

Nonnegative Matrix Factorization (NNMF)

Note Title

LECTURE 20

★ What is NNMF?

- a type of **low-rank approximation** of a given matrix $A \in \mathbb{R}^{m \times n}$ where **$a_{ij} \geq 0 \quad \forall i, j$** .
- Factors must be **nonnegative**.
- Certain applications (e.g., text mining, chemometrics, etc.) require nonnegativity in all the factors involved.
- SVD, PCA cannot be used because they involve negative coefficients, negative entries in the factors (i.e., entries of U, V etc.)

★ NNMF Objective

Given a nonnegative matrix $A \in \mathbb{R}^{m \times n}$ and $k < \min(m, n)$, find nonnegative matrices $W \in \mathbb{R}^{m \times k}$, $H \in \mathbb{R}^{k \times n}$ to minimize the objective function

$$J_{\text{NNMF}}(W, H) := \frac{1}{2} \|A - WH\|_F^2$$

The product WH is called an (approximate) **NNMF** of A .

The choice of k is critical in practice, but often $k \ll \min(m, n)$
 $\Rightarrow$ a **Compressed approx.** of A .

LECTURE 20

★ Numerical Approaches for NNMF

- Minimization of J_{NNMF} is difficult:
 - Many local minima exist in J_{NNMF} in both W & H .

- Lack of a unique solution

Consider $D \in \mathbb{R}^{k \times k}$, nonnegative and nonsingular, and suppose D^{-1} is also nonnegative (e.g., D could be $\text{diag}(d_1, \dots, d_k)$ with $d_j > 0 \quad 1 \leq j \leq k$.)

Then if WH is an NNMF of A , so is $WDD^{-1}H$.

- Many algorithms have been proposed. We'll discuss only one of them based on the so-called

Alternating Least Squares (ALS).

Algorithm (ALS-NNMF)

rand(m,k)

returns

m x k

random

matrix

whose

entries

are uniformly

distributed

on the unit

interval (0,1).

- Initialize W by $W = \text{rand}(m, k)$.

- For $j = 1 : \text{maxiter}$

- Solve for H in $W^T W H = W^T A$.
- Set all negative entries of H to 0.
- Solve for W in $H H^T W^T = H A^T$.
- Set all negative entries of W to 0.

Compare it with "rand n": the standard normal distribution!

LECTURE 20

Notes: (1) Convergence is not guaranteed yet this algorithm usually works in practice.

(2) $W^T W H = W^T A$, $H H^T W^T = H A^T$
are just a bunch of normal eqn's,
e.g., $W^T W h_i = W^T a_i$, $i = 1:n$.

$H H^T W^T = H A^T$ comes from the following:

$$\begin{aligned} & \|A - WH\|_F^2 \rightarrow \min. \\ \Leftrightarrow & \|A^T - H^T W^T\|_F^2 \rightarrow \min. \\ \Leftrightarrow & ((H^T)^T H^T) W^T = (H^T)^T A^T \\ \Leftrightarrow & H H^T W^T = H A^T. \end{aligned}$$

(3) Random initialization like the original algorithm may not be efficient. We can use the following algorithm to initialize the matrix W :

- Compute the first k singular values and the corresponding vectors by
} using Arpack

$$\{ U, S, V = \text{svds}(A, \text{nsv} = k)[1];$$

- Then do the following:

$$W[:, 1] = U[:, 1];$$

for $j = 2:k$

$$C = U[:, j] * V[:, j]'; \quad \text{you can also write}$$

$$C = C .* (C >= 0); \quad \leftarrow C .* = C. >= 0$$

$$u, s, v = \text{svds}(C, \text{nsv} = 1)[1];$$

$$W[:, j] = u;$$

end

LECTURE 20

The reasoning behind this initialization is the following :

Good exercise!

