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"

CHATGPT

picture

Visible Variables = Weights * Hidden Variables

Image Processing

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

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

Problem v2

Optimization Problem

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

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

Note that $F(W, H) = \|A - WH\|$ 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 $\|A\|^2 \geq \|WH\|^2$, which is only fulfilled at the exact factorization.

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^T.

minimal rank

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

Exact Factorization

Lemma

$$rank(A) \leq rank^+_{WH}(A) \leq \min(m, n)$$

existence of exact factorization of A of rank $r \iff$ determining $rank^+_{WH}(A)$

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 the Gradient Descent with a step size

$$\epsilon(W_{ia}^t) = \frac{w_{ia}^t}{[W^t H^t (H^t)^T]_{ia}} \quad \forall i, a$$

$$\epsilon(H_{bj}^t) = \frac{h_{bj}^t}{[(W^t)^T W^t H^t]_{bj}} \forall b, j$$

Initialization: $w_{ia}^1, h_{bj}^1 > 0 \quad \forall i, j, a, b$;

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

$$w_{ia}^{t+1} = w_{ia}^t \frac{(A(H^t)^T)_{ia}}{[W^t H^t (H^t)^T]_{ia}} \quad \forall i, a;$$

$$h_{bj}^{t+1} = h_{bj}^t \frac{((W^{t+1})^T A)_{ia}}{[(W^{t+1})^T W^{t+1} H^t]_{bj}} \quad \forall j, b;$$

end

Algorithm 2: The Multiplicative Update Rule

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

Proof of convergence of the Multiplicative Update Rule Algorithm

Theorem Lee and Seung, 2001

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

$$w_{ia} \leftarrow w_{ia} \frac{(AH^t)_{ia}}{[WH(H)^T]_{ia}}, \qquad h_{bj} \leftarrow h_{bj} \frac{(W^T A)_{ia}}{[(W)^T WH]_{bj}}$$

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

Weaknesses and modifications:

Lee and Seung claim that the limit of the sequence $\{W^t, H^t\}$ is a stationary point (i.e., a point satisfying the KKT condition). However, Gonzales and Zhang (2005) indicate that this claim is wrong as having the cost function non increasing under the update may not imply the convergence.

Therefore, this multiplicative update method still lacks optimization properties.

we can only make the following statement about the convergence of the Lee and Seung multiplicative update 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.

A number of modifications to the original Lee and Seung algorithms have been introduced with the objective to overcome these shortcomings.

For instance, Lin in 2007 proposed a modification that is guaranteed to converge to a stationary point, this algorithm requires more work per iteration than the already slow Lee and Seung algorithm.

Fast Multiplicative Update Rule Algorithm by Li-Xin Li, Lin Wu, Hui-Sheng Zhang, and Fang-Xiang Wu (2014) Li-Xin Li, Lin Wu, Hui-Sheng Zhang, and Fang-Xiang Wu Presented a comparison between the three Algorithms in there paper in 2014.
Three algorithms are programmed in MATLAB R2013a and run on a computer with the following specifications: a processor of Intel Core 2 Quad CPU Q9450 at 2.66 GHZ 2.67 GHZ and a RAM of 4 GB (3.72 GB usable) **Note:**

	MU Algorithm	Updated MU	Fast MU
initial values in case I			
CPUTime(s)	110.14	129.80	**106.60**
Iteration	2730.37	2978.53	**776.80**
OBJ.ave	149450.1	149354.3	**148681.0**
OBJ.std	35.39	43.57	**27.79**
initial values in case II			
CPUTime(s)	91.66	130.12	**88.18**
Iteration	2518.97	3330.40	**741.97**
OBJ.ave	149914.2	149290.5	**148639.6**
OBJ.std	34.86	48.45	**22.33**

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 $W^T WH = W^T A$;

Set all negative elements of in H to 0;

Solve for W the LS equation $WHH^T = AH^T$;

Set all negative elements of in W to 0;

end

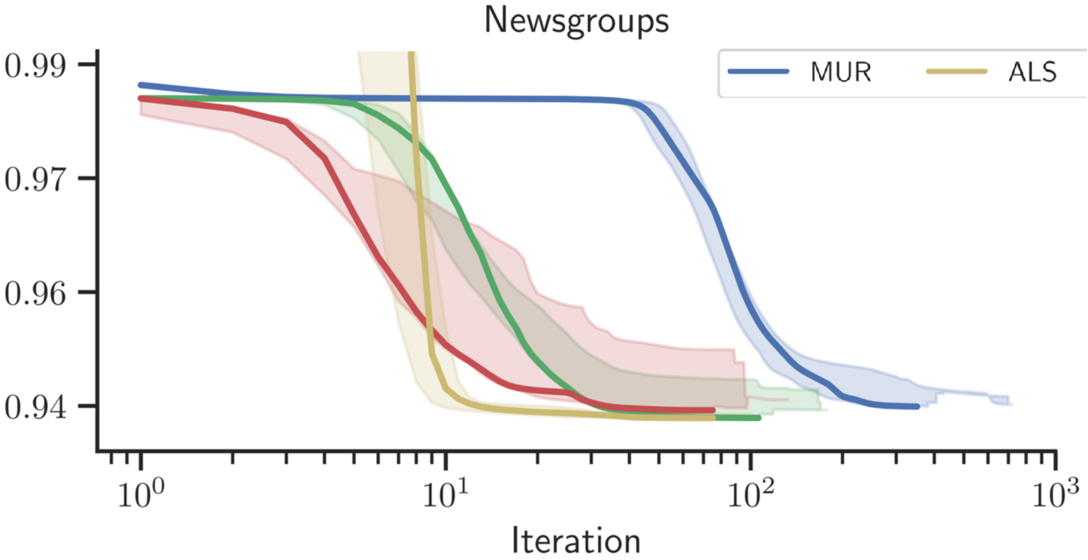
Algorithm 3: Basic ALS for NMF

Theorem

Any limit point of the sequence $\{W^t, H^t\}$ *generated by ALS Algorithm is a stationary point of (1).*

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 -the work per iteration is high since each iter requires $O(mnk)$ work.


Newsgroups
MUR
ALS
0.99
0.97
0.96
0.94
10^0
10^1
10^2
10^3
Iteration