An Interpretable Joint Nonnegative Matrix Factorization-Based Point Cloud Distance Measure

by Jamie Haddock
(Harvey Mudd College, Department of Mathematics)
on March 23, 2023,
Conference on Information Sciences and Systems (CISS)

joint with Hannah Friedman, Amani R. Maina-Kilaas, Julianna Schalkwyk, and Hina Ahmed (graduating Harvey Mudd College and Pitzer College seniors)

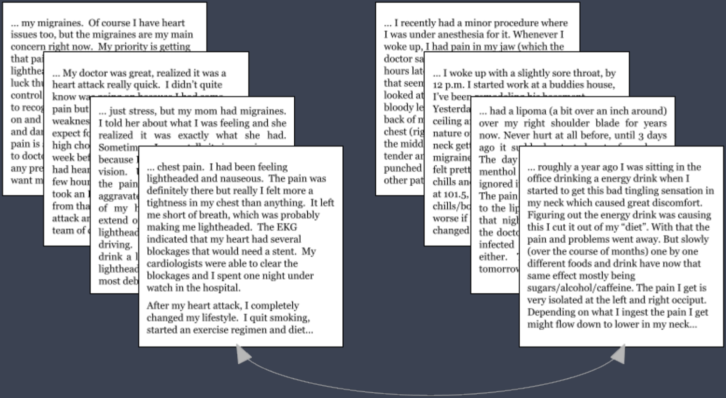
supported by NSF DMS #2211318

matH of Algorithms, Data & Decisions

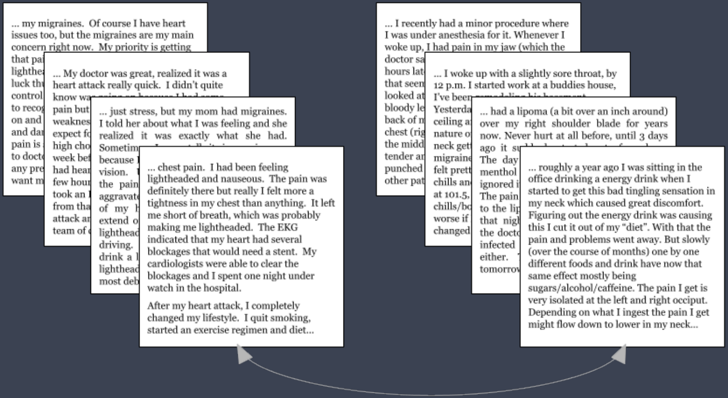
research group @ HARVEY MUDD COLLEGE

Motivation

» Dataset similarity

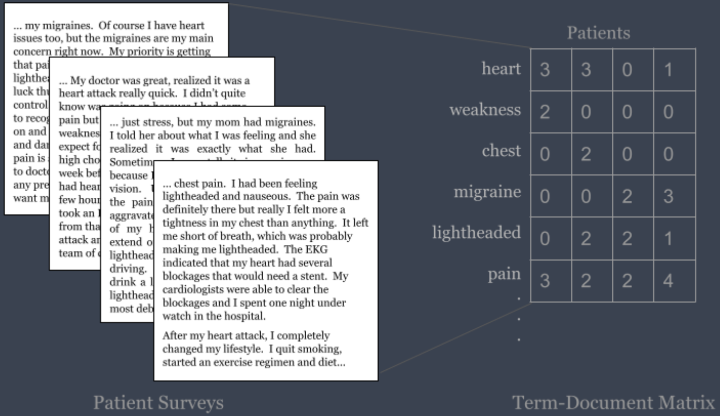


» Dataset similarity



Understanding the similarities and differences between datasets arises in many contexts: e.g., transfer learning, plagiarism/manipulation detection, and data denoising.

» Dataset similarity



» Point Cloud Distances

Chamfer's distance:

$$d_{\mathsf{cham}}(X_1, X_2) = \frac{1}{|X_1|}\sum_{\mathbf{x}\in X_1} \min_{\mathbf{y}\in X_2} \|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{1}{|X_2|}\sum_{\mathbf{y}\in X_2} \min_{\mathbf{x}\in X_1} \|\mathbf{x}-\mathbf{y}\|_2^2$$

» Point Cloud Distances

Chamfer's distance:

$$d_{\mathsf{cham}}(X_1, X_2) = \frac{1}{|X_1|}\sum_{\mathbf{x}\in X_1}\min_{\mathbf{y}\in X_2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \frac{1}{|X_2|}\sum_{\mathbf{y}\in X_2}\min_{\mathbf{x}\in X_1}\|\mathbf{x}-\mathbf{y}\|_2^2$$

We seek a distance that is:

» Point Cloud Distances

Chamfer's distance:

$$d_{\mathrm{cham}}(X_1, X_2) = \frac{1}{|X_1|}\sum_{\mathbf{x}\in X_1}\min_{\mathbf{y}\in X_2}\|\mathbf{y}-\mathbf{x}\|_2^2+\frac{1}{|X_2|}\sum_{\mathbf{y}\in X_2}\min_{\mathbf{x}\in X_1}\|\mathbf{x}-\mathbf{y}\|_2^2$$

We seek a distance that is:

* More robust to outliers.

» Point Cloud Distances

Chamfer's distance:

$$d_{\mathrm{cham}}(X_1, X_2) = \frac{1}{|X_1|} \sum_{\mathbf{x} \in X_1} \min_{\mathbf{y} \in X_2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{1}{|X_2|} \sum_{\mathbf{y} \in X_2} \min_{\mathbf{x} \in X_1} \|\mathbf{x} - \mathbf{y}\|_2^2$$

We seek a distance that is:

* More robust to outliers.
* Utilizes the structure of data.

» Point Cloud Distances

Chamfer's distance:

$$d_{\mathrm{cham}}(X_1, X_2) = \frac{1}{|X_1|} \sum_{\mathbf{x} \in X_1} \min_{\mathbf{y} \in X_2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{1}{|X_2|} \sum_{\mathbf{y} \in X_2} \min_{\mathbf{x} \in X_1} \|\mathbf{x} - \mathbf{y}\|_2^2$$

We seek a distance that is:

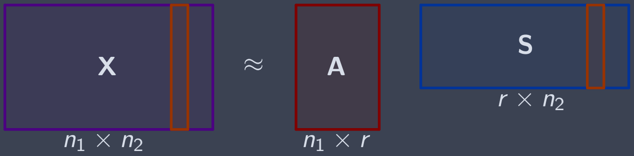
* More robust to outliers.
* Utilizes the structure of data.
* Helps illustrate how the data is similar or dissimilar.

Introduction

» Nonnegative Matrix Factorization (NMF)

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

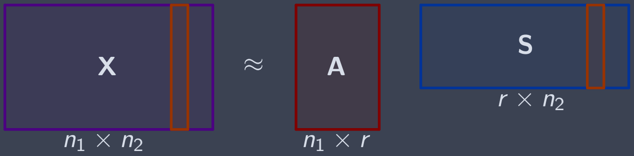
$$\mathbf{X} \approx \mathbf{AS}.$$



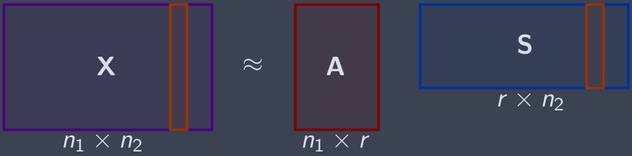
» Nonnegative Matrix Factorization (NMF)

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

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

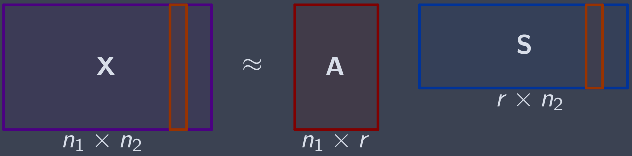


» Nonnegative Matrix Factorization (NMF)



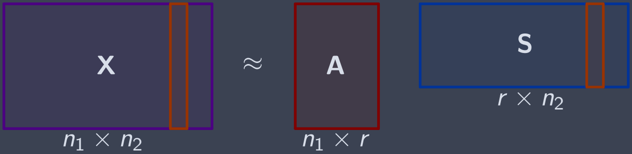
- ▷ Popularized by [Lee & Seung 1999]

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

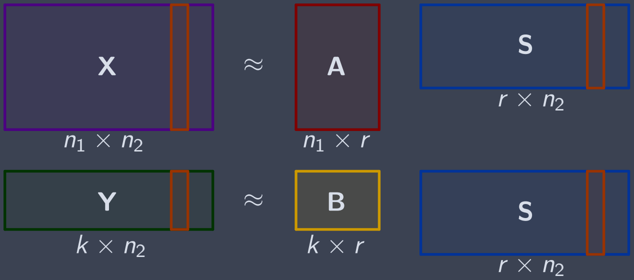
» Joint NMF

Model: Jointly factorize two nonnegative matrices $\mathbf{X}_1$ and $\mathbf{X}_2$, sharing one factor matrix between the factorizations.

» Joint NMF

Model: Jointly factorize two nonnegative matrices $\mathbf{X}_1$ and $\mathbf{X}_2$, sharing one factor matrix between the factorizations.

Example: Semi-supervised NMF



`Often applied in classification!`

[1]Lee, H., Yoo, J., and Choi, S. "Semi-supervised nonnegative matrix factorization." IEEE Signal Processing Letters 17.1 (2009): 4-7.
H., et al. "Semi-supervised Nonnegative Matrix Factorization for Document Classification." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

» Joint NMF

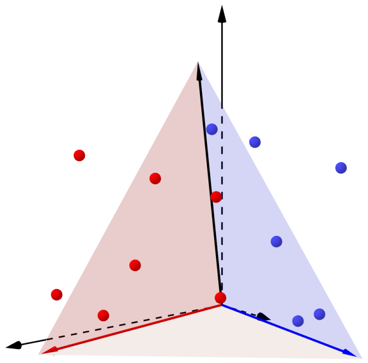
Model: Jointly factorize two nonnegative matrices $\mathbf{X}_1$ and $\mathbf{X}_2$, sharing one factor matrix between the factorizations.

Example: Joint NMF/Guided NMF

`Intuition: many columns of` A `used in representing` X_1 `and` X_2 `indicates dataset similarity.`

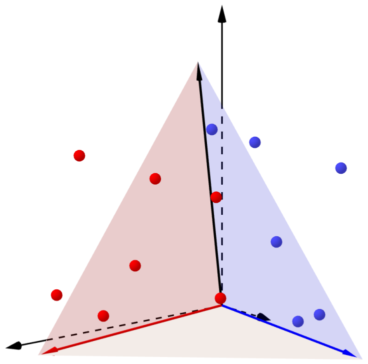
[1] Kim, H., et al. "Simultaneous discovery of common and discriminative topics via joint nonnegative matrix factorization." Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining. 2015.
Vendrow, J., **H.**, et al. "On a guided nonnegative matrix factorization." IEEE Int. Conf. Acoust. Speech Sig. Process. (ICASSP), 2021.

» Joint NMF (jNMF) for Similarity

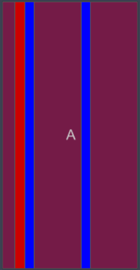


NMF learns a conic representation of data

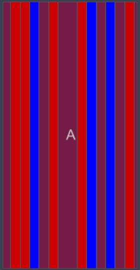
» Joint NMF (jNMF) for Similarity



NMF learns a conic representation of data



Similar



Different

Our Method and Distance

» Our jNMF Similarity Method

Intuition: use the entries of S_1 and S_2 to measure how much topics are shared between datasets.

Method:

» Our jNMF Similarity Method

Intuition: use the entries of S_1 and S_2 to measure how much topics are shared between datasets.

Method:

* Scale each column in X_1, X_2 to be mean one.
* Learn rank-k jNMF approximation, $[X_1 \ X_2] \approx A[S_1 \ S_2]$.

» Our jNMF Similarity Method

`Intuition: use the entries of` S_1 `and` S_2 `to measure how much topics are shared between datasets.`

Method:

* Scale each column in X_1, X_2 to be mean one.
* Learn rank-k jNMF approximation, $[X_1\ X_2] \approx A[S_1\ S_2]$.
* For $i = 1, \cdots, k$, define

$$s_i = \max\left(\{s_{ij}^{(1)}\}_{j=1}^{n_1} \cup \{s_{ij}^{(2)}\}_{j=1}^{n_2}\right)$$

where $s_{i1}^{(1)}, s_{i2}^{(1)}, \cdots, s_{in_1}^{(1)}$ and $s_{i1}^{(2)}, s_{i2}^{(2)}, \cdots, s_{in_2}^{(2)}$ are the entries of the ith rows of S_1 and S_2, respectively.

» Our jNMF Similarity Method

Intuition: use the entries of S_1 and S_2 to measure how much topics are shared between datasets.

Method:

* Scale each column in X_1, X_2 to be mean one.
* Learn rank-k jNMF approximation, $[X_1\ X_2] \approx A[S_1\ S_2]$.
* For $i = 1, \cdots, k$, define

$$s_i = \max\left(\{s_{ij}^{(1)}\}_{j=1}^{n_1} \cup \{s_{ij}^{(2)}\}_{j=1}^{n_2}\right)$$

  where $s_{i1}^{(1)}, s_{i2}^{(1)}, \cdots, s_{in_1}^{(1)}$ and $s_{i1}^{(2)}, s_{i2}^{(2)}, \cdots, s_{in_2}^{(2)}$ are the entries of the ith rows of S_1 and S_2, respectively.
* For $j = 1, ..., K$

» Our jNMF Similarity Method

`Intuition: use the entries of` S_1 `and` S_2 `to measure how much topics are shared between datasets.`

Method:

* Scale each column in X_1, X_2 to be mean one.
* Learn rank-k jNMF approximation, $[X_1\ X_2] \approx A[S_1\ S_2]$.
* For $i = 1, \cdots, k$, define

$$s_i = \max\left(\{s_{ij}^{(1)}\}_{j=1}^{n_1} \cup \{s_{ij}^{(2)}\}_{j=1}^{n_2}\right)$$

  where $s_{i1}^{(1)}, s_{i2}^{(1)}, \cdots, s_{in_1}^{(1)}$ and $s_{i1}^{(2)}, s_{i2}^{(2)}, \cdots, s_{in_2}^{(2)}$ are the entries of the ith rows of S_1 and S_2, respectively.
* For $j = 1, ..., K$
  * Choose $T_i \sim \text{unif}([0, s_i])$ for $i = 1, 2, \cdots, k$.

» Our jNMF Similarity Method

`Intuition: use the entries of` S_1 `and` S_2 `to measure how much topics are shared between datasets.`

Method:

* Scale each column in X_1, X_2 to be mean one.
* Learn rank-k jNMF approximation, $[X_1\ X_2] \approx A[S_1\ S_2]$.
* For $i = 1, \cdots, k$, define

$$s_i = \max\left(\{s_{ij}^{(1)}\}_{j=1}^{n_1} \cup \{s_{ij}^{(2)}\}_{j=1}^{n_2}\right)$$

  where $s_{i1}^{(1)}, s_{i2}^{(1)}, \cdots, s_{in_1}^{(1)}$ and $s_{i1}^{(2)}, s_{i2}^{(2)}, \cdots, s_{in_2}^{(2)}$ are the entries of the ith rows of S_1 and S_2, respectively.
* For $j = 1, ..., K$
  * Choose $T_i \sim \text{unif}([0, s_i])$ for $i = 1, 2, \cdots, k$.
  * Compute $\boldsymbol{p}_i^{(j)} := F_i^{(2)}(T_i) - F_i^{(1)}(T_i)$, where

$$F_i^{(1)}(T_i) := \frac{1}{n_1}\sum_{j=1}^{n_1} \mathbf{1}[s_{ij}^{(1)} < T_i] \text{ and } F_i^{(2)}(T_i) := \frac{1}{n_2}\sum_{j=1}^{n_2} \mathbf{1}[s_{ij}^{(2)} < T_i].$$

» Our jNMF Similarity Method

Intuition: use the entries of S_1 and S_2 to measure how much topics are shared between datasets.

Method:

* Scale each column in X_1, X_2 to be mean one.
* Learn rank-k jNMF approximation, $[X_1\ X_2] \approx A[S_1\ S_2]$.
* For $i = 1, \cdots, k$, define

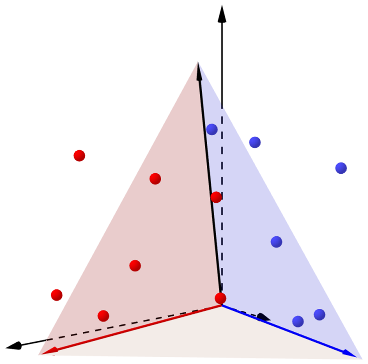
$$s_i = \max\left(\{s_{ij}^{(1)}\}_{j=1}^{n_1} \cup \{s_{ij}^{(2)}\}_{j=1}^{n_2}\right)$$

  where $s_{i1}^{(1)}, s_{i2}^{(1)}, \cdots, s_{in_1}^{(1)}$ and $s_{i1}^{(2)}, s_{i2}^{(2)}, \cdots, s_{in_2}^{(2)}$ are the entries of the ith rows of S_1 and S_2, respectively.
* For $j = 1, ..., K$
  * Choose $T_i \sim \text{unif}([0, s_i])$ for $i = 1, 2, \cdots, k$.
  * Compute $\boldsymbol{p}_i^{(j)} := F_i^{(2)}(T_i) - F_i^{(1)}(T_i)$, where

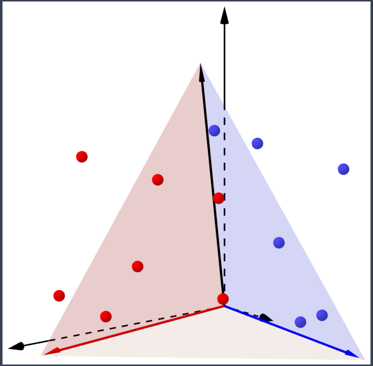
$$F_i^{(1)}(T_i) := \frac{1}{n_1}\sum_{j=1}^{n_1} \mathbf{1}[s_{ij}^{(1)} < T_i] \text{ and } F_i^{(2)}(T_i) := \frac{1}{n_2}\sum_{j=1}^{n_2} \mathbf{1}[s_{ij}^{(2)} < T_i].$$

