Non-negative Matrix Factorization with Regularization

Andersen Ang

Mathématique et de Recherche opérationnelle
Faculté polytechnique de Mons
UMONS
Mons, Belgium

email: manshun.ang@umons.ac.be
homepage: *angms.science*

June 6, 2017

Overview

1. Introduction to regularized NMF Problem
2. Orthogonal NMF

The regularized NMF Problem

Given a non-neagtive matrices $X \in \mathbb{R}_{+}^{m\times n}$, find two non-neagtive matrices $W \in \mathbb{R}_{+}^{m\times r}$ and $H \in \mathbb{R}_{+}^{r\times n}$ such that

$$[W * H*] = \arg \min_{W\geq 0, H\geq 0} \frac{1}{2}\|X - WH\|_F^2 + \lambda g\left(W, H\right)$$

$f\left(W, H\right)$ is the primary objective function and $g\left(W, H\right)$ is the regularization and λ is the weight.

The Gradient Descent update schemes

Gradient Descent update schemes can be used:

$$W^{k+1} = W^k - t_W^k \nabla_W (f + \lambda g)(W, H)$$

$$H^{k+1} = H^k - t_H^k \nabla_H (f + \lambda g)(W, H)$$

The terms $\nabla f(W, H)$ are already derived and thus

$$W^{k+1} = W^k - t_W^k (W^k H^k - X)(H^k)^T - t_W^k \nabla_W \lambda g(W, H)$$

$$H^{k+1} = H^k - t_H^k (W^k)^T (W^k H^k - X) - t_H^k \nabla_H \lambda g(W, H)$$

The Gradient Descent update

Step size t can be defined in a similar way as in NMF and we get the multiplicative update

$$W^{k+1} = W^k \otimes \Big(\frac{[X(H^k)^T]}{[W^k H^k (H^k)^T]} - \frac{[W^k]}{[W^k H^k (H^k)^T]} \nabla_W \lambda g(W,H) \Big)$$

$$H^{k+1} = H^k \otimes \Big(\frac{[W^k X]}{[(W^k)^T W^k H^k]} - \frac{[H^k]}{[(W^k)^T W^k H^k]} \nabla_H \lambda g(W,H) \Big)$$

where $\otimes$ and $\frac{[.]}{[.]}$ are Hadamard product and division repsectively.

In compact form:

$$W^{k+1} = W^k \otimes \frac{[X(H^k)^T - \nabla_W \lambda g(W,H)]}{[W^k H^k (H^k)^T]}$$

$$H^{k+1} = H^k \otimes \frac{[W^k X - \nabla_H \lambda g(W,H)]}{[(W^k)^T W^k H^k]}$$

The overall picture

Hence, the regularized optimization problem:

$$[W * H*] = \arg \min_{W \geq 0, H \geq 0} \frac{1}{2} \|X - WH\|_F^2 + \lambda g\left(W, H\right)$$

Can be solved by the following multiplicative update

$$\begin{aligned} W^{k+1} &= W^k \otimes \frac{\left[X(H^k)^T - \lambda \nabla_W g(W, H)\right]}{\left[W^k H^k (H^k)^T\right]} \\ H^{k+1} &= H^k \otimes \frac{\left[W^k X - \lambda \nabla_H g(W, H)\right]}{\left[(W^k)^T W^k H^k\right]} \end{aligned}$$

Now the key is on $g(W, H)$ and ∇g.

Orthogonal NMF

Consider the orthogonality constraints

$$\min_{W\geq 0, H\geq 0} \frac{1}{2}\|X - WH\|_F^2 \text{ s.t. } W^TW = I \text{ and } H^TH = I$$

The optimization formulation becomes

$$\min_{W\geq 0, H\geq 0} \frac{1}{2}\|X - WH\|_F^2 + \lambda\|W^TW - I\|_F^2 + \mu\|H^TH - I\|_F^2$$

So in this case $g(W, H)$ can be speparated into two independent functions $\|W^TW - I\|_F^2$ and $\|H^TH - I\|_F^2$.

Now we need to know $\nabla_W\|W^TW - I\|_F^2$ and $\nabla_H\|H^TH - I\|_F^2$.

Evaluating gradient of $\|W^TW - I\|_F^2$

$$
\begin{aligned}
\|W^TW - I\|_F^2 &= \mathsf{Tr}\left((W^TW - I)^T(W^TW - I)\right) \\
&= \mathsf{Tr}\left((W^TW)^T(W^TW) - 2W^TW + I\right) \\
\nabla\|W^TW - I\|_F^2 &= \nabla\mathsf{Tr}\left((W^TW)^T(W^TW)\right) - 2\nabla\mathsf{Tr}\left(W^TW\right) \\
\nabla\mathsf{Tr}\left(W^TW\right) &= 2W \\
\nabla\mathsf{Tr}\left((W^TW)^T(W^TW)\right) &= \frac{\partial\mathsf{Tr}\left((W^TW)^TW^TW\right)}{\partial W} \\
&= \frac{\partial\mathsf{Tr}\left((W^TW)^TW^TW\right)}{\partial W^TW}\frac{\partial W^TW}{\partial W} \\
&= \nabla_U\mathsf{Tr}\left(U^TU\right)\nabla_W\mathsf{Tr}\left(W^TW\right), U = W^TW \\
&= 2U2W \\
&= 4W^TWW \\
\therefore \nabla\|W^TW - I\|_F^2 &= 4(W^TW - I)W
\end{aligned}
$$

Orthogonal NMF Updating

Now we have

$$
\begin{aligned}
\nabla_W \|W^T W - I\|_F^2 &= 4(W^T W - I)W \\
\nabla_H \|H^T H - I\|_F^2 &= 4(H^T H - I)H
\end{aligned}
$$

Hence, for the orthogonal NMF problem

$$
\min_{W \geq 0, H \geq 0} \frac{1}{2}\|X - WH\|_F^2 + \lambda \|W^T W - I\|_F^2 + \mu \|H^T H - I\|_F^2
$$

Thus the update are

$$
\begin{aligned}
W^{k+1} &= W^k \otimes \frac{\left[X(H^k)^T - 4\lambda((W^k)^T W^k - I)W^k\right]}{[W^k H^k (H^k)^T]} \\
H^{k+1} &= H^k \otimes \frac{\left[W^k X - 4\mu(H^k)^T H^k - I)H^k\right]}{[(W^k)^T W^k H^k]}
\end{aligned}
$$

Last page - summary

- Introduction of regularized NMF

$$[W*H*] = \arg\min_{W\geq 0, H\geq 0} \frac{1}{2}\|X - WH\|_F^2 + \lambda g\left(W, H\right)$$

- Multiplicative update for regularized NMF

$$\begin{aligned} W^{k+1} &= W^k \otimes \frac{\left[X(H^k)^T - \lambda \nabla_W g(W,H)\right]}{\left[W^k H^k (H^k)^T\right]} \\ H^{k+1} &= H^k \otimes \frac{\left[W^k X - \lambda \nabla_H g(W,H)\right]}{\left[(W^k)^T W^k H^k\right]} \end{aligned}$$

- Orthogonal NMF

$$\min_{W\geq 0, H\geq 0} \frac{1}{2}\|X - WH\|_F^2 + \lambda\|W^TW - I\|_F^2 + \mu\|H^TH - I\|_F^2$$

- Derivation of the multiplicative update for orthogonal NMF

End of document