The Why and How of Nonnegative Matrix Factorization

Topic Presentation

Group 02

LINