The Why and How of Nonnegative Matrix Factorization

Group 2

In a few words

NMF is a powerful tool for the analysis of **high-dimensional** data as it automatically extracts **sparse** and **meaningful** features from a set of **nonnegative** vectors. NMF was introduced by Paatero and Tapper in 1994 and further developped by Lee and Seung in 1999.

Nonnegative Matrix Factorization : definition and properties

Nonnegative matrix factorization (NMF) is a Linear dimensionality reduction (LDR). As a reminder, a LDR computes a set of $r < min(p, n)$ basis elements $w_k \in R^p$ for $1 \leq k \leq r$ to approximate a set of data points $x_j \in R^p$ for $1 \leq j \leq n$ such that $\forall j, x_j \approx \sum_{k=1}^{r} w_k h_j(k)$, for some weights $h_j \in R^r$.

LDR is equivalent to low-rank matrix approximation where: $X \approx WH$ where

- $X \in R^{p\times n} : X(:, j) = x_j$ for $1 \leq j \leq n$. Each column of the matrix X is a data point.
- $W \in R^{p\times r} : W(:, k) = w_k$ for $1 \leq k \leq r$. Each column of the matrix W is a basis element.
- $H \in R^{r\times n} : H(:, j) = h_j$ for $1 \leq j \leq n$. Each column of the matrix H gives the coordinates of a data point X(:,j) in the basis W.

NMF : decomposing a given nonnegative data matrix X as $X \approx WH$ where $W \geq 0$ and $H \geq 0$. W and H are thus component-wise nonnegative.

Applications

Image processing

Goal: facial feature extraction

The **data matrix** $X \in_{+}^{p\times n}$ carries information about n face images. Each image has q pixels and therefore each column of X represents an image of a face.

$\longrightarrow$ The (i, j)th entry of X represents the gray-level of the ith pixel in the jth face.

The **nonnegative matrix factorization** can be interpreted as follows:

- Each column of the matrix W represents a facial feature
- The (k, j)th entry of H represents the importance of the kth feature in the jth face

presentation/NMF_app1.png

Text Mining

Goal: topic recovery and document classification

The **data matrix** $X \in_{+}^{n\times n}$ encodes the frequency of some words in a list of documents.

$\longrightarrow$ The (i, j)th entry of X represents the number of times the ith word appears in the jth document.

The **nonnegative matrix factorization** can be interpreted as follows:

- Each colum of the matrix W represents a topic
- The (k, j)th entry of H represents the importance of the kth topic in the jth document

presentation/NMF_app2.png

Algorithms and Difficulties

Optimization Problem

We want to solve $\min_{W\in\mathbb{R}^{p\times r}, H\in\mathbb{R}^{r\times n}} ||X - WH||_F^2$, such that $W \geqslant 0$, $H \geqslant 0$.

Our use of the Frobenius norm implies the assumption that the noise is Gaussian. This is not always the best choice; in practice, depending on the application, other choices are possible too, such as:

- Kullback–Leibler divergence used in text mining;
- Itakura–Saito distance used in music analysis;
- ℓ_1 norm used to improve robustness against outliers;
- etc.

Difficulties

NMF is not a trivial task:

- NMF is **NP-hard**, but in practice, this is rarely problematic.
- NMF is **ill-posed**: if (W, H) is an NMF of X, then so is $(W', H') = (WQ, Q^{-1}H)$, where $WQ \geqslant 0$, $Q^{-1}H \geqslant 0$. This can be solved by
  - using **priors** for W and H, such as sparsity;
  - adding **regularization** terms.

  Finding application-specific solutions is an active area of research.

Links to other problems

TO DO

Conclusion

TO DO

References