* Return $\bar{\boldsymbol{p}} = \frac{1}{K}\sum_{j=1}^{K} \mathbf{p}^{(j)}$

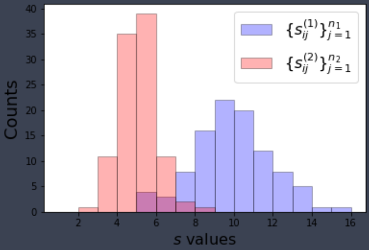
» jNMF Similarity Method

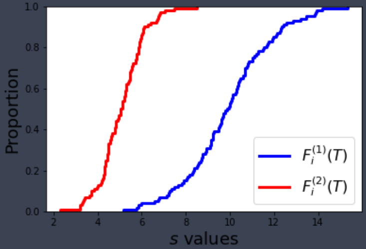


» jNMF Similarity Method

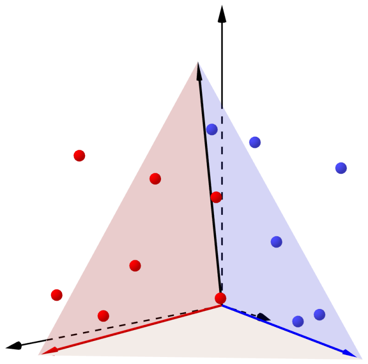


$$p_i^{(j)} := F_i^{(2)}(T_i) - F_i^{(1)}(T_i)$$

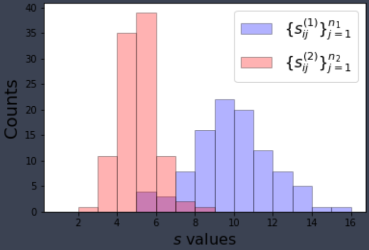


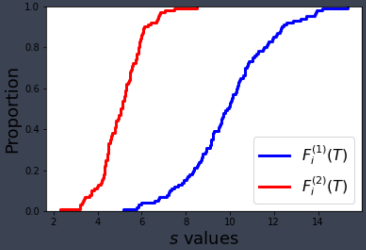


» jNMF Similarity Method



$$p_i^{(j)} := F_i^{(2)}(T_i) - F_i^{(1)}(T_i)$$



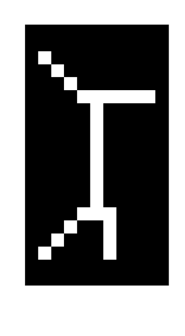


$$d(X_1, X_2) := \|\bar{\mathbf{p}}\|_1$$

Experiments

» Swimmer Image Dataset

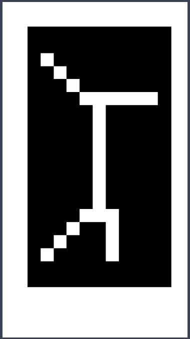
Swimmer Images

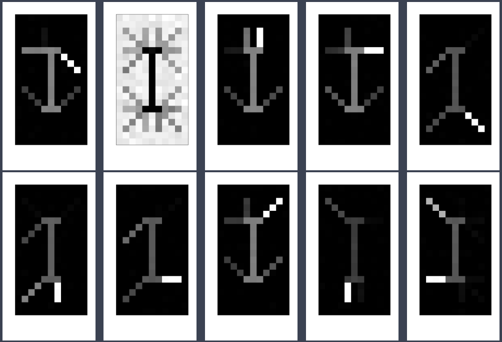


[1]Donoho, D., and Stodden, V. "When does non-negative matrix factorization give a correct decomposition into parts?." NeurIPS (2003).

» Swimmer Image Dataset

Swimmer Images



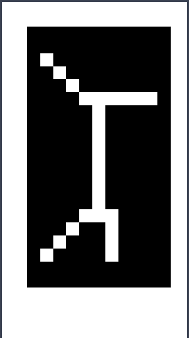


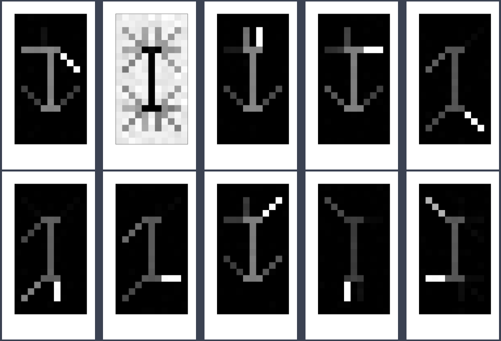
Basis vectors learned by jNMF on Swimmer dataset $X_1, X_1 + N$ where N is uniform noise.

[1]Donoho, D., and Stodden, V. "When does non-negative matrix factorization give a correct decomposition into parts?." NeurIPS (2003).

