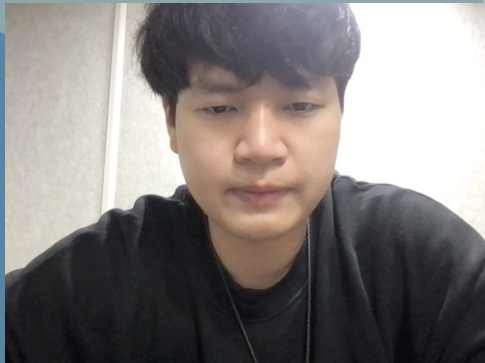
Week 14 – Dimensionality reduction

L14-05. Nonnegative matrix factorization (What & When & How)

Donghee and Jungwoo



All linear dimension reduction can be thought of as…

Matrix decompostition (Factorization)

in a way that retains some variance of the original data...

$$\underset{n \times p}{X} \approx \underset{n \times k}{A}\,\underset{k \times p}{B} \qquad k < \min(n, p)$$

it is also called, low-rank approximation

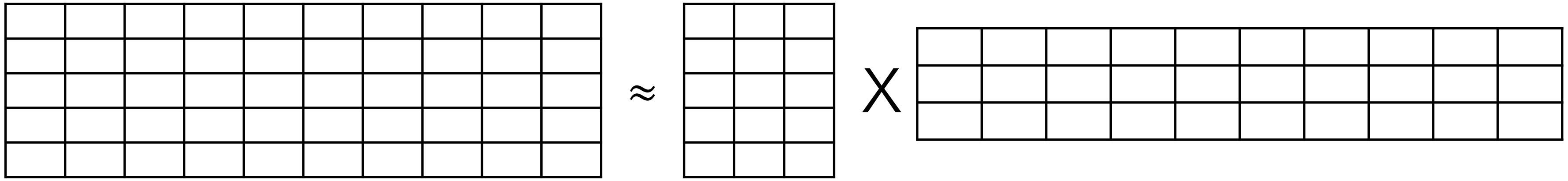
Nonnegative matrix factorization

PCA: Decomposition of data covariance matrix!

ICA: Decomposing data to spatially or temporally independent components!

NNMF: Decomposing matrix with **"non-negative"** elements

Mathematically speaking…



A ≈ W X H

A: 5 x 10

W: 5 x 3

H: 3 x 10

$\min(|A - WH|^2_F)$

All elements in W and H are … non-negative (larger or equal to zero)

5: number of data
10: number of features (or dimensions)
3: reduced dimension

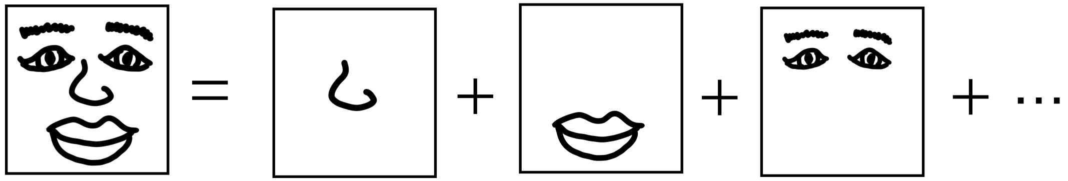
Meaning depends on how you define your A.
(Rows and columns could be transposed.)

W can be thoughts of as basis and *H* can be thought of as weight matrix…

What's the point of being non-negative?

Interpretability!

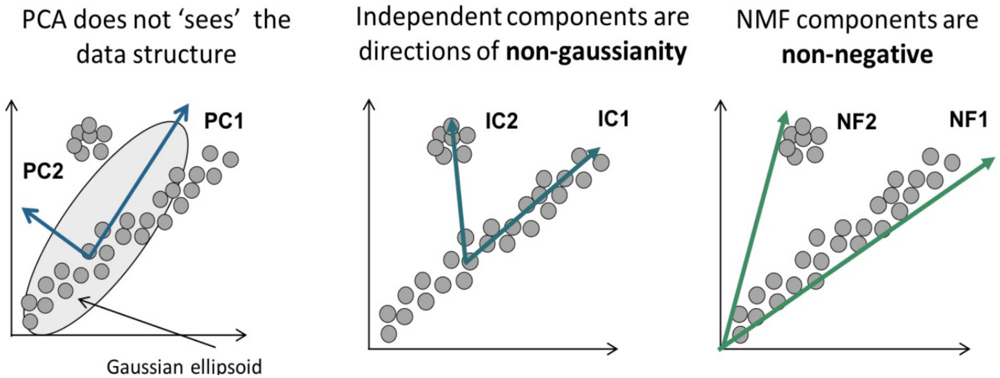
Additive…



Face as sum of
nose + mouse + eyes + …

Doesn't have to be orthogonal!

More capable of
retaining data structure!



https://urszulaczerwinska.github.io/DeconICA/DeconICA_introduction.html

Nonnegative matrix factorization

How to find W and H?

starting from random and optimize through gradient descent.

might fall in local minimum which means no absolute solution…(yield different output every time it runs)
so there are new ways to avoid this.
(Also to lower computational costs)

Default of 'nnmf' function in MATLAB →

- Multiplicative update rules: Lee and Seung, 99, Modified multiplicative update: Lin 07
- Alternating least squares (ALS): Berry et al 06
- Alternating nonnegative least squares (ANLS)
  - Lin, 07, Projected gradient descent
  - D. Kim et al., 07, Quasi-Newton
  - H. Kim and Park, 08, Active-set
  - J. Kim and Park, 08, Block principal pivoting
  - Han et al., 09, Projected Barzilai-Borwein
- Other algorithms and variants
  - Cichocki et al., 07, Hierarchical ALS (HALS)
    Ho, 08, Rank-one Residue Iteration (RRI)
  - Gillis and Glineur, 12, Accelerated multiplicative updates and HALS/multilevel approach
  - Hsieh and Dhillon, 11, Coordinate descent with variable selection
  - Zdunek, Cichocki, Amari 06, Quasi-Newton
  - Chu and Lin, 07, Low dim polytope approx.
  - Other rank-1 deflation based algorithms (Vavasis,..)
  - C. Ding, T. Li, tri-factor NMF, orthogonal NMF, ...
  - Cichocki, Zdunek, Phan, Amari: NMF and NTF: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation, Wiley, 09
  - Andersson and Bro, Nonnegative Tensor Factorization, 00
  - And MANY MORE...

So many…

How to find W and H? (deeper…)

$$\|X - WH\|_F^2 = tr((X - WH)^T(X - WH))$$

$$H := H - \eta_H \circ \nabla_H \|X - WH\|_F^2$$

$$W := W - \eta_W \circ \nabla_W \|X - WH\|_F^2$$

$$H := H + \eta_H \circ (W^T X - W^T W H)$$

$$W := W + \eta_W \circ (X H^T - W H H^T)$$

$$\eta_H = \frac{H}{W^T W H}$$

$$\eta_W = \frac{W}{W H H^T}$$

$$H := H \circ \frac{W^T X}{W^T W H}$$

$$W := W \circ \frac{X H^T}{W H H^T}$$

Gradient descent…!

Excellent explanation in Korean: https://angeloyeo.github.io/2020/10/15/NMF.html

Original paper: https://proceedings.neurips.cc/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf

Fun facts… Movie recommending system

Learning from Incomplete Ratings Using
Non-negative Matrix Factorization

Sheng Zhang, Weihong Wang, James Ford, Fillia Makedon
{clap, whwang, jford, makedon}@cs.dartmouth.edu
Department of Computer Science, Dartmouth College, Hanover, NH 03755

Netflix ratings

	M1	M2	M3	M4	M5
	1	3	2	5	4
	2	1	1	1	5
	3	2	3	1	5
	2	4	1	5	2

Matrix Factorization

	M1	M2	M3	M4	M5
Comedy	3	1	1	3	1
Action	1	2	4	1	3

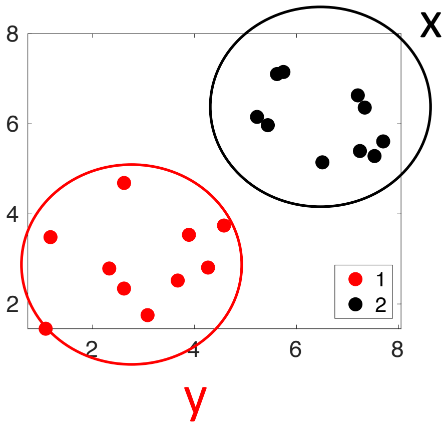
	Comedy	Action
A	✓	✗
B	✗	✓
C	✓	✗
D	✓	✓

	M1	M2	M3	M4	M5
	3	1	1	3	1
	1	2	4	1	3
	3	1	1	3	1
	4	3	5	4	4

https://www.youtube.com/watch?v=ZspR5PZemcs&t=1333s

Fun facts… relationship with K-means clustering

K-means clustering can be thought of as…
more contrained version of NMF.



$$\underset{A}{\begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}} \approx \underset{W}{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} \underset{H}{\begin{pmatrix} \bar{x}_1 & \bar{x}_2 \\ \bar{y}_1 & \bar{y}_2 \end{pmatrix}}$$

Sparse Nonnegative Matrix Factorization for Clustering

Jingu Kim and Haesun Park *

College of Computing
Georgia Institute of Technology
266 Ferst Drive, Atlanta, GA 30332, USA
{jingu, hpark}@cc.gatech.edu

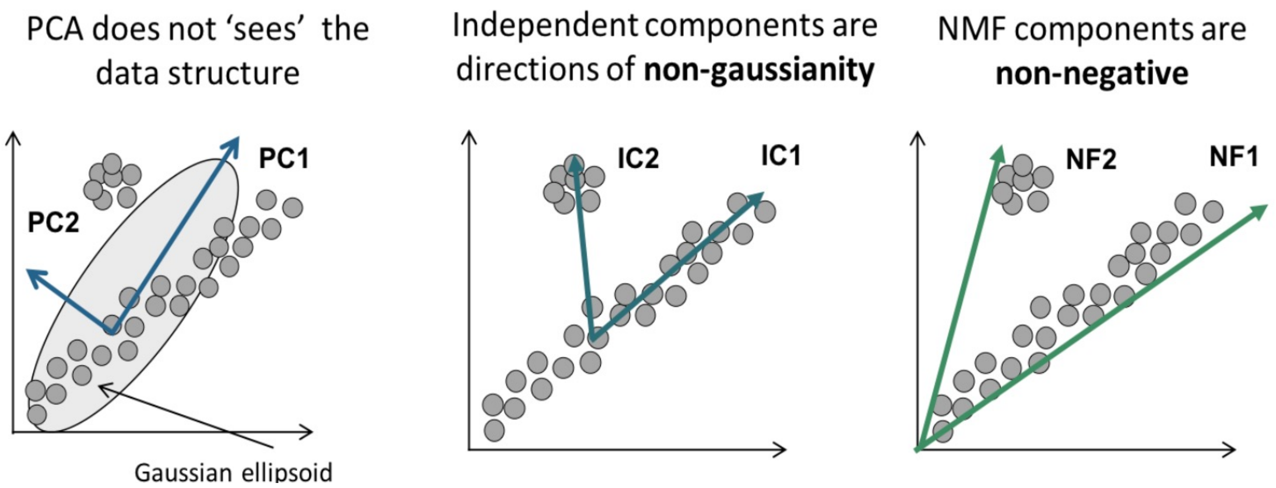
How about in fMRI data⋯?

Since fMRI data includes negative values⋯ it is not widely used.

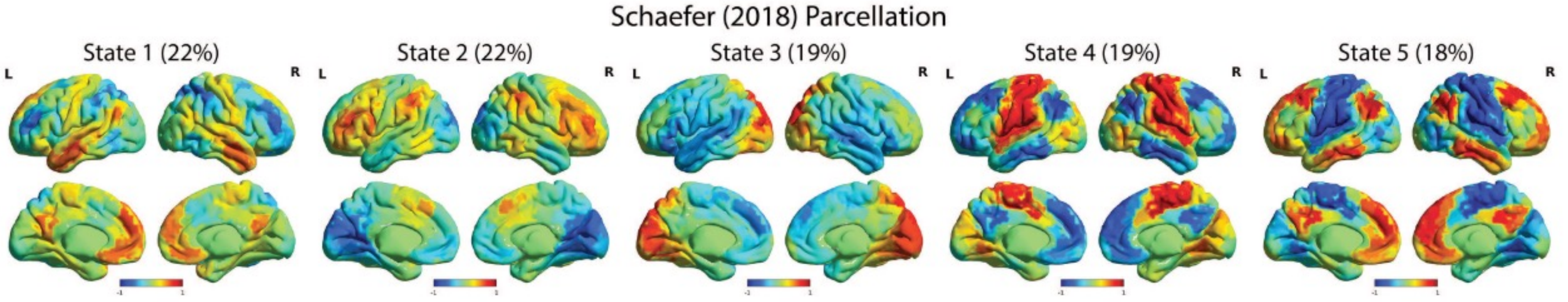
Still methodological issues of how to apply this technique!

Let’s apply it ourselves in the code.

Clustering with NMF



Nonnegative matrix factorization



Zachary et al. 2021

5 optimal clusters with k-means clustering!
How will it results using NMF?

Note: There will be no thorough inspection of using NMF, just going to focus on how it can be applied through code!

Nonnegative matrix factorization

No parcellation, whole brain, single subject…

```
>> size(obj.dat)

ans =

      211119         914
```

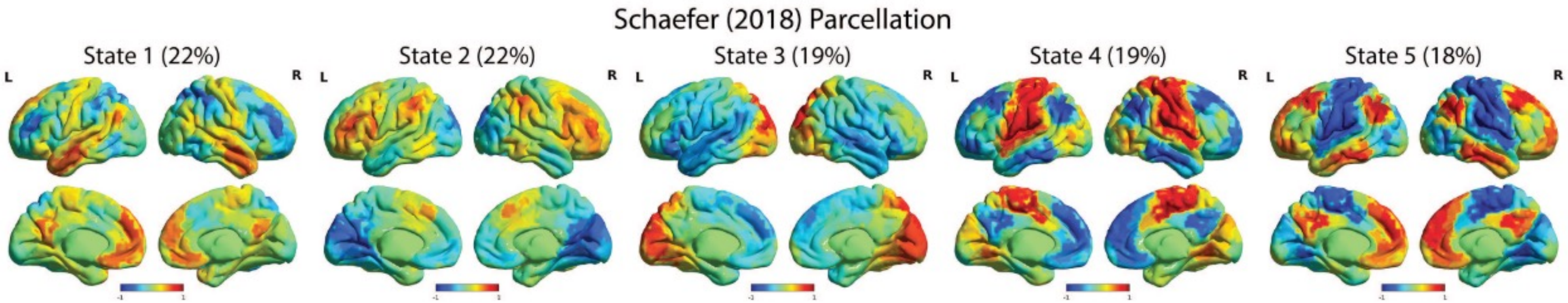
One simple block will do!

```
[W1, H1] = nnmf(obj.dat, 5);
```

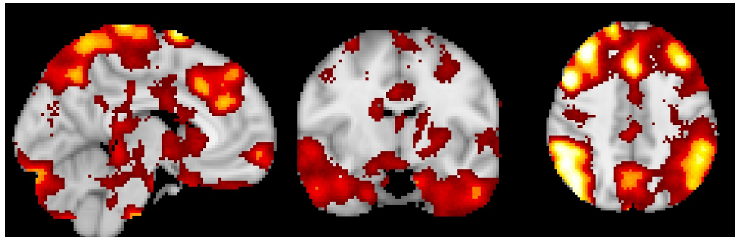
Note: default 'nnmf' function uses 'ALS' algorithms

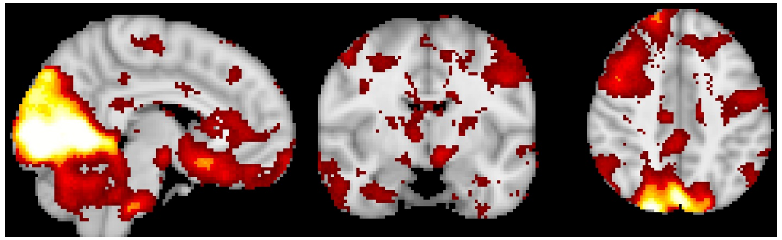
Alternative NMFs! : https://github.com/kimjingu/nonnegfac-matlab/blob/master/nmf.m

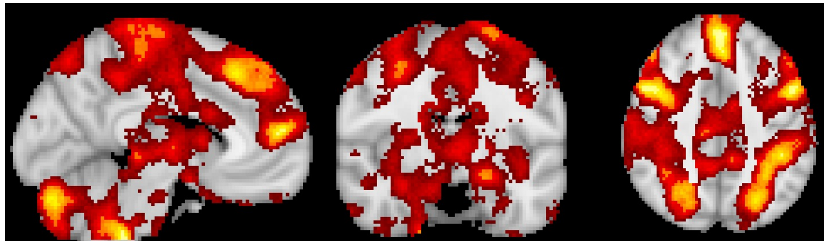
Nonnegative matrix factorization

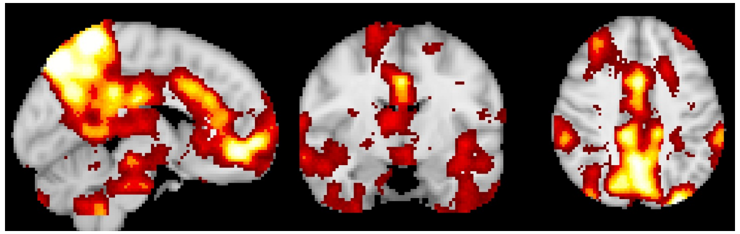


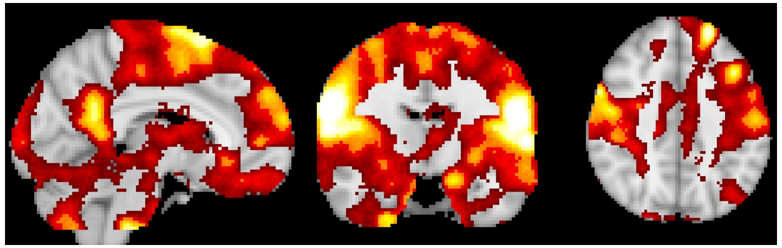
Results in 'W1' matrix … See any correspondence?







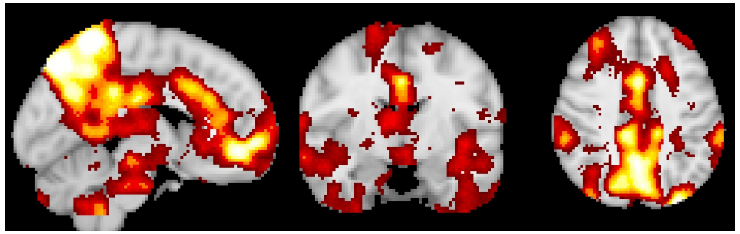


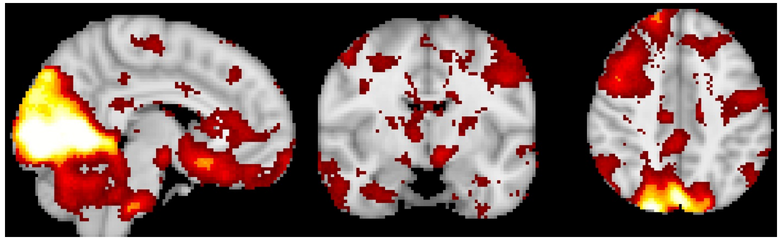


Nonnegative matrix factorization

Of course, direct comparison between two results is not plausible!

1. What are the empty spots?





2. How should we interpret the result?

Nonnegative matrix factorization

1. What are the empty spots?

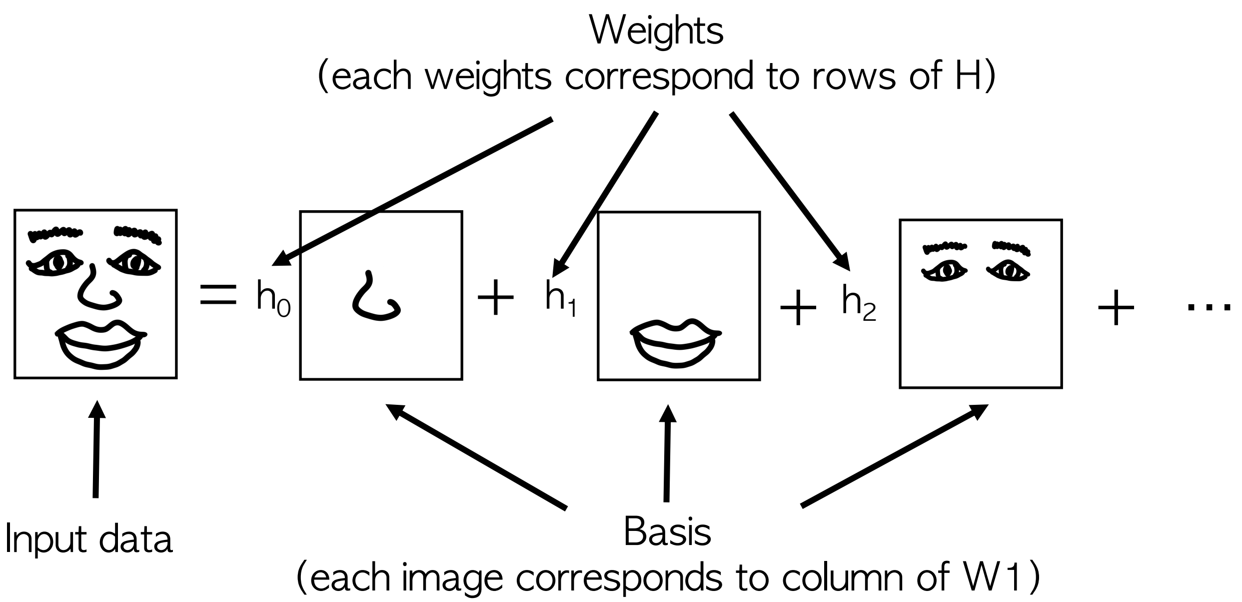
I added matrix with negative values in 'obj.dat' variable…

```
>> size(obj.dat)

ans =

      211119         914
```

Since all elements in W and H are positive…
There are cases when no solution to give negative values…which would results in zero…!

But, it depends on constraints, input data and etc…

2. How to interpret the result?



NMF itself is not difficult, but further analysis based on W and H is possible would be an usual way.

Nonnegative matrix factorization

Comments···

Have its potential in that
it has good **interpretability** which is quite crucial for interpreting fMRI data!

However, NMF is not super widely used since it has a big constraint that we should give the algorithm a positive element inputs. Otherwise, it would be difficult to give plausible interpretation like in the example we just saw···

For actual fMRI data, which have abundant features, techniques such as regularization could also be an option!

Cocoan 101

https://cocoanlab.github.io