Nonnegative Matrix Factorization

Be Positive!

Abdelbast Nassiri
Maximilian Stollmayer
Manuel Wissiak

30.01.2023

NMF

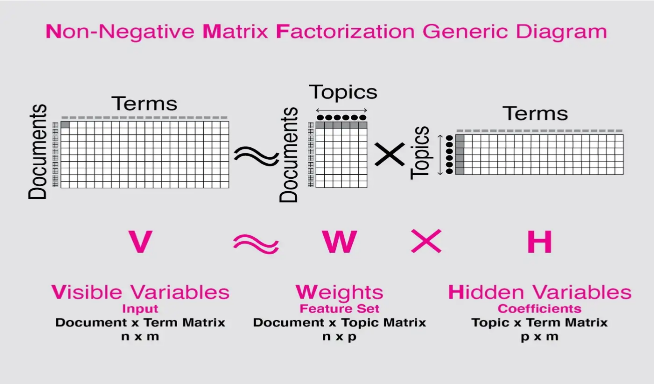
Problem

Given $A \in \mathbb{R}_{+}^{m \times n}$ non-negative and *rank* $r \leq \min(m, n)$.
Find $W \in \mathbb{R}_{+}^{m \times r}$ and $H \in \mathbb{R}_{+}^{r \times n}$ both non-negative s.t.:

$$A \approx WH$$

Text Mining

"Words which are similar in meaning occur in similar contexts"



Source: "NMF — A visual explainer and Python Implementation", Anupama Garla

Image Processing

Given vectorized gray-levels $X \in \mathbb{R}_+^{p \times n}$ of n facial images.

Problem: Facial Feature Extraction

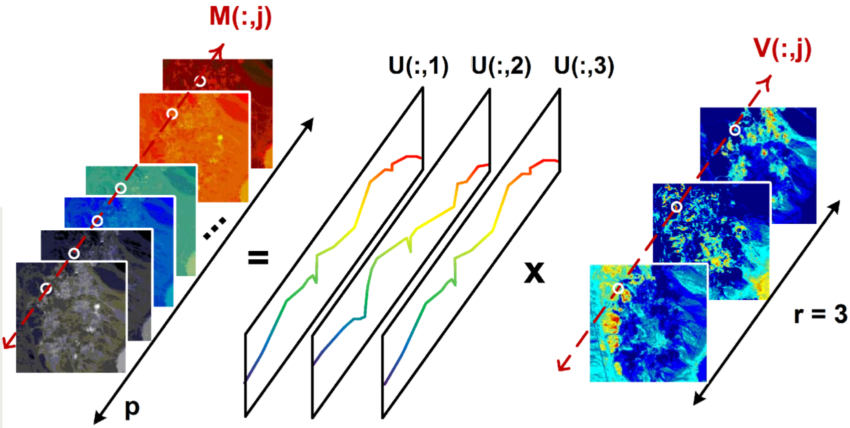
$X(:, j) \approx$	$\sum_{k=1}^{r} W(:, k)$	$H(k, j)$
j-th facial image	facial features	importance of feature in j-th image

Hyperspectral Unmixing

Given vectorized spectral signature $X \in \mathbb{R}_+^{p \times n}$ of an image.
Problem: Identify Endmembers (Grass, Stone,...)

$X(:,j) \approx$	$\sum_{k=1}^r W(:,k)$	$H(k,j)$
spectral signature	spectral signature of k-th endmember	abundance of k-th endmember

Hyperspectral Unmixing



Source: "NMF — A visual explainer and Python Implementation", Anupama Garla

Problem Reformulation

Optimization Problem

Given $A \in \mathbb{R}_+^{m\times n}$ non-negative and *rank* $r \leq \min(m,n)$.

$$\min_{W\in\mathbb{R}_+^{m\times r},\, H\in\mathbb{R}_+^{r\times n}} \quad \|A - WH\|_F^2$$

Note that $F(W,H) = \|A - WH\|_F^2$ is convex in U and convex in V, but not in both!

Stationary Points & KKT-Conditions

Checking the KKT-Conditions for $F(W, H)$ yields the following:

$$W \geq 0,\ \nabla_W F = WHH^T - AH^T \geq 0,\ \nabla_W F * W = 0$$

$$H \geq 0,\ \nabla_W F = W^T WH - W^T A \geq 0,\ \nabla_H F * H = 0$$

Stationary Points

A pair (U, V) is called a *stationary Point*, if and only if U and V satisfy the KKT-Conditions.

Stationary Points & KKT-Conditions

From the KKT-Conditions simple characteristics of the solutions can be derived:

Theorem

Suppose (W, H) be a stationary point of the problem, then it holds:

$$\frac{1}{2}\|A - WH\|^2 = \frac{1}{2}(\|A\|^2 - \|WH\|^2)$$

This furthermore implies that $\|A\|^2 \geq \|WH\|^2$, with equality attained only at the exact factoriztion.

Coordinate Descent

For Ω (pointwise) convex

$$solve \ \min_{x \in \Omega} f(x)$$

Initialization: $x \in \mathbb{R}^n$

for $t \leftarrow 1, 2, ..., n$ **do**

$\quad solve \ x_i = \arg\min_{\zeta in \Omega_i} f(x_1, ..., x_{i-1}, \zeta, x_{i+1}, ..., x_n)$

end

Algorithm 1: General Coordinate Descent

"Convergence" Theorem

"Convergence" to stationary Points

Suppose f is continuously differentiable and furthermore that

$$\forall i\, \forall x : \min_{\zeta \in \Omega_i} f(x_1, \ldots, x_{i-1}, \zeta, x_{i+1}, \ldots, x_n)$$

is uniquely attained. Let $\{x^k\}$ be the sequence generated by the *Coordinate Descent*, then every limit point is a stationary point.

