The separability assumption in NMF

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft : Janurary 19, 2019
Last update : January 20, 2019

Non-negative matrix factorization is NP-Hard

Non-egative matrix factorization (NMF) :

Given $(\mathbf{X}, r)$, where,

- $\mathbf{X}$ is a $m \times n$ matrix with non-negative entries : $\mathbf{X} \in \mathbb{R}_{+}^{m \times n}$.
- r is an integer : $1 \leq r \leq \min(m, n)$, called *the factorization rank*.

Find a pair of matrices $(\mathbf{W}, \mathbf{H})$ that

- $\mathbf{W} \in \mathbb{R}_{+}^{m \times r}$,
- $\mathbf{H} \in \mathbb{R}_{+}^{r \times n}$,

such that $\mathbf{X} = \mathbf{W}\mathbf{H}$.

NMF is NP-Hard : Stephen A. Vavasis showed NMF is equivalent to a problem in polyhedral combinatorics called Intermediate Simplex, which is NP-hard.

Due to the NP-hardness of NMF, most of the algorithms that solve NMF are based on local improvement heuristics without (global) optimality grantee.

Separable NMF

Separable NMF imposes an additional assumption on top of NMF :

Given $(\mathbf{X}, r)$, where,

- $\mathbf{X}$ is a $m \times n$ matrix with non-negative entries : $\mathbf{X} \in \mathbb{R}_+^{m \times n}$.
- r is an integer : $1 \leq r \leq \min(m, n)$, called *the factorization rank*.

Find a pair of matrices $(\mathbf{W}, \mathbf{H})$ that

- $\mathbf{W} \in \mathbb{R}_+^{m \times r}$,
- $\mathbf{H} \in \mathbb{R}_+^{r \times n}$,
- The separability condition is true

such that $\mathbf{X} = \mathbf{WH}$.

Separability condition in words

Consider the given matrix $\mathbf{X}$ are a collection of "data columns" $\mathbf{x}_i$.

Definition : consider matrix $\mathbf{H}$, for every row $i \in \{1, 2, \ldots, r\}$ of $\mathbf{H}$, there is a column $j \in \{1, 2, \ldots, n\}$ in $\mathbf{H}$ such that the only non-zero in such column is in the i^{th} row.

This definition comes from : David Donoho and Victoria Stodden. "When does non-negative matrix factorization give a correct decomposition into parts?" Advances in neural information processing systems (NIPS), 2004

This condition can be expressed in linear algebra terms.

Separability condition in linear algebra terms

Separability condition $(\mathcal{S})$ means matrix $\mathbf{H}$ has the structure

$$(\mathcal{S}) : \ \mathbf{H} = [\mathbf{D}_r \ \mathbf{H}']\mathbf{\Pi}_n,$$

where $\mathbf{D}_r$ is r-by-r diagonal matrix, $\mathbf{H}' \in \mathbb{R}_+^{r\times(n-r)}$, and $\mathbf{\Pi}_n$ is a n-by-n permutation matrix.

Denote the j^{th} column of $\mathbf{H}$ be $\mathbf{h}_{:,j}$ and let the permutation map of an integer i under permutation $\mathbf{\Pi}_n$ be $\pi(i)$. Under such notation, $(\mathcal{S})$ says $\mathbf{H}$ has (at least) r "special columns" :

- These columns are labelled as $\mathbf{h}_{:,\pi(1)}, \dots, \mathbf{h}_{:,\pi(r)}$.
- Each vector $\mathbf{h}_{:,\pi(i)}$ has only have 1 non-zero element.
  That is, all $\mathbf{h}_{:,\pi(i)}$ are $(r-1)$-sparse.
- The only non-zero element of $\mathbf{h}_{:,\pi(i)}$ is exactly at the i^{th} position.
- All $\mathbf{h}_{:,\pi(i)}$ are linearly independent to each other.
  For any two vectors $\mathbf{h}_{:,\pi(p)}, \mathbf{h}_{:,\pi(q)}$, they do not share the same non-zero position.
- These columns $\mathbf{h}_{:,\pi(i)}$ form a diagonal matrix $\mathbf{D}$ of order r

Example

Suppose $r = 3, n = 5$ and $\mathbf{H} = \begin{bmatrix} 0.8 & 0 & 0 & 2 & 0.5 \\ 5 & 0 & 1 & 0.4 & 0 \\ 0.1 & 3 & 0 & 0.6 & 0 \end{bmatrix}$.

