Convergence analysis of NMF algorithm

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

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

First draft : June 6, 2017
Last update : February 27, 2018

Overview

Non-negative Matrix Factorization

Given a non-negative matrix $X \in \mathbb{R}_+^{m \times n}$, NMF aims to find $W \in \mathbb{R}_+^{m \times r}$ and $H \in \mathbb{R}_+^{r \times n}$ by solving the optimization problem

$$[W^* H^*] = \underset{W \geq 0, H \geq 0}{\arg\min} \ f(W, H) = \frac{1}{2} \|X - WH\|_F^2$$

The following multiplicative update formulae can be derived from gradient descent

$$\begin{aligned} \left[W^{k+1}\right]_{ij} &= \left[W^k\right]_{ij} \frac{\left[X (H^k)^T\right]_{ij}}{[W^k H^k (H^k)^T]_{ij}} \\ \left[H^{k+1}\right]_{ij} &= \left[H^k\right]_{ij} \frac{\left[W^k X\right]_{ij}}{[(W^k)^T W^k H^k]_{ij}} \end{aligned}$$

The theorem of non-increasing error norm of NMF update

Theorem (Lee and Seung 2001)[1] The Eucledian norm $\frac{1}{2}\|X - WH\|_F$ is *non-increasing* under the following updates

$$\left[W^{k+1}\right]_{ij} = \left[W^{k}\right]_{ij} \frac{\left[X(H^k)^T\right]_{ij}}{[W^k H^k (H^k)^T]_{ij}} \quad (1)$$

$$\left[H^{k+1}\right]_{ij} = \left[H^{k}\right]_{ij} \frac{\left[W^k X\right]_{ij}}{[(W^k)^T W^k H^k]_{ij}} \quad (2)$$

The aim of this document is to study the proof of this theorem.

[1] Lee, D. D., Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In Advances in neural information processing systems (pp. 556-562).

The proof of the theorem - idea ... (1/3)

For simplicity, we can first focus on updating the objective function $\frac{1}{2}\|X - WH\|_F$ on a row in H, denoted as h, for a fixed W. In this case $f(h)$It can be expressed as

$$f(h) = \frac{1}{2}\|Wh - x\|_2^2.$$

At the k^{th} iteration, the second order Taylor expansion around h is

$$F(h) = f(h^k) + \nabla f^T(h^k)(h - h^k) + \frac{1}{2}(h - h^k)^T \nabla^2 f(h^k)(h - h^k),$$

where h^k is a constant (usually treated as the previous iterate of h).

The proof uses the technique of Majorization Minimization : a surrogate function (upper bound function) G has to be constructed that majorizes (upper bound) F. A simple way to form G is

$$G(h|h^k) = f(h^k) + \nabla f^T(h^k)(h - h^k) + \frac{1}{2}(h - h^k)^T M(h - h^k)$$

The proof of the theorem - idea ... (2/3)

We now have

$$\begin{aligned} F(h) &= f(h^k) + \nabla f^T(h^k)(h-h^k) + \frac{1}{2}(h-h^k)^T \nabla^2 f(h^k)(h-h^k), \\ G(h|h^k) &= f(h^k) + \nabla f^T(h^k)(h-h^k) + \frac{1}{2}(h-h^k)^T M (h-h^k). \end{aligned}$$

For the surrogate G to upper bound the original function F, we need to have the matrix

$$M \succeq \nabla^2 f.$$

How to construct M?

The proof of the theorem - idea ... (3/3)

To build M that $M \succeq \nabla^2 f$, we first need to know more about ∇f :

For $f(h) = \frac{1}{2}\|x - Wh\|_2$, $\nabla_h^2 \frac{1}{2}\|x - Wh\|_2^2 = W^T W$.

- $W^T W$ is symmetric
- W is non-negative $\implies$ $W^T W$ is also non-negative.

These properties of $W^T W$ lead to the following lemma to construct M.

Lemma. For a non-negative symmetric matrix $A \in \mathbb{R}_+^d$ and a positive vector $x \in \mathbb{R}_{++}^d$, the following matrix is positive semi-definite:

$$\hat{A} := \text{Diag}\left(\frac{[Ax]_i}{[x]_i}\right) - A$$

A lemma ... (1/3)

Lemma. For a non-negative symmetric matrix $A \in \mathbb{R}^d_+$ and a positive vector $x \in \mathbb{R}^d_{++}$, the following matrix is positive semi-definite:

$$\hat{A} := \text{Diag}\left(\frac{[Ax]_i}{[x]_i}\right) - A$$

Proof. Let vector $y = Ax$. Then

$$\begin{aligned}
\text{Diag}\left(\frac{[Ax]_i}{[x]_i}\right) &= \text{Diag}\left(\frac{[y]_i}{[x]_i}\right) \\
&= \frac{\text{Diag}(y)}{\text{Diag}(x)} \\
&= D_x^{-1} D_y \\
&= \begin{bmatrix} \frac{y_1}{x_1} & 0 & \dots & 0 \\ 0 & \frac{y_2}{x_2} & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{y_d}{x_d} \end{bmatrix}
\end{aligned}$$

where $D_x = \text{Diag}(x)$

A lemma ... (2/3)

Consider $D_x \hat{A} D_x$

$$\begin{aligned}
D_x \hat{A} D_x &= D_x \text{Diag}\left(\frac{[Ax]_i}{[x]_i}\right) D_x - D_x A D_x \\
(\text{As } D_x = \text{Diag}(x)) &= \text{Diag}\left([Ax]_i\right) D_x - D_x A D_x \\
(*) &= \text{Diag}\left([Ax]_i\right) D_x - D_x^2 A \\
(\text{As } y = Ax) &= \begin{bmatrix} y_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & y_d \end{bmatrix} \begin{bmatrix} x_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & x_d \end{bmatrix} - D_x^2 A \\
&= \begin{bmatrix} y_1 x_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & y_d x_d \end{bmatrix} - \begin{bmatrix} x_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & x_d^2 \end{bmatrix} A
\end{aligned}$$

(*) : diagonal matrix D commute with symmetric matrix A : we have $DAD = D(AD) = D(DA) = D^2 A$

A lemma ... (3/3)

$$D_x \hat{A} D_x = \begin{bmatrix} y_1 x_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & y_d x_d \end{bmatrix} - \begin{bmatrix} x_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & x_d^2 \end{bmatrix} A$$

Note $D_x \hat{A} D_x$ is *diagonally dominant*. That is

$$\left| [D_x \hat{A} D_x]_{ii} \right| \geq \sum_{j \neq i} \left| [D_x \hat{A} D_x]_{ij} \right| \quad \forall i$$

Fact : a symmetric diagonally dominant real matrix with nonnegative diagonal entries is positive semidefinite (psd). So $D_x \hat{A} D_x$ is psd.

As D_x is diagonal matrix with positive element, hence D_x^{-1} is also a diagonal matrix with positive element, hence $\hat{A}$ is also psd. □

The proof of the NMF theorem

Based on the lemma, $M \succeq \nabla^2 f$ for the following M and ∇f

$$\begin{aligned} \nabla^2 f(h^{(k)}) &= W^T W \\ M &= \text{Diag}\left(\frac{[W^T W h^{(k)}]_i}{[h^{(k)}]_i}\right) \end{aligned}$$

Hence $G\left(h|h^{(k)}\right)$ (the surrogate) and F (the second order Taylor expansion of the objective function of NMF on variable h) :

$$F(h) = f(h^{(k)}) + \nabla f^T(h^{(k)})(h - h^{(k)}) + \frac{1}{2}(h - h^{(k)})^T \nabla^2 f(h^{(k)})(h - h^{(k)})$$

$$G\left(h|h^{(k)}\right) = f(h^{(k)}) + \nabla f^T(h^{(k)})(h - h^{(k)}) + \frac{1}{2}(h - h^{(k)})^T M (h - h^{(k)})$$

satisfy

$$F(h^{k+1}) = G\left(h^{(k+1)}|h^{(k)}\right) \leq G\left(h|h^{(k)}\right) = F(h^k)$$

which proves the theorem on variable h. The proof of the theorem on variable w will be similar.

The update

For the update, $h^{(k+1)}$ can be obtained by solving

$$\frac{\partial G(h|h^{(k)})}{\partial h} = 0$$

$$G\left(h|h^{(k)}\right) = f(h^{(k)}) + \nabla f^T(h^{(k)})(h - h^{(k)}) + \frac{1}{2}(h - h^{(k)})^T M(h - h^{(k)})$$

Gives $\dfrac{\partial G(h|h^{(k)})}{\partial h} = \nabla f(h^{(k)}) + M(h - h^{(k)})$

Thus $h = h^{(k)} - M^{-1}\nabla f(h^{(k)})$

The update

As $M = \text{Diag}\left(\frac{[W^T W h^{(k)}]_i}{[h^{(k)}]_i}\right) \implies M^{-1} = \text{Diag}\left(\frac{[h^{(k)}]_i}{[W^T W h^{(k)}]_i}\right)$

And $\nabla f(h^{(k)}) = W^T(Wh^{(k)} - x)$

Hence

$$\begin{aligned} h &= h^{(k)} - M^{-1}\nabla f(h^{(k)}) \\ &= h^{(k)} - \text{Diag}\left(\frac{[h^{(k)}]_i}{[W^T W h^{(k)}]_i}\right) W^T(Wh^{(k)} - x) \\ &= h^{(k)} - h^{(k)} \otimes \frac{\left[W^T(Wh^{(k)} - x)\right]_{ij}}{[W^T W h^{(k)}]_{ij}} \\ &= h^{(k)} \otimes \frac{\left[W^T x\right]_{ij}}{[W^T W h^{(k)}]_{ij}} \end{aligned}$$

which is the update formula stated in the theorem

Last page - summary

- The non-decreasing norm theorem of NMF update
- A lemma on diagonal matrix from non-negative matrix
- The proof of the NMF thorem

End of document