If A is nonnegative, then its first singular vectors u_1 & v_1 are also nonnegative. So, it's good to use $W[:, 1] = u_1$ and $H[1, :] = v_1^T$

Unfortunately, u_2, v_2 contains negative entries due to the orthogonality $u_1 \perp u_2$, $v_1 \perp v_2$.

So, construct $C = u_2 v_2^T$, and set all the negative entries of C to 0. Then this C is nonnegative, so can compute the first singular vectors of this C , which are nonnegative and good approximations to u_2, v_2 . Then set the first left singular vector as the 2nd column of W . We can repeat this procedure until we fill W .

LECTURE 20

Example : Problem 2 of HW #1.

Consider the following set of terms (words) and documents (or rather book titles):

Terms	Documents
T1: Book (Handbook, BOOK)	D1: The Princeton Companion to Mathematics
T2: Equation (Equations)	D2: NIST Handbook of Mathematical Functions
T3: Function (Functions)	D3: Table of Integrals, Series, and Products
T4: Integral (Integrals)	D4: Linear Integral Equations
T5: Linear	D5: Proofs from THE BOOK
T6: Mathematics (Mathematical)	D6: The Book of Numbers
T7: Number (Numbers)	D7: Number Theory in Science and Communication
T8: Series	D8: Green's Functions and Boundary Value Problems
	D9: Discourse on Fourier Series
	D10: Basic Linear Partial Differential Equations
	D11: Mathematical Physics, An Advanced Course

Term-Document Matrix

$A =$

	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	D_{11}
T_1	0	1	0	0	1	1	0	0	0	0	0
T_2	0	0	0	1	0	0	0	0	0	1	0
T_3	0	1	0	0	0	0	0	1	0	0	0
T_4	0	0	1	1	0	0	0	0	0	0	0
T_5	0	0	0	1	0	0	0	0	0	1	0
T_6	1	1	0	0	0	0	0	0	0	0	1
T_7	0	0	0	0	0	1	1	0	0	0	0
T_8	0	0	1	0	0	0	0	0	1	0	0

Let's compute the NMF of A with $k=3$, using NMF.jl :

```

{ using NMF
  result = nnmf(A, 3);
  W = result.W ; H = result.H
}

```

LECTURE 20

The resulting matrices are :

$$W = \begin{bmatrix} 0.0000 & 1.2841 & 0.0000 \\ 0.9159 & 0.0000 & 0.0000 \\ 0.0000 & 0.8381 & 0.0000 \\ 0.4275 & 0.0000 & 0.8038 \\ 0.9159 & 0.0000 & 0.0000 \\ 0.0000 & 1.1292 & 0.0000 \\ 0.0000 & 0.4460 & 0.0000 \\ 0.0000 & 0.0000 & 1.1972 \end{bmatrix} \begin{matrix} T1 \\ T2 \\ T3 \\ T4 \\ T5 \\ T6 \\ T7 \\ T8 \end{matrix}$$

W_3

$$H = \begin{matrix} & D_1 & D_2 & D_3 & D_4 & D_5 & D_6 & D_7 & D_8 & D_9 & D_{10} & D_{11} \end{matrix} \begin{bmatrix} 0.0000 & 0.0000 & 0.0537 & 1.1789 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.9845 & 0.0000 \\ 0.2952 & 0.8500 & 0.0000 & 0.0000 & 0.3357 & 0.4523 & 0.1166 & 0.2191 & 0.0000 & 0.0000 & 0.2952 \\ 0.0000 & 0.0000 & 0.9534 & 0.1917 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5758 & 0.0000 & 0.0000 \end{bmatrix}$$

Let's interpret the results !

W_3 has large entries corresponding to $T4$ (Integral / Integrals) and $T8$ (Series).

The responses of the documents to W_3 is the 3rd row of H .

You can see that D_3 and D_9 have high responses, which are reasonable:

D_3 = Table of Integrals, Series, and Products

D_9 = Discourse on Fourier Series

Exercise : Do interpret W_1 and W_2 yourself !!