Such $\mathbf{H}$ fulfils separability as we can re-arrange the columns of $\mathbf{H}$ as

$$\begin{aligned}
\mathbf{H}^{\text{New}} &= \mathbf{H}\mathbf{\Pi}_5 \\
&= \begin{bmatrix} 0.8 & 0 & 0 & 2 & 0.5 \\ 5 & 0 & 1 & 0.4 & 0 \\ 0.1 & 3 & 0 & 0.6 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0.5 & 0 & 0 & 0.8 & 2 \\ 0 & 1 & 0 & 5 & 0.4 \\ 0 & 0 & 3 & 0.1 & 0.6 \end{bmatrix} = [\mathbf{D}_3 \ \mathbf{H}'].
\end{aligned}$$

Now we have $\mathbf{H}^{\text{New}} = \mathbf{H}\mathbf{\Pi}_5$ so

$$\mathbf{H} = \mathbf{H}^{\text{New}}\mathbf{\Pi}_5^{-1} \overset{*}{=} [\mathbf{D}_3 \ \mathbf{H}']\mathbf{\Pi}_5,$$

where * used the fact that $\mathbf{\Pi}^{-1} = \mathbf{\Pi}$.

What separability condition does say and does not say

Separability condition $(\mathcal{S})$ says $\mathbf{H}$ has the structure

$$(\mathcal{S}): \ \mathbf{H} = [\mathbf{D}_r \ \mathbf{H}']\mathbf{\Pi}_n,$$

where $\mathbf{D}_r$ is r-by-r diagonal matrix, $\mathbf{H}' \in \mathbb{R}_+^{r\times(n-r)}$ and $\mathbf{\Pi}_n$ is a n-by-n permutation matrix.

It does not say :

- The elements in $\mathbf{D}_r$ are upper bounded[†]
  The elements in $\mathbf{D}_r$ can take any values in $[0 \ +\infty)$
  It does not say $\mathbf{D}_r$ has to be an identity matrix $\mathbf{I}_r$
- The norm of each column in $\mathbf{H}$ is upper bounded[†]
  Such norm can take any values in $[0 \ +\infty)$
  It does not say the coefficient in $\mathbf{H}$ forms a convex combination of $\mathbf{W}$.
- The expression of $\mathbf{H} = [\mathbf{D}_r \ \mathbf{H}']\mathbf{\Pi}_n$ is unique

† Due to non-negativity, these values are lower bounded by 0.

The non-uniqueness in $\mathbf{H} = [\mathbf{D}_r \ \mathbf{H}']\boldsymbol{\Pi}_n$

The re-arrangement of $\mathbf{H}$ in the raw form to the "separable form" is not unique : we can rearrange

$$\mathbf{H} = \begin{bmatrix} 0.8 & 0 & 0 & 2 & 0.5 \\ 5 & 0 & 1 & 0.4 & 0 \\ 0.1 & 3 & 0 & 0.6 & 0 \end{bmatrix}$$

as

$$\mathbf{H}^{\text{New}} = \begin{bmatrix} 0.5 & 0 & 0 & 0.8 & 2 \\ 0 & 1 & 0 & 5 & 0.4 \\ 0 & 0 & 3 & 0.1 & 0.6 \end{bmatrix}.$$

But also we can arrange $\mathbf{H}$ as

$$\mathbf{H}^{\text{New}} = \begin{bmatrix} 0.5 & 0 & 0 & 2 & 0.8 \\ 0 & 1 & 0 & 0.4 & 5 \\ 0 & 0 & 3 & 0.6 & 0.1 \end{bmatrix}$$

The non-uniqueness in $\mathbf{H} = [\mathbf{D}_r \ \mathbf{H}']\mathbf{\Pi}_n$

The re-arrangement of $\mathbf{H}$ in the raw form to the "separable form" is not unique : suppose we have

$$\mathbf{H} = \begin{bmatrix} 0.8 & 0 & 0 & 2 & 0.5 \\ 5 & 0 & 1 & 0 & 0 \\ 0.1 & 3 & 0 & 0 & 0 \end{bmatrix}$$

Then we can have

$$\mathbf{H}^{\text{New}} = \begin{bmatrix} 0.5 & 0 & 0 & 2 & 0.8 \\ 0 & 1 & 0 & 0 & 5 \\ 0 & 0 & 3 & 0 & 0.1 \end{bmatrix}$$

or

$$\mathbf{H}^{\text{New}} = \begin{bmatrix} 2 & 0 & 0 & 5 & 0.8 \\ 0 & 1 & 0 & 0 & 5 \\ 0 & 0 & 3 & 0 & 0.1 \end{bmatrix}$$

Understanding what separability condition means

With the structure $\mathbf{H} = [\mathbf{D}_r \ \mathbf{H}']\mathbf{\Pi}_n$, the expression $\mathbf{X} = \mathbf{W}\mathbf{H}$ becomes

$$\mathbf{X} = \mathbf{W}[\mathbf{D}_r \ \mathbf{H}']\mathbf{\Pi}_n = [\mathbf{W}\mathbf{D}_r \ \mathbf{W}\mathbf{H}']\mathbf{\Pi}_n.$$

What does it means?

To understand this expression, we can consider the following three special cases :

- No permutation ($\mathbf{\Pi}_n = \mathbf{I}_n$)
- The diagonal matrix $\mathbf{D}_r$ becomes identity matrix $\mathbf{I}_r$
- Both permutation and diagonal matrix become identity matrix

When both permutation matrix and diagonal matrix are identity :
$\mathbf{H} = [\mathbf{I}_r \ \mathbf{H}']\mathbf{I}_n$, the expression $\mathbf{X} = \mathbf{W}\mathbf{H}$ becomes

$$\mathbf{X} = \mathbf{W}[\mathbf{I}_r \ \mathbf{H}']\mathbf{I}_n = [\mathbf{W}\mathbf{I}_r \ \mathbf{W}\mathbf{H}'] = [\mathbf{W} \ \mathbf{W}\mathbf{H}'].$$

Recall that $\mathbf{W}$ has r columns. The expression $\mathbf{X} = [\mathbf{W} \ \mathbf{W}\mathbf{H}']$ means $\mathbf{W}$ equals to the first r columns of $\mathbf{X}$, that is, we have

$$\mathbf{X} = [\mathbf{X}(:, 1:r) \ \ \mathbf{X}(:, 1:r)\mathbf{H}'].$$

Recall that $\mathbf{H}'$ is non-negative. So $\mathbf{X}(:, 1:r)\mathbf{H}'$ describes a collection of points spanned by the first r columns in $\mathbf{X}$, with *conical* weighting coefficient $\mathbf{H}'$.

Geometrically speaking, $\mathbf{X} = \mathbf{W}\mathbf{H}$ now describes a non-negative cone : a collection of non-negative points are encapsulated inside a *convex cone* with basis (extreme rays) $\mathbf{W}$, where $\mathbf{W}$ is the first r columns of $\mathbf{X}$. All the other $n - r$ columns of $\mathbf{X}$ are some conical combinations of the first r columns of $\mathbf{X}$.

Understanding separability condition . . . (2/3)

The same geometric interpretation of non-negative cone still applies if only permutation matrix become identity : $\mathbf{H} = [\mathbf{D}_r \ \mathbf{H}']$ and the expression $\mathbf{X} = \mathbf{WH}$ becomes

$$\mathbf{X} = \mathbf{W}[\mathbf{D}_r \ \mathbf{H}'] = [\mathbf{WD}_r \ \mathbf{WH}'].$$

As $\mathbf{D}_r$ is diagonal so $\mathbf{WD}_r$ is just a scaled version of each column of $\mathbf{W}$ by $\mathbf{D}$, so the first column of $[\mathbf{WD}_r \ \mathbf{WH}']$, which is is $d_1\mathbf{w}_{:1}$, should be equal to the first column of $\mathbf{X}$, that is

$$\mathbf{x}_1 = d_1\mathbf{w}_{:1} \iff \mathbf{w}_{:1} = \frac{1}{d_1}\mathbf{x}_1.$$

In this case $\mathbf{W}$ is just a scaled version of the first r columns of $\mathbf{X}$, where the amount of scaling is encoded by the diagonal element of $\mathbf{D}_r$. We have

$$\mathbf{W} = \mathbf{X}(:, 1:r)\mathbf{D}_r^{-1},$$

and $\mathbf{X} = \mathbf{WH}$ becomes

$$\mathbf{X} = [\mathbf{X}(:, 1:r) \ \ \mathbf{X}(:, 1:r)\mathbf{D}_r^{-1}\mathbf{H}'] \stackrel{*}{=} [\mathbf{X}(:, 1:r) \ \ \mathbf{X}(:, 1:r)\mathbf{H}'].$$

where in $*$ the matrix $\mathbf{D}_r^{-1}$ is absorbed into $\mathbf{H}'$. As $\mathbf{D}_r$ is non-negative diagonal matrix, hence $\mathbf{D}_r^{-1}\mathbf{H}'$ is still a valid non-negative coefficient matrix.

If only diagonal matrix become identity : $\mathbf{H} = [\mathbf{I}_r \ \mathbf{H}']\mathbf{\Pi}_n$ and the expression $\mathbf{X} = \mathbf{W}\mathbf{H}$ becomes

$$\mathbf{X} = \mathbf{W}[\mathbf{I}_r \ \mathbf{H}']\mathbf{\Pi}_n = [\mathbf{W} \ \mathbf{W}\mathbf{H}']\mathbf{\Pi}_n.$$

In this case, we have

$$[\mathbf{W} \ \mathbf{W}\mathbf{H}'] = \mathbf{X}\mathbf{\Pi}_n^{-1} = \mathbf{X}\mathbf{\Pi}_n,$$

meaning that $\mathbf{W}$ is the first r columns of $\mathbf{X}\mathbf{\Pi}_n$. Or in other words, $\mathbf{W}$ come from a r-subset of columns of $\mathbf{X}$, which can be expressed as

$$\mathbf{W} = \mathbf{X}(:, \mathcal{K}),$$

where $\mathcal{K}$ holds the permutation indices $\pi(1), \ldots, \pi(r)$.

The same geometric interpretation of non-negative cone still applies.

Separability condition means self-expressiveness

From the expressions

$$
\begin{aligned}
\text{No permutation nor scaling} \quad & \mathbf{W} = \mathbf{X}(:, 1:r) \\
\text{No permutation} \quad & \mathbf{W} = \mathbf{X}(:, 1:r)\mathbf{D}_r^{-1} \\
\text{No scaling} \quad & \mathbf{W} = \mathbf{X}(:, \mathcal{K}) \\
\text{General case} \quad & \mathbf{W} = \mathbf{X}(:, \mathcal{K})\mathbf{D}_r^{-1}
\end{aligned}
$$

we can see that the basis $\mathbf{W}$ are a scaled, permuted version of *certain r columns of* $\mathbf{X}$ .That is, the basis *comes from* the data : we have a *self-expressive dictionary* where there are r specific points among the dataset (conically) span all the other $n - r$ data points.

Therefore the key is to find such index set $\mathcal{K}$.

Simplifying the separability condition

In the general form, separability condition requires $\mathbf{H} = [\mathbf{D}_r \; \mathbf{H}']$. where the permutation is moved into $\mathbf{W}$ such that $\mathbf{W} = \mathbf{X}(:, \mathcal{K})\mathbf{D}_r^{-1}$.
Hence the expression $\mathbf{X} = \mathbf{W}\mathbf{H}$ becomes

$$\mathbf{X} = \mathbf{W}\mathbf{H} = \mathbf{X}(:, \mathcal{K})\mathbf{D}_r^{-1}[\mathbf{D}_r \; \mathbf{H}'] = [\mathbf{X}(:, \mathcal{K}) \; \mathbf{X}(:, \mathcal{K})\mathbf{D}_r^{-1}\mathbf{H}'],$$

which, again we absorb $\mathbf{D}_r^{-1}$ into $\mathbf{H}'$, we get

$$\mathbf{X} = [\mathbf{X}(:, \mathcal{K}) \;\; \mathbf{X}(:, \mathcal{K})\mathbf{H}'].$$

In this case, let $\mathbf{W} = \mathbf{X}(:, \mathcal{K})$, we can says that, without lost of generality, separability condition means $\mathbf{H}$ has the following structure

$$\mathbf{H} = [\mathbf{I}_r \;\; \mathbf{H}']\mathbf{\Pi}.$$

That is, we can kind of "ignore" the scaling $\mathbf{D}_r$ and concentrate only on finding the indices set $\mathcal{K}$. This the most important thing in separable NMF is to find $\mathcal{K}$.

And finding such $\mathcal{K}$ is not NP-hard.

End of document