

# NONNEGATIVE MATRIX FACTORIZATION

SOFÍA LLAVAYOL

ABSTRACT. This document presents the final project for the course *Numerical Linear Algebra for Statistical Learning* at Universidad de la República, Uruguay. It outlines the fundamental concepts of Nonnegative Matrix Factorization based on the reference [1], and includes selected experiments implemented in Python to illustrate key ideas.

## 1. INTRODUCTION

*Nonnegative matrix factorization (NMF)* is an easily interpretable *linear dimensionality reduction (LDR)* technique for nonnegative data. We first introduce the general concept of LDR, followed by a more detailed discussion of NMF.

**1.1. LDR techniques for Data Analysis.** Extracting the underlying structure within data sets is one of the central problems in data science, and numerous techniques exist to perform this task. One of the oldest approaches is LDR. The idea of LDR is to represent each data point as a linear combination of a small number of basis elements.

Mathematically, given a dataset of  $n$  data points  $x_1, \dots, x_n \in \mathbb{R}^m$ , LDR looks for  $r \ll \min\{n, m\}$  basis vectors  $w_1, \dots, w_r \in \mathbb{R}^m$  such that each data point  $x_j$  is well-approximated by a linear combination of these basis vectors:

$$x_j \approx w_1 \cdot h_{1j} + \dots + w_r \cdot h_{rj} = [w_1 \cdots w_r] \begin{bmatrix} h_{1j} \\ \vdots \\ h_{rj} \end{bmatrix} = Wh_j,$$

for some  $h_j = [h_{1j}, \dots, h_{rj}]^T \in \mathbb{R}^r$ .

Note that this is equivalent to *low-rank matrix approximation (LRMA)* –that is, expressing  $X \approx WH$  where

- each column of  $X \in \mathbb{R}^{m \times n}$  is a data point,  $X(:, j) = x_j$ ;
- each column of  $W \in \mathbb{R}^{m \times r}$  is a basis element,  $W(:, j) = w_j$ ;
- each column of  $H \in \mathbb{R}^{r \times n}$  contains the coordinates of a data point  $x_j$  in the basis  $W$ ,  $H(:, j) = h_j$ .

Hence LDR provides a rank- $r$  approximation  $WH$  of  $X$ , which can be written as:

$$[x_1 \cdots x_n] \approx [w_1 \cdots w_r] [h_1 \cdots h_n].$$

In order to compute  $W$  and  $H$  given  $X$  and  $r$ , one needs to define an error measure. For example, when  $(W, H)$  minimizes the Frobenius norm

$$\|X - WH\|_F^2 = \sum_{i,j} (X - WH)_{ij}^2,$$

then LRMA is equivalent to *principal component analysis (PCA)*, which can be computed via the *singular value decomposition (SVD)*.

LRMA models are used to compress the data, filter the noise, reduce the computational effort for further manipulation of the data, or to directly identify hidden structure in a data set. Many variants of LRMA have been developed, and they differ in two key aspects: (1) the error measure can vary and should be chosen depending on the noise statistic assumed on the data, (2) different constraints can be imposed on the factors  $W$  and  $H$ .

**1.2. NMF, an LDR technique for nonnegative data.** Among LRMA models, nonnegative matrix factorization requires the factor matrices  $W$  and  $H$  to be componentwise nonnegative, which we denote  $W \geq 0$  and  $H \geq 0$ . In Section 2, we discuss an application where these nonnegativity constraints are natural and meaningful.

Formally, the NMF problem is defined as follows.

**NMF Problem.** Given a nonnegative matrix  $X \in \mathbb{R}^{m \times n}$ , a factorization rank  $r$ , and a distance measure between matrices  $d(\cdot, \cdot)$ , solve

$$(1) \quad \min_{\substack{W \in \mathbb{R}^{m \times r} \\ H \in \mathbb{R}^{r \times n}}} d(X, WH) \quad \text{such that } W \geq 0 \text{ and } H \geq 0.$$

In Section 3, we discuss an algorithm to approximate this minimum for the case where the distance measure is induced by the Frobenius norm. An application of this algorithm to image processing is presented in Section 2.

## 2. APPLICATION ON FACIAL FEATURE EXTRACTION

...

## 3. ALGORITHM WITH MULTIPLICATIVE UPDATES

Let  $X \in \mathbb{R}^{m \times n}$  and  $r \ll \min\{n, m\}$  be given. In this section, we focus on the following optimization problem

$$(2) \quad \min_{\substack{W \in \mathbb{R}^{m \times r} \\ H \in \mathbb{R}^{r \times n}}} f(W, H) \quad \text{such that } W \geq 0 \text{ and } H \geq 0,$$

where  $f(W, H) = \frac{1}{2} \|X - WH\|_F^2$ .

## REFERENCES

- [1] Gillis, N. (2021). Nonnegative matrix factorization. Society for Industrial and Applied Mathematics. <https://licn.loc.gov/2020042037>