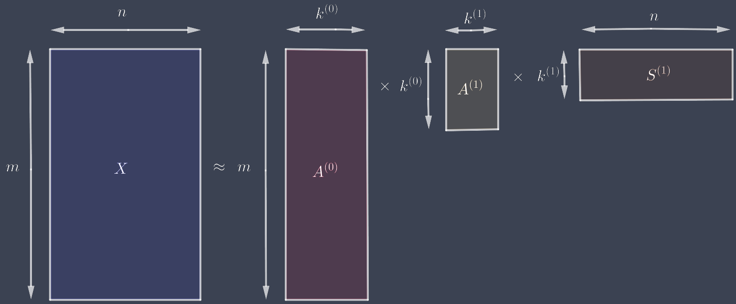


Cichocki, Andrzej, and Rafal Zdunek. "Multilayer nonnegative matrix factorisation." ELECTRONICS LETTERS-IEE 42.16 (2006): 947.

» Hierarchical NMF

Model: Sequentially factorize

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



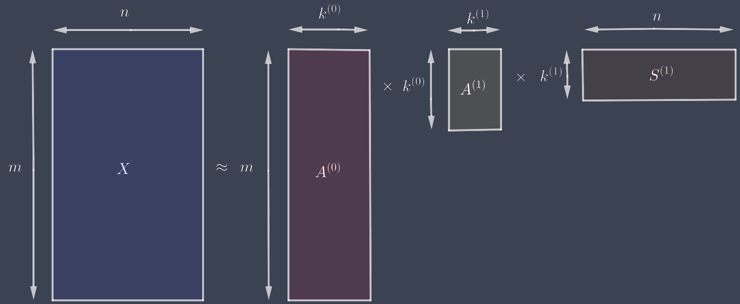
▷ $k^{(\ell)}$: supertopics collecting $k^{(\ell-1)}$ subtopics

Cichocki, Andrzej, and Rafal Zdunek. "Multilayer nonnegative matrix factorisation." ELECTRONICS LETTERS-IEE 42.16 (2006): 947.

» Hierarchical NMF

Model: Sequentially factorize

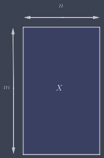
$$X \approx A^{(0)}S^{(0)}, S^{(0)} \approx A^{(1)}S^{(1)}, S^{(1)} \approx A^{(2)}S^{(2)}, ..., S^{(\mathcal{L}-1)} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}.$$



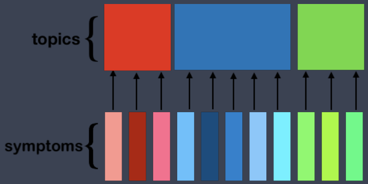
- $k^{(\ell)}$: supertopics collecting $k^{(\ell-1)}$ subtopics
- provides relationship between data matrix modes and $k^{(\ell)}$ topics

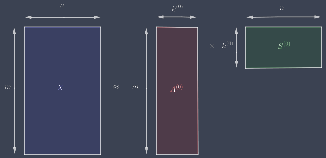
Cichocki, Andrzej, and Rafal Zdunek. "Multilayer nonnegative matrix factorisation." ELECTRONICS LETTERS-IEE 42.16 (2006): 947.

» Hierarchical NMF

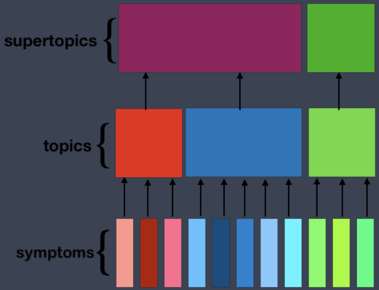


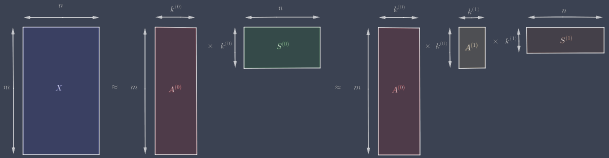
» Hierarchical NMF



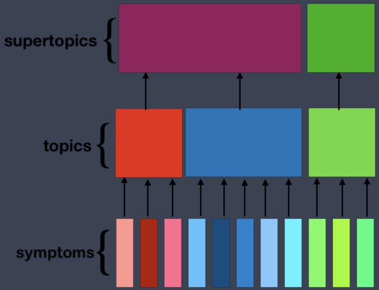


» Hierarchical NMF

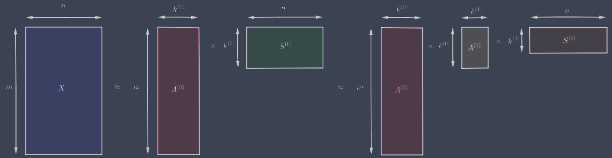




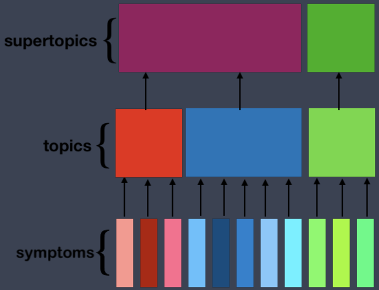
» Hierarchical NMF



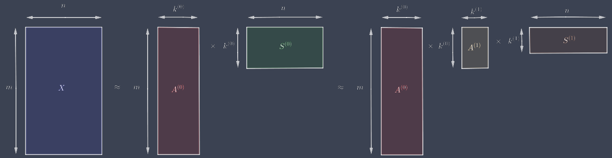
- ▷ elucidates the hierarchical relationships of learned topics



» Hierarchical NMF



- ▷ elucidates the hierarchical relationships of learned topics
- ▷ no need to choose a fixed model rank (number of topics)



Hierarchical Models

» Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

» Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

- Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.

» Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

- ▷ Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- ▷ Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- ▷ Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.

- ▷ Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization." International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

» Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

- Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.

- Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization." International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

Results depend upon hyperparameter choice (mode).

» Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

- ▷ Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- ▷ Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- ▷ Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.

- ▷ Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization." International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

Results depend upon hyperparameter choice (mode).

- * Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

» Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

- ▷ Vasilescu, M. Alex O., and Eric Kim. “Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors.” arXiv preprint arXiv:1911.04180 (2019).
- ▷ Song, Le, et al. “Hierarchical tensor decomposition of latent tree graphical models.” International Conference on Machine Learning. PMLR, 2013.
- ▷ Grasedyck, Lars. “Hierarchical singular value decomposition of tensors.” SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.

- ▷ Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. “Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization.” International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

Results depend upon hyperparameter choice (mode).

- * Vendrow, Joshua, Jamie Haddock, and Deanna Needell. “Neural nonnegative CP decomposition for hierarchical tensor analysis.” 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

Not a single hierarchical relationship, good training method.

» Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

- ▷ Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- ▷ Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- ▷ Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.

- ▷ Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization." International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

Results depend upon hyperparameter choice (mode).

- * Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

Not a single hierarchical relationship, good training method.

- * Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

» Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

- ▷ Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- ▷ Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- ▷ Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.

- ▷ Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization." International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

Results depend upon hyperparameter choice (mode).

- * Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

Not a single hierarchical relationship, good training method.

- * Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

Single hierarchical relationship, naive training method.

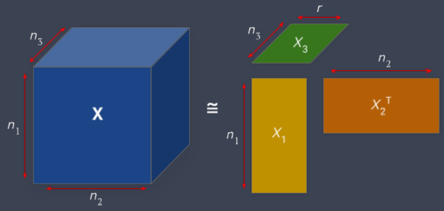
» Hierarchical NCPD Model (Take 1)

Vendrow, H., Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

» Hierarchical NCPD Model (Take 1)

Learn an initial rank-r NCPD model,

$$\mathcal{X} \approx [\![\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_k]\!]$$



Vendrow, H., Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

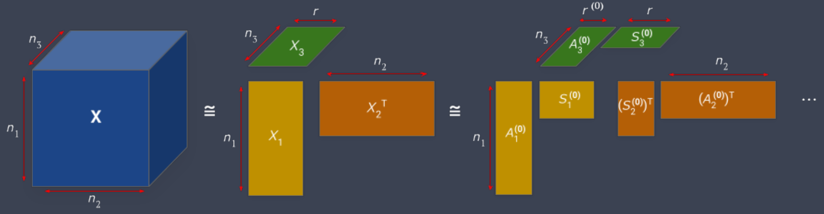
» Hierarchical NCPD Model (Take 1)

Learn an initial rank-r NCPD model,

$$\mathcal{X} \approx [\![\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_k]\!]$$

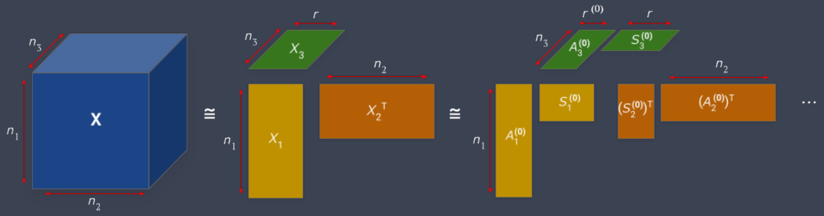
and apply a hierarchical NMF model independently to each factor matrix,

$$\boldsymbol{X}_i \approx \boldsymbol{A}_i^{(0)} \boldsymbol{A}_i^{(1)} \cdots \boldsymbol{A}_i^{(l)} \boldsymbol{S}_i^{(l)}.$$



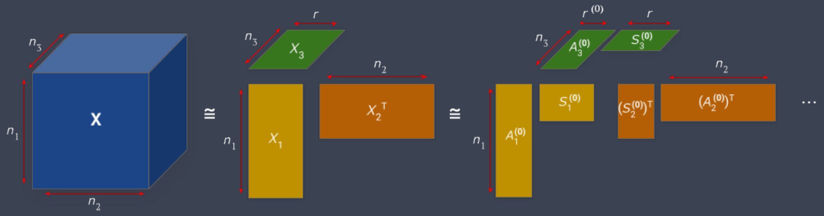
Vendrow, H., Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

» Hierarchical NCPD Model (Take 1)



Vendrow, H., Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

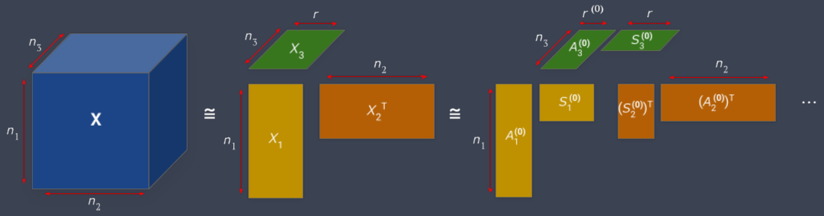
» Hierarchical NCPD Model (Take 1)



- ▷ can extend good training method for HNMF (Neural NMF) → Neural NCPD (later in this talk!)

Vendrow, H., Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

» Hierarchical NCPD Model (Take 1)



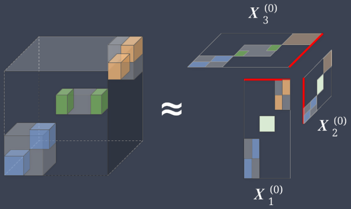
- ▷ can extend good training method for HNMF (Neural NMF) $\rightarrow$ Neural NCPD (later in this talk!)
- ▷ `Different hierarchy across tensor modes.  :(`

Vendrow, H., Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

» Multi-HNTF Model (Take 2)

This model learns

$$\mathcal{X} \approx [\![X_1^{(0)}, X_2^{(0)}, \cdots, X_k^{(0)}]\!]$$



Vendrow, H., Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

» Multi-HNTF Model (Take 2)

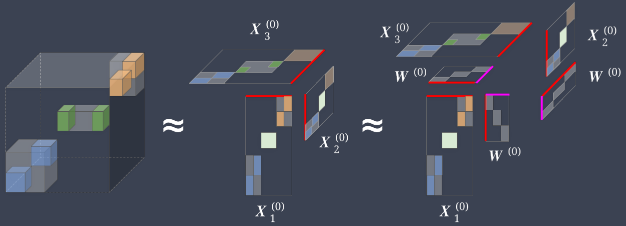
This model learns

$$\mathcal{X} \approx [\![X_1^{(0)}, X_2^{(0)}, \cdots, X_k^{(0)}]\!] \approx [\![X_1^{(1)}, X_2^{(1)}, \cdots, X_k^{(1)}]\!]$$

where

$$X_j^{(\ell+1)} = X_j^{(\ell)} W^{(\ell)},$$

and $W^{(\ell)} \in \mathbb{R}_{\geq 0}^{r^{(\ell-1)} \times r^{(\ell)}}$.



Vendrow, H., Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

» Multi-HNTF Model (Take 2)

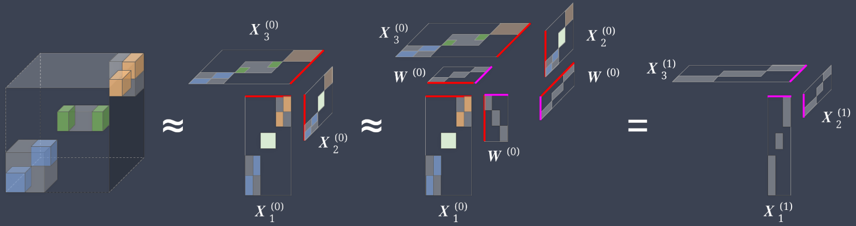
This model learns

$$\mathcal{X} \approx [\![X_1^{(0)}, X_2^{(0)}, \cdots, X_k^{(0)}]\!] \approx [\![X_1^{(1)}, X_2^{(1)}, \cdots, X_k^{(1)}]\!]$$

where

$$X_j^{(\ell+1)} = X_j^{(\ell)} W^{(\ell)},$$

and $W^{(\ell)} \in \mathbb{R}_{\geq 0}^{r^{(\ell-1)} \times r^{(\ell)}}$.



Vendrow, H., Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

» Multi-HNTF Model (Take 2)

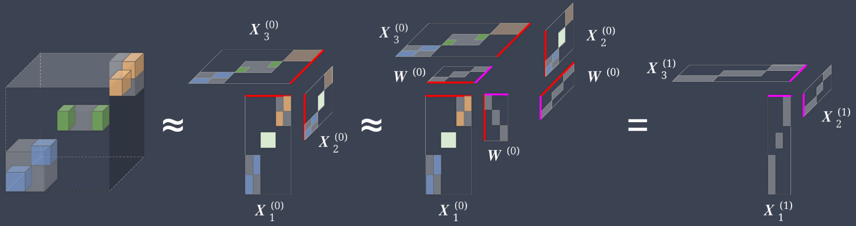
This model learns

$$\mathcal{X} \approx [\![X_1^{(0)}, X_2^{(0)}, \dots, X_k^{(0)}]\!] \approx [\![X_1^{(1)}, X_2^{(1)}, \dots, X_k^{(1)}]\!] \approx \cdots$$
$$\approx [\![X_1^{(\mathcal{L}-1)}, X_2^{(\mathcal{L}-1)}, \dots, X_k^{(\mathcal{L}-1)}]\!]$$

where

$$X_j^{(\ell+1)} = X_j^{(\ell)} W^{(\ell)},$$

and $W^{(\ell)} \in \mathbb{R}_{\geq 0}^{r^{(\ell-1)} \times r^{(\ell)}}$.



Vendrow, H., Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

» Multi-HNTF Model (Take 2)

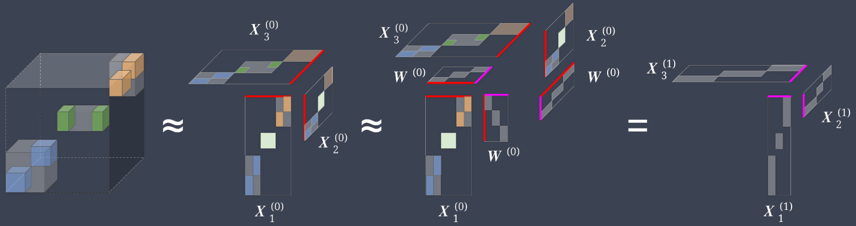
This model learns

$$\mathcal{X} \approx [\![X_1^{(0)}, X_2^{(0)}, \cdots, X_k^{(0)}]\!] \approx [\![X_1^{(1)}, X_2^{(1)}, \cdots, X_k^{(1)}]\!] \approx \cdots$$
$$\approx [\![X_1^{(\mathcal{L}-1)}, X_2^{(\mathcal{L}-1)}, \cdots, X_k^{(\mathcal{L}-1)}]\!]$$

where

$$X_j^{(\ell+1)} = X_j^{(\ell)} W^{(\ell)},$$

and $W^{(\ell)} \in \mathbb{R}_{\geq 0}^{r^{(\ell-1)} \times r^{(\ell)}}$.



A single hierarchical relationship for all modes!

Vendrow, H., Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

» Training Process

procedure $\textsc{Multi-HNTF}(\mathcal{X})$

$\quad \{X_i^{(0)}\}_{i=1}^{k} \leftarrow \mathrm{NCPD}(\mathcal{X}, r_0)$

$\quad$ **for** $\ell = 0 \dots \mathcal{L}$ **do**

$\quad\quad W^{(\ell)} \leftarrow \operatorname{argmin}_{W \in \mathbb{R}_+^{r_\ell \times r_{\ell+1}}} \|\mathcal{X} - [\![X_1^{(\ell)} W, \dots, X_k^{(\ell)} W]\!]\|$

$\quad\quad$ **for** $i = 0 \dots k$ **do**

$\quad\quad\quad X_i^{(\ell+1)} = X_i^{(\ell)} W^{(\ell)}$

» Training Process

procedure MULTI-HNTF($\mathcal{X}$)
  $\{\boldsymbol{X}_i^{(0)}\}_{i=1}^{k} \leftarrow \text{NCPD}(\mathcal{X}, r_0)$
  for $\ell = 0 \ldots \mathcal{L}$ **do**
    $\boldsymbol{W}^{(\ell)} \leftarrow \text{argmin}_{\boldsymbol{W} \in \mathbb{R}_+^{r_\ell \times r_{\ell+1}}} \|\mathcal{X} - [\![\boldsymbol{X}_1^{(\ell)}\boldsymbol{W}, \ldots, \boldsymbol{X}_k^{(\ell)}\boldsymbol{W}]\!]\|$
    for $i = 0 \ldots k$ **do**
      $\boldsymbol{X}_i^{(\ell+1)} = \boldsymbol{X}_i^{(\ell)}\boldsymbol{W}^{(\ell)}$

- ▷ Can be approximated via NMF method on each mode with averaging of learned $\boldsymbol{W}$ matrix across modes.

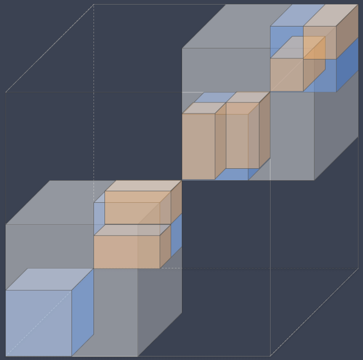
» Training Process

1: **procedure** MULTI-HNTF($\mathcal{X}$)
2: $\{X_i^{(0)}\}_{i=1}^k \leftarrow \text{NCPD}(\mathcal{X}, r_0)$
3: **for** $\ell = 0 \dots \mathcal{L}$ **do**
4: $W^{(\ell)} \leftarrow \text{argmin}_{W \in \mathbb{R}_+^{r_\ell \times r_{\ell+1}}} \|\mathcal{X} - [\![X_1^{(\ell)} W, \dots, X_k^{(\ell)} W]\!]\|$
5: **for** $i = 0 \dots k$ **do**
6: $X_i^{(\ell+1)} = X_i^{(\ell)} W^{(\ell)}$

- ▷ Can be approximated via NMF method on each mode with averaging of learned W matrix across modes.
- ▷ Could/should also be trained in a neural network framework.

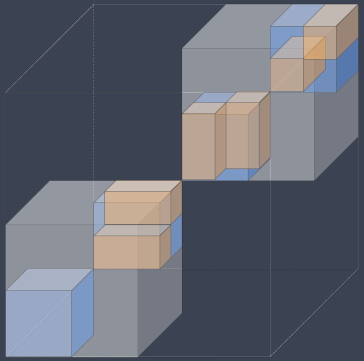
Experiments

» Synthetic Tensor



The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.

» Synthetic Tensor

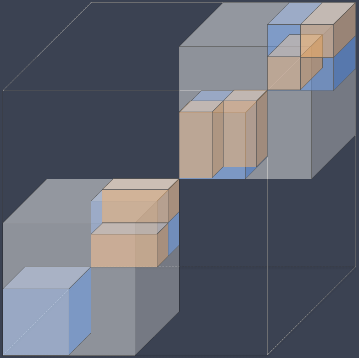


The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.

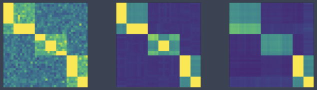
Relative reconstruction error.

Method	$r_0 = 7$	$r_1 = 4$	$r_2 = 2$
Multi-HNTF	0.454	0.548	0.721
~~Neural HNCPD [Vendrow, et. al.]~~	~~0.454~~	~~**0.508**~~	~~**0.714**~~
Standard HNCPD [Vendrow, et. al.]	0.454	0.612	0.892
HNTF-1 [Cichocki, et. al.]	0.454	0.576	0.781
HNTF-2 [Cichocki, et. al.]	0.454	0.587	0.765
HNTF-3 [Cichocki, et. al.]	0.454	0.560	0.747

» Synthetic Tensor



The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.



Projections of tensor approximation at each layer of Multi-HNTF.

Relative reconstruction error.

Method	$r_0 = 7$	$r_1 = 4$	$r_2 = 2$
Multi-HNTF	0.454	0.548	0.721
~~Neural HNCPD [Vendrow, et. al.]~~	~~0.454~~	~~**0.508**~~	~~**0.714**~~
Standard HNCPD [Vendrow, et. al.]	0.454	0.612	0.892
HNTF-1 [Cichocki, et. al.]	0.454	0.576	0.781
HNTF-2 [Cichocki, et. al.]	0.454	0.587	0.765
HNTF-3 [Cichocki, et. al.]	0.454	0.560	0.747

» Political Twitter Data

- A data set of tweets sent by political candidates during the 2016 election season
- We subset the tweets from eight politicians, four Republicans and four Democrats:

  Hillary Clinton, Tim Kaine, Martin O'Malley, Bernie Sanders, Ted Cruz, John Kasich, Marco Rubio, and Donald Trump.

» Political Twitter Data

Rank 8 Topics

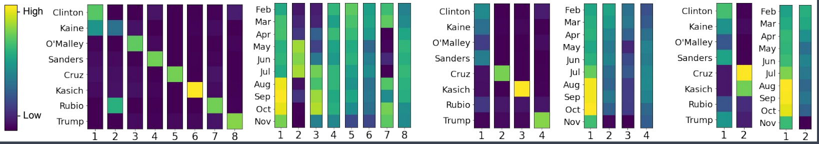
Topic 1	Topic 2	Topic 3	Topic 4
trump	senate	martinomalley	berniesanders
hillary	florida	hillaryclinton	people
donald	zika	realdonaldtrump	bernie
president	venezuela	campaigning	must
timkaine	nicolasmaduro	maryland	change
Topic 5	**Topic 6**	**Topic 7**	**Topic 8**
tedcruz	johnkasich	marcorubio	crooked
cruz	kasich	teammarco	hillary
ted	ohio	vote	thank
internet	john	flsen	great
choosecruz	gov	click	clinton

Rank 4 Topics

Topic 1	Topic 2
trump	tedcruz
hillary	cruz
vote	ted
people	internet
berniesanders	choosecruz
Topic 3	**Topic 4**
johnkasich	crooked
kasich	hillary
ohio	thank
john	great
gov	realdonaldtrump

Rank 2 Topics

Topic 1
trump
hillary
people
vote
realdonaldtrump
Topic 2
tedcruz
cruz
ted
johnkasich
kasich



» Political Twitter Data

Rank 8 Topics

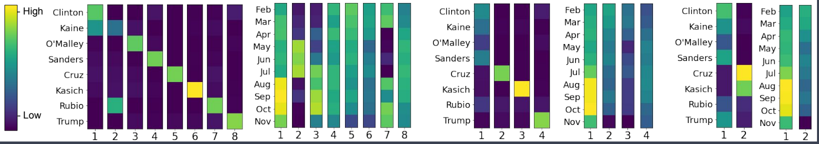
Topic 1	Topic 2	Topic 3	Topic 4
trump	senate	martinomalley	berniesanders
hillary	florida	hillaryclinton	people
donald	zika	realdonaldtrump	bernie
president	venezuela	campaigning	must
timkaine	nicolasmaduro	maryland	change
Topic 5	**Topic 6**	**Topic 7**	**Topic 8**
tedcruz	johnkasich	marcorubio	crooked
cruz	kasich	teammarco	hillary
ted	ohio	vote	thank
internet	john	flsen	great
choosecruz	gov	click	clinton

Rank 4 Topics

Topic 1	Topic 2
trump	tedcruz
hillary	cruz
vote	ted
people	internet
berniesanders	choosecruz
Topic 3	**Topic 4**
johnkasich	crooked
kasich	hillary
ohio	thank
john	great
gov	realdonaldtrump

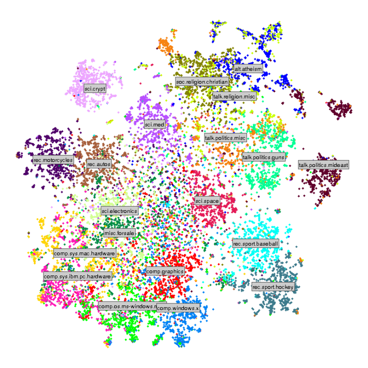
Rank 2 Topics

Topic 1
trump
hillary
people
vote
realdonaldtrump
Topic 2
tedcruz
cruz
ted
johnkasich
kasich

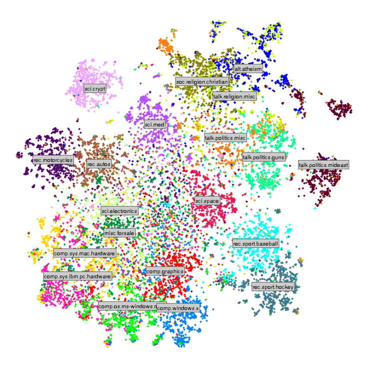


Method	$r_0 = 8$	$r_1 = 4$	$r_2 = 2$
Multi-HNTF	0.834	0.887	0.920
~~Neural HNCPD [Vendrow, et. al.]~~	~~0.834~~	~~**0.883**~~	~~**0.918**~~
Standard HNCPD [Vendrow, et. al.]	0.834	0.889	0.919
Standard NCPD	0.834	0.931	0.950
HNTF-1 [Cichocki, et. al.]	0.834	0.890	0.927
HNTF-2 [Cichocki, et. al.]	0.834	0.909	0.956
HNTF-3 [Cichocki, et. al.]	0.834	0.895	0.942

» 20 Newsgroups Data



» 20 Newsgroups Data



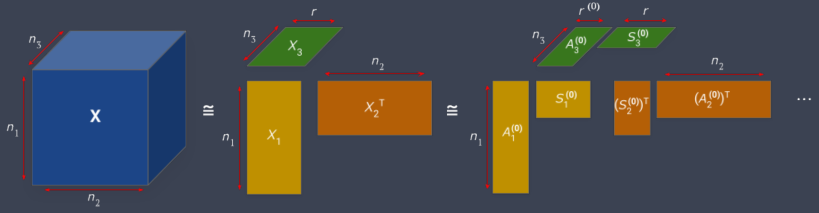
Reconstruction loss and classification accuracy at the second layer of two layer Multi-HNTF and HNMF on the 20 newsgroup data set.

Method	Recon Loss		Accuracy	
	Unsup.	Sup.	Unsup.	Sup.
Multi-HNTF	**30.81**	**30.91**	**0.516**	**0.737**
HNMF	30.82	31.45	0.507	0.636

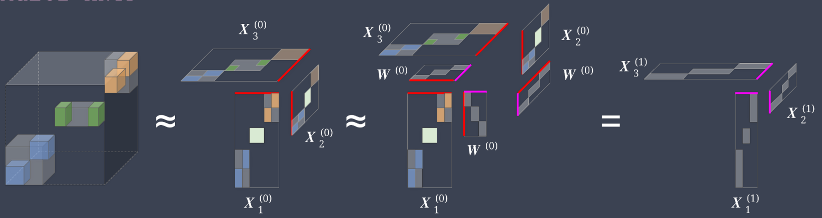
Backpropagation

» Hierarchical Tensor Decompositions

Hierarchical NCPD

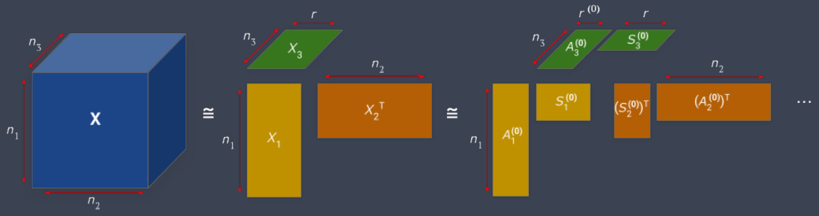


Multi-HNTF

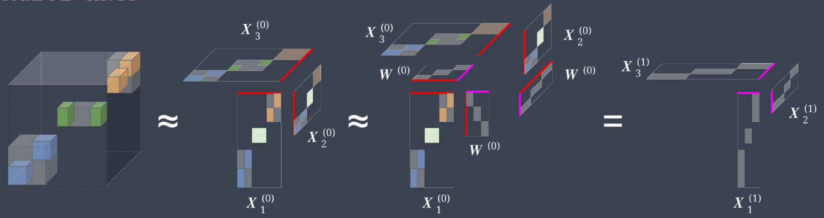


» Hierarchical Tensor Decompositions

Hierarchical NCPD



Multi-HNTF

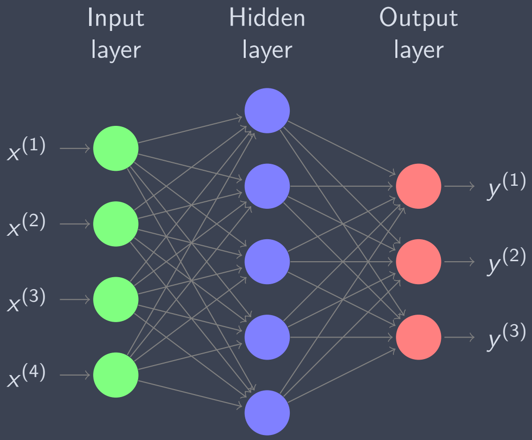


Devastating error propagation through layers!

» Reminder

Neural Network: Learn weights $W^{(1)}, W^{(2)}, ..., W^{(L)}$ to minimize model error

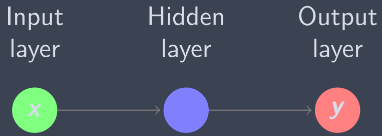
$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} f(\mathbf{y}(\mathbf{x}_n, \{W^{(i)}\}), \mathbf{x}_n, \mathbf{t}_n).$$



» Reminder

Neural Network: Learn weights $W^{(1)}, W^{(2)}, \ldots, W^{(L)}$ to minimize model error

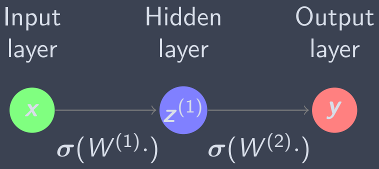
$$E(\{W^{(i)}\}) = \sum_{n=1}^{N} f(\boldsymbol{y}(\boldsymbol{x}_n, \{W^{(i)}\}), \boldsymbol{x}_n, \boldsymbol{t}_n).$$



» Reminder

Neural Network: Learn weights $W^{(1)}, W^{(2)}, \dots, W^{(L)}$ to minimize model error

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



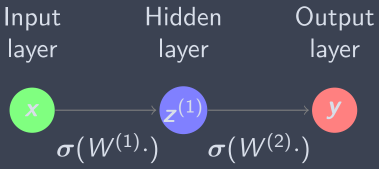
Training:

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

» Reminder

Neural Network: Learn weights $W^{(1)}, W^{(2)}, \dots, W^{(L)}$ to minimize model error

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

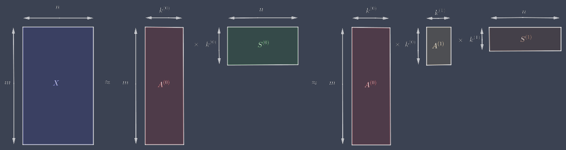


Training:

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

» Training via backpropagation

Neural NMF: Forward and back propagation algorithms for hNMF.

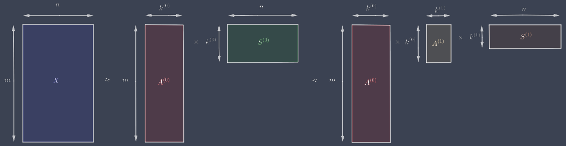


Gao, Mengdi, et al. "Neural nonnegative matrix factorization for hierarchical multilayer topic modeling." 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019.

Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Training via backpropagation

Neural NMF: Forward and back propagation algorithms for hNMF.



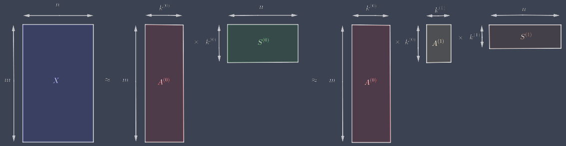
- ▷ Regard the A matrices as independent variables, determine the S matrices from the A matrices.

Gao, Mengdi, et al. “Neural nonnegative matrix factorization for hierarchical multilayer topic modeling.” 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019.

Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Training via backpropagation

Neural NMF: Forward and back propagation algorithms for hNMF.



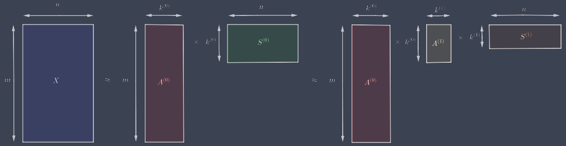
- ▷ Regard the A matrices as independent variables, determine the S matrices from the A matrices.
- ▷ Define $q(X, A) := \text{argmin}_{S \geq 0} \|X - AS\|_F^2$ (least-squares).

Gao, Mengdi, et al. "Neural nonnegative matrix factorization for hierarchical multilayer topic modeling." 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019.

Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Training via backpropagation

Neural NMF: Forward and back propagation algorithms for hNMF.



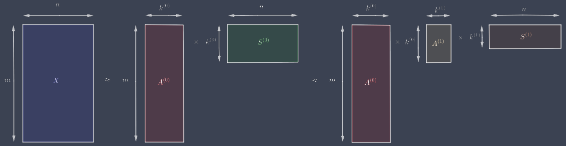
- ▷ Regard the A matrices as independent variables, determine the S matrices from the A matrices.
- ▷ Define $q(X, A) := \text{argmin}_{S \geq 0} \|X - AS\|_F^2$ (least-squares).
- ▷ Pin the values of S to those of A by recursively setting $S^{(\ell)} := q(S^{(\ell-1)}, A^{(\ell)})$.

Gao, Mengdi, et al. "Neural nonnegative matrix factorization for hierarchical multilayer topic modeling." 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019.

Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Training via backpropagation

Neural NMF: Forward and back propagation algorithms for hNMF.



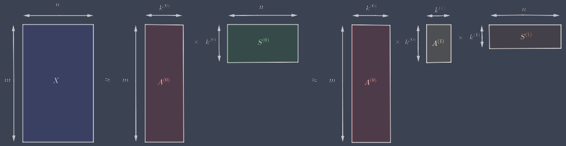
- Regard the A matrices as independent variables, determine the S matrices from the A matrices.
- Define $q(X, A) := \text{argmin}_{S \geq 0} \|X - AS\|_F^2$ (least-squares).
- Pin the values of S to those of A by recursively setting $S^{(\ell)} := q(S^{(\ell-1)}, A^{(\ell)})$.

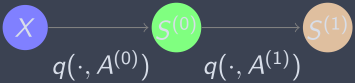
Gao, Mengdi, et al. "Neural nonnegative matrix factorization for hierarchical multilayer topic modeling." 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019.

Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Training via backpropagation

Neural NMF: Forward and back propagation algorithms for hNMF.



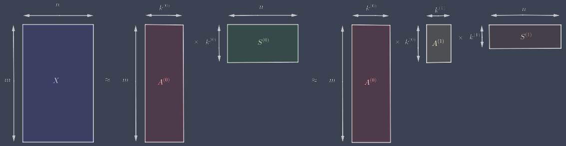


Gao, Mengdi, et al. "Neural nonnegative matrix factorization for hierarchical multilayer topic modeling." 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019.

Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Training via backpropagation

Neural NMF: Forward and back propagation algorithms for hNMF.



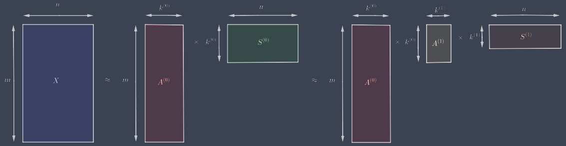
Training:

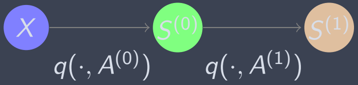
Gao, Mengdi, et al. "Neural nonnegative matrix factorization for hierarchical multilayer topic modeling." 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019.

Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Training via backpropagation

Neural NMF: Forward and back propagation algorithms for hNMF.





Training:

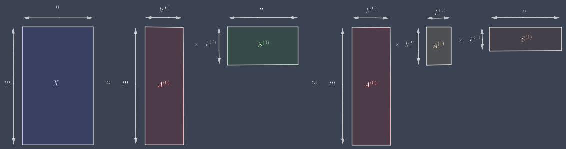
- forward propagation:
  $S^{(0)} = q(X, A^{(0)})$,
  $S^{(1)} = q(S^{(0)}, A^{(1)})$, ...,
  $S^{(L)} = q(S^{(L-1)}, A^{(L)})$
- back propagation: update $\{A^{(i)}\}$ with $\nabla E(\{A^{(i)}\})$

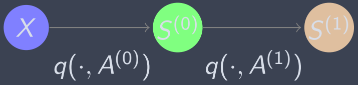
Gao, Mengdi, et al. "Neural nonnegative matrix factorization for hierarchical multilayer topic modeling." 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019.

Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Training via backpropagation

Neural NMF: 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 $\nabla E(\{A^{(i)}\})$

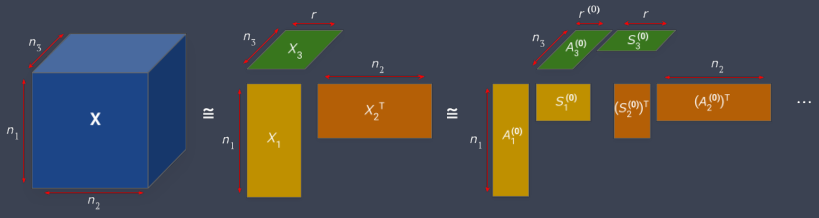
Apply this approach to each mode of HNCPD!

Gao, Mengdi, et al. "Neural nonnegative matrix factorization for hierarchical multilayer topic modeling." 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019.

Related work: [Flenner, Hunter 2018], [Trigeorgis, Bousmalis, Zafeiriou, Schuller 2016], [Le Roux, Hershey, Weninger 2015], [Sun, Nasrabadi, Tran 2017]

» Neural NCPD

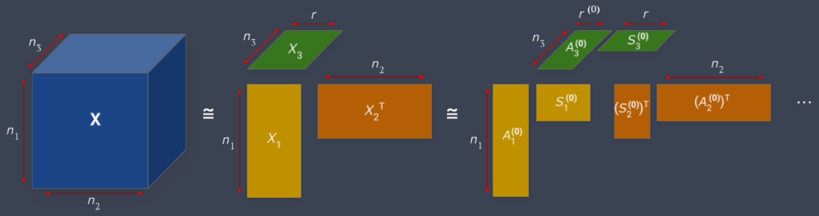
Train independent neural NMF models for each mode of tensor from fixed NCPD factor matrices.



Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

» Neural NCPD

Train independent neural NMF models for each mode of tensor from fixed NCPD factor matrices.



Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

» Gradient Calculation

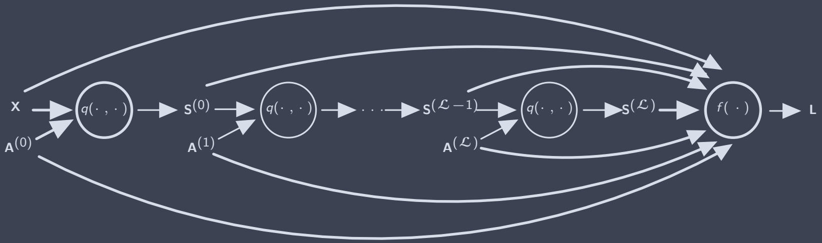
Theorem-ish [Will, Zhang, Sadovnik, Gao, Vendrow, H., Molitor, Needell, 22+]

Given knowledge of the support of $q(\boldsymbol{A}, \boldsymbol{X})$, the gradient $\nabla_{\boldsymbol{A}} q(\boldsymbol{A}, \boldsymbol{X})$ has a closed-form expression almost everywhere in the space of real-valued matrix pairs. This gradient expression is inherited from unconstrained least-squares.

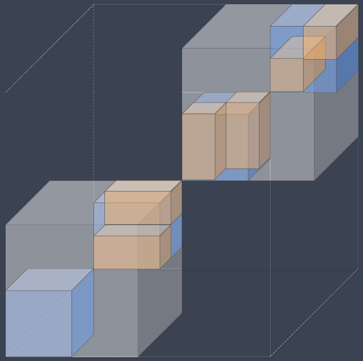
» Gradient Calculation

Theorem-ish [Will, Zhang, Sadovnik, Gao, Vendrow, H., Molitor, Needell, 22+]

Given knowledge of the support of $q(\boldsymbol{A}, \boldsymbol{X})$, the gradient $\nabla_{\boldsymbol{A}} q(\boldsymbol{A}, \boldsymbol{X})$ has a closed-form expression almost everywhere in the space of real-valued matrix pairs. This gradient expression is inherited from unconstrained least-squares.

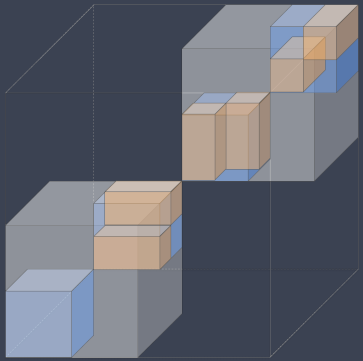


» Synthetic Tensor



The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.

» Synthetic Tensor

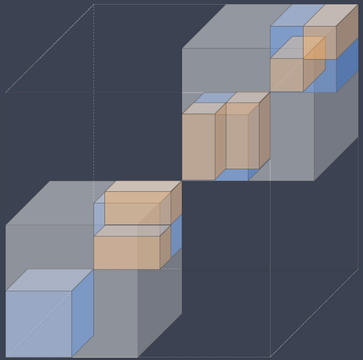


The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.

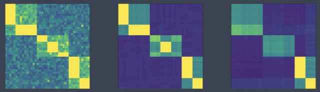
Relative reconstruction error.

Method	$r_0 = 7$	$r_1 = 4$	$r_2 = 2$
Multi-HNTF	0.454	0.548	0.721
Neural HNCPD [Vendrow, et. al.]	0.454	**0.508**	**0.714**
Standard HNCPD [Vendrow, et. al.]	0.454	0.612	0.892
HNTF-1 [Cichocki, et. al.]	0.454	0.576	0.781
HNTF-2 [Cichocki, et. al.]	0.454	0.587	0.765
HNTF-3 [Cichocki, et. al.]	0.454	0.560	0.747

» Synthetic Tensor



The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.



Projections of tensor approximation at each layer of Multi-HNTF.

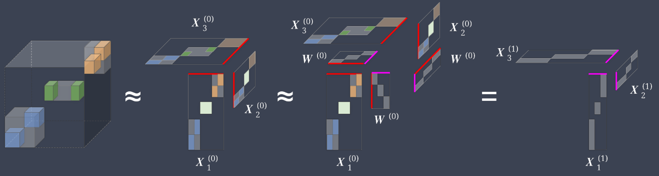
Relative reconstruction error.

Method	$r_0 = 7$	$r_1 = 4$	$r_2 = 2$
Multi-HNTF	0.454	0.548	0.721
Neural HNCPD [Vendrow, et. al.]	0.454	**0.508**	**0.714**
Standard HNCPD [Vendrow, et. al.]	0.454	0.612	0.892
HNTF-1 [Cichocki, et. al.]	0.454	0.576	0.781
HNTF-2 [Cichocki, et. al.]	0.454	0.587	0.765
HNTF-3 [Cichocki, et. al.]	0.454	0.560	0.747

Conclusions

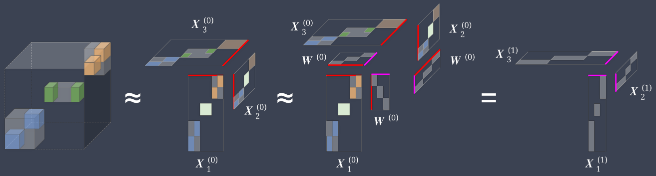
» Conclusions

- ▷ Multi-HNTF is a hierarchical tensor decomposition model that generalizes hierarchical NMF.



» Conclusions

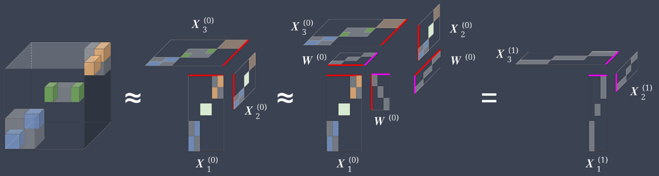
- Multi-HNTF is a hierarchical tensor decomposition model that `generalizes` hierarchical NMF.



- Model can be trained by your `favorite NMF method` with an additional projection step.

» Conclusions

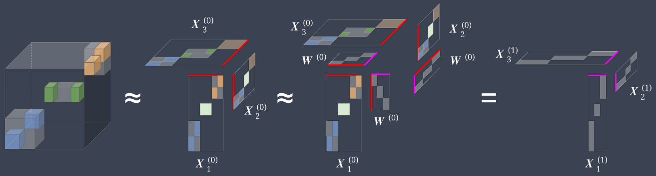
- Multi-HNTF is a hierarchical tensor decomposition model that `generalizes` hierarchical NMF.



- Model can be trained by your `favorite NMF method` with an additional projection step.
- Neural NMF and Neural NCPD can help `mitigate` devastating error propagation through multi-layer decomposition models.

» Conclusions

- Multi-HNTF is a hierarchical tensor decomposition model that `generalizes` hierarchical NMF.



- Model can be trained by your `favorite NMF method` with an additional projection step.
- Neural NMF and Neural NCPD can help `mitigate` devastating error propagation through multi-layer decomposition models.

- Develop backpropagation framework for Multi-HNTF and first layer NCPD.

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

[3] J. Vendrow, J. Haddock, and D. Needell. A generalized hierarchical nonnegative tensor decomposition. In Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2022.

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

[5] J Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika, 35(3):283–319, 1970.

[6] Richard A Harshman et al. Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis. 1970.

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