# Exercises on Non Negative Matrix Factorization

M-L. Cauwet

December 2, 2018

The following exercises are inspired from [1, 2, 3].

## Rank(s).

Consider $X$ a $n \times p$ non-negative matrix (i.e., every coefficient of $X$ is non-negative) with at least one positive coefficient. The **non-negative rank** of a $n \times p$ non-negative matrix $X$ is the smallest integer $k$ such that $X = AB$ where $A$ is a $n \times k$ non-negative matrix and $B$ is a $k \times p$ non-negative matrix. The non-negative rank will be denote by $rank_+(X)$. In the following, $X_{i,j}$ denotes the coefficient on the $i^{th}$ row and $j^{th}$ column of $X$. We denote by $rank(X)$ the rank of $X$.

1. Proof that $rank(X) \leq rank_+(X) \leq \min(n, p)$.
2. We say that $X_{i,j}$ and $X_{k,\ell}$ are *independent* if

   $X_{i,j} X_{k,\ell} > 0$ and $X_{i,\ell} X_{k,j} = 0$.

   Show that if $X_{i,j}$ and $X_{k,\ell}$ are independent, then $rank_+(X) \geq 2$. *Hint: write the system satisfied by the coefficients of $A$ and $B$ if $rank_+(X) = 1$.*

   More generally, we could show that, if $X$ contains a set of $q$ pairwise independent coefficients, then $rank_+(X) \geq q$. (Result admitted).
3. Consider:

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

(a) Give the rank of $X$.

(b) We can show that $X_{1,1}$, $X_{2,3}$, $X_{3,2}$ and $X_{4,4}$ are pairwise independent. Deduce the non-negative rank of $X$.

## Exact factorization: particular case of the symmetric factorization

Assume that $X$ is a $n \times n$ matrix which is:

- non-negative;
- symmetric (i.e. $X = X^T$, where $X^T$ is the transpose of $X$);
- positive semidefinite (i.e. $v^T X v \geq 0$ for all vector $v$);
- $rank(X) = 2$.

In this particular case, there exists a $n \times 2$ non-negative matrix $W$ such that $X = WW^T$. The purpose of this exercise is to build such $W$.

1. Show that $X$ has exactly 2 non-zero (and positive) eigenvalues that we will denote $\lambda_1$ and $\lambda_2$ with $\lambda_1 \geq \lambda_2$. We denote $\mathbf{v_1}$ and $\mathbf{v_2}$ their corresponding orthonormal eigenvectors (column vectors).

2. Let $D$ be the diagonal matrix $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ and $P$ the $n \times 2$ matrix $P = \begin{pmatrix} \mathbf{v_1} & \mathbf{v_2} \end{pmatrix}$. Show that $X = PDP^T$.

3. We define $V$ by $V = P\sqrt{D}$, where $\sqrt{D} = \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix}$. Show that for any rotation matrix $R$, $X = WW^T$, with $W = VR$. *We recall that a rotation matrix $R$ is defined by $RR^T = R^T R = I$ and $det(R) = 1$.*

4. $R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$. Graphically, how to choose $\theta$ in order to have $W$ non-negative?

## On the equivalence of Non-Negative Matrix Factorization and K-means

We consider $n$ points $\mathbf{x_1}, \dots, \mathbf{x_n}$ (columns vectors). In the following, $\| \cdot \|$ stands for the Euclidean norm. A K-means clustering over these $n$ points consists into finding $k$ clusters $\mathcal{C}_1, \dots, \mathcal{C}_k$ that minimize the inertia $J_K$, defined by:

$$J_K = \sum_{k=1}^{K} \sum_{i \in \mathcal{C}_k} \|\mathbf{x_i} - \mathbf{m_k}\|^2, \tag{1}$$

where $\mathbf{m_k}$ is the centroid of $\mathcal{C}_k$: $\mathbf{m_k} = \frac{1}{n_k} \sum_{i \in \mathcal{C}_k} \mathbf{x_i}$ and $n_k = Card\,(\mathcal{C}_k)$.

1. Show that Eq. 1 can be rewritten:

$$J_K = \sum_{i} \|\mathbf{x_i}\|^2 - \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i,j \in \mathcal{C}_k} \mathbf{x_i}^T \mathbf{x_j}. \tag{2}$$

2. The rescaled indicator vectors $\delta_k$ represent a given clustering as follows:

$$\begin{cases} \delta_k^i = \frac{1}{\sqrt{n_k}} & \text{if } \mathbf{x_i} \in \mathcal{C}_k \\ 0 & \text{otherwise} \end{cases}$$

   where $\delta_k^i$ is the $i^{th}$ component of column vector $\delta_k$.

   We denote $X = (\mathbf{x_1}| \dots |\mathbf{x_n})$ the $p \times n$ matrix containing the data and $\Delta = (\delta_1| \dots |\delta_K)$ the $n \times K$ matrix containing the indicator vectors.

   Show that Eq. 2 can be rewritten:

$$J_K = Tr(X^T X) - Tr(\Delta^T X^T X \Delta), \tag{3}$$

   where $Tr(A)$ is the trace of matrix $A$, i.e. the sum of the elements of its diagonal.

3. Deduce from question 2 that minimizing $J_K$ boils down to solve:

$$\max_{\Delta^T \Delta = I_K, \Delta \geq 0} Tr(\Delta^T W \Delta), \text{ where } W = X^T X. \tag{4}$$

4. Prove that Eq. 4 can be rewritten as:

$$\min_{\Delta^T\Delta = I, \Delta \geq 0} \|W - \Delta\Delta^T\|_F^2 \tag{5}$$

where $\|\cdot\|_F^2$ is the Frobenius norm, i.e. the sum of the squared coefficient of the matrix. In particular, $\|A\|_F^2 = Tr(AA^T)$.

5. Show that $\Delta\Delta^T$ is a diagonal bloc matrix of the form:

$$\Delta\Delta^T = \begin{pmatrix} \boxed{\mathbf{1}_{n_1}} & 0 & \dots & 0 \\ 0 & \boxed{\mathbf{1}_{n_2}} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \boxed{\mathbf{1}_{n_K}} \end{pmatrix}$$

where $\mathbf{1}_{n_k}$ is a $k \times k$ matrix with every coefficients equal to 1.

6. In question 4, we proved that K-means clustering is a non-negative matrix factorization with the additional constraints of:

   - symmetry (i.e. $W = AB$ with $B = A^T$);
   - orthogonality (i.e. $\Delta^T\Delta = I$).

   What is this impact on the clusters if the orthogonality constraint is relaxed?

# References


[1] Joel Cohen and Uriel G. Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. 190:149–168, 09 1993.

[2] Vassilis Kalofolias and Efstratios Gallopoulos. Computing symmetric nonnegative rank factorizations. *Linear Algebra and its Applications*, 436(EPFL-ARTICLE-198764):421–435, 2012.

[3] Chris Ding, Xiaofeng He, and Horst D Simon. On the equivalence of nonnegative matrix factorization and spectral clustering. In *Proceedings of the 2005 SIAM International Conference on Data Mining*, pages 606–610. SIAM, 2005.