» Swimmer Image Dataset

Swimmer Images





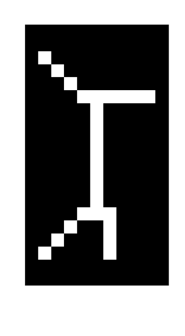
Basis vectors learned by jNMF on Swimmer dataset $X_1, X_1 + N$ where N is uniform noise.

$$\bar{\boldsymbol{p}} = [0.063, -0.901, 0.076, 0.065, 0.069, 0.058, 0.058, 0.069, 0.079, 0.079]$$

[1]Donoho, D., and Stodden, V. "When does non-negative matrix factorization give a correct decomposition into parts?." NeurIPS (2003).

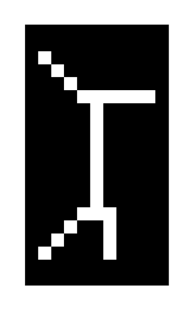
» Swimmer Image Dataset

Swimmer Images



» Swimmer Image Dataset

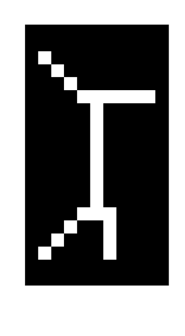
Swimmer Images



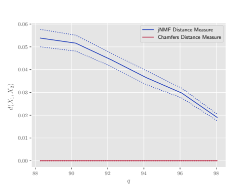
X_2	X_1	$X_1 P_\pi$	λX_1	$\tilde{X_1}$	$X_1 + N$	N
$d(X_1, X_2)$	0.000	0.000	0.000	0.052	1.509	2.297
$d_{\text{cham}}(X_1, X_2)$	0.000	0.000	0.000	0.000	0.741	1.560

» Swimmer Image Dataset

Swimmer Images

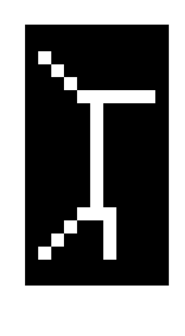


X_2	X_1	$X_1 P_\pi$	λX_1	$\tilde{X_1}$	$X_1 + N$	N
$d(X_1, X_2)$	0.000	0.000	0.000	0.052	1.509	2.297
$d_{\text{cham}}(X_1, X_2)$	0.000	0.000	0.000	0.000	0.741	1.560

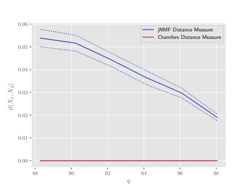


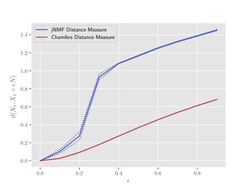
» Swimmer Image Dataset

Swimmer Images



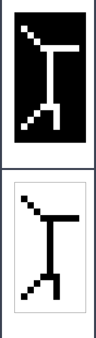
X_2	X_1	$X_1 P_\pi$	λX_1	$\tilde{X_1}$	$X_1 + N$	N
$d(X_1, X_2)$	0.000	0.000	0.000	0.052	1.509	2.297
$d_{\text{cham}}(X_1, X_2)$	0.000	0.000	0.000	0.000	0.741	1.560





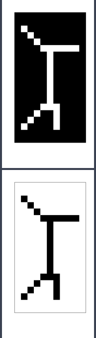
» Swimmer Image Dataset

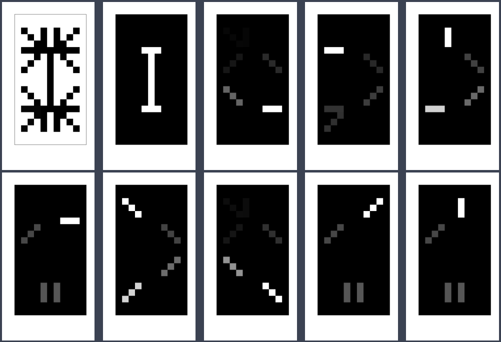
Swimmer and
Inverse Swimmer:



» Swimmer Image Dataset

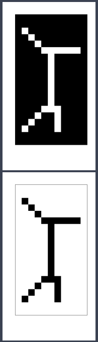
Swimmer and
Inverse Swimmer:

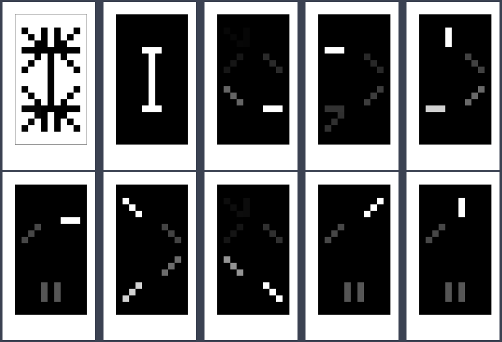




» Swimmer Image Dataset

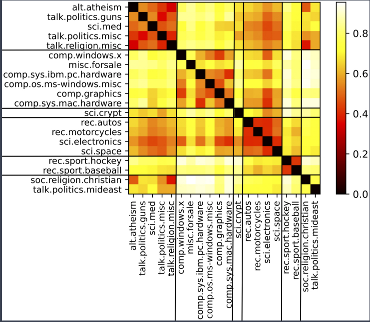
Swimmer and Inverse Swimmer:





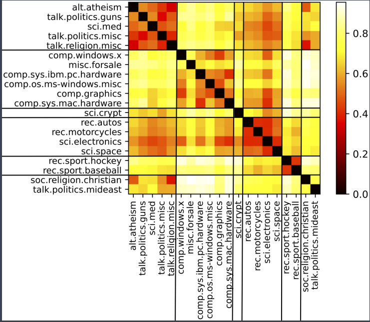
$$\bar{\mathbf{p}} = [-0.999, 1.000, 0.010, -0.017, 0.003, -0.004, 0.015, 0.004, -0.001, -0.000]$$

» 20 Newsgroups Dataset

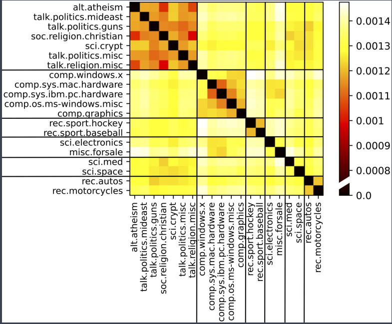


jNMF Distance

» 20 Newsgroups Dataset



jNMF Distance



Chamfer Distance

Conclusions

» Conclusions

- ▷ jNMF provides information about dataset similarity and dissimilarity

» Conclusions

- jNMF provides information about dataset similarity and dissimilarity

- we can aggregate this information using samples from the empirical distribution function to form a vector indicating which dataset learned basis vectors represent, $\bar{\boldsymbol{p}}$

Conclusions

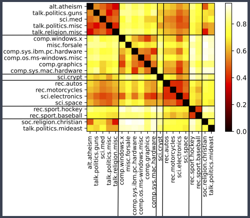
- jNMF provides information about dataset similarity and dissimilarity

- we can aggregate this information using samples from the empirical distribution function to form a vector indicating which dataset learned basis vectors represent, $\bar{\boldsymbol{p}}$
- the information can be further aggregated to yield a distance, $d(X_1, X_2) = \|\bar{\boldsymbol{p}}\|_1$

Conclusions

- jNMF provides information about dataset similarity and dissimilarity

- we can aggregate this information using samples from the empirical distribution function to form a vector indicating which dataset learned basis vectors represent, $\bar{\boldsymbol{p}}$
- the information can be further aggregated to yield a distance, $d(X_1, X_2) = \|\bar{\boldsymbol{p}}\|_1$
- initial experiments are promising



» Thanks for listening! Questions?

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

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

[3] 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.

[4] Hannah Kim, Jaegul Choo, Jingu Kim, Chandan K Reddy, and Haesun Park. Simultaneous discovery of common and discriminative topics via joint nonnegative matrix factorization. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 567–576, 2015.

[5] 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.

[6] David Donoho and Victoria Stodden. When does non-negative matrix factorization give a correct decomposition into parts? Advances in neural information processing systems, 16, 2003.