Exact Factorization

Now consider the case where A is exactly factorized by WH.

minimal rank

The smallest r such that exist $W \in \mathbb{R}_+^{m \times r}$ *and* $H \in \mathbb{R}_+^{r \times n}$ such that $A = WH$, is called the *inner rank* and denoted by $rank_{WH}^+(A)$.

Determining the inner rank

One algorithm to determine if A can be factorized with *inner rank r* would be the Renegar algorithm, which scales $(6mn)^{\mathcal{O}(mn)}$.

Since $rank^{+}_{WH}(A) \leq \min(m, n)$ this can be done in finite time.

Vavasis, 2008

- exact *NMF* is *NP-hard*
- $\exists$ polynomial time local search heuristics

Algorithms for NMF

The Multiplicative Update Rule

$$\min_{W,H>0} f(W,H) = \min_{W,H>0} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} (A_{ij} - (WH)_{ij})^2 \tag{1}$$

The most used approach to minimize (1) is a simple multiplicative update method proposed by Lee and Seung (2001):

This algorithm is just a special case of Gradient Descent with a step size

$$\epsilon(W^t) = \frac{W^t}{W^t H^t (H^t)^T}$$

$$\epsilon(H^t) = \frac{H^t}{(W^t)^T W^t H^t}$$

The Multiplicative Update Rule

Initialization: $W^1, H^1 > 0$;

for $t \leftarrow 1, 2.. \ldots$ **do**

$$W^{t+1} = W^t \frac{A(H^t)^T}{W^t H^t (H^t)^T};$$

$$H^{t+1} = H^t \frac{(W^{t+1})^T A}{(W^{t+1})^T W^{t+1} H^t};$$

end

Algorithm 2: MUR Algorithm

The Multiplicative Update Rule

This algorithm is a type of fixed-point method, meaning that if $(H^t)^T H^t W^t \neq 0$ and $W^{t+1} = W^t$, then $A(H^t)^T = W^t H^t (H^t)^T$ implies $\nabla_W f(W^t, H^t) = 0$,
which is part of the KKT condition.

"Convergence" Theorem

Theorem Lee and Seung

-The Euclidean distance $||A - WH||$ is non-increasing under the update rules

$$W \leftarrow W \frac{AH^T}{WHH^T}, \qquad H \leftarrow H \frac{W^T A}{W^T WH}$$

-The Euclidean distance is invariant under these updates if and only if W and H are at a stationary point of the distance.

Weaknesses

- Lee and Seung claim that this algorithm "converges" to a stationary point. However, it has been shown in 2005 that this claim is wrong, as having the cost function non-increasing may not imply convergence.

- Therefore, the algorithm still lacks optimization properties.

Weaknesses

- We can only make the following statement about the convergence of these algorithms: "When the algorithm has converged to a limit point, this point is a stationary point."

- Also it has been repeatedly shown that the convergence is notoriously slow.

Modifications: Convergence vs. speed trade-off

- Lin in 2007 proposed a modification that is guaranteed to converge to a stationary point. However, it requires more work per iteration so it is even slower.

- The Fast Multiplicative Update Rule Algorithm from 2014 is faster than the two algorithms in the case of convergence.

Comparison of the the three Algorithms

table.jpg

Source: "A Fast Algorithm for Non-negative Matrix Factorization and Its Convergence", Li, Wu, and Zhang

Alternating Non-negative Least Squares (ANLS)

- In this algorithms, a least squares step is followed by another least squares step in an alternating fashion, thus giving rise to the ALS name.
- The Alternating Least Squares - ALS algorithms were first introduced by Paatero 1994 , who initially invented the whole NMF Theory.

Initialization: $W > 0$;

for $t \leftarrow 1, 2 \ldots$ **do**

Solve for H the LS equation $W^T W H = W^T A$;

Set all negative elements in H to 0;

Solve for W the LS equation $W H H^T = A H^T$;

Set all negative elements in W to 0;

end

Algorithm 3: Basic ALS for NMF

Convergence Theorem

Theorem

Any limit point of the sequence $\{W^t, H^t\}$ *generated by ALS Algorithm is a stationary point.*

Comparison between ANLS and MU

ANLS	MU
+ Can be very fast depending on the implementation + Aids sparsity	+ Easy to use
- Once an element in W or H becomes 0, it must remain 0.	- Notoriously slow - Lacks optimization properties.

Comparison between ANLS and MU

comparison.png

Source: "Fast optimization of non-negative matrix tri-factorization", Zupan, Zitnik

References I

- Michael W. Berry, Murray Browne, Amy N. Langville, V. Paul Pauca, and Robert J. Plemmons.
  Algorithms and applications for approximate nonnegative matrix factorization.
  Computational Statistics Data Analysis, 52(1):155–173, 2007.
- Nicolas Gillis.
  The why and how of nonnegative matrix factorization.
  Regularization, Optimization, Kernels, and Support Vector Machines, 12, 01 2014.
- Nicolas Gillis.
  Introduction to nonnegative matrix factorization.
  03 2017.

References II

Ngoc-Diep Ho.
Nonnegative matrix factorization algorithms and applications.
PhD thesis, 06 2008.

Li-Xin Li, Lin Wu, Hui-Sheng Zhang, and Fang-Xiang Wu.
A fast algorithm for nonnegative matrix factorization and its convergence.
IEEE Transactions on Neural Networks and Learning Systems, 25(10):1855–1863, 2014.

Stephen A. Vavasis.
On the complexity of nonnegative matrix factorization.
SIAM Journal on Optimization, 20(3):1364–1377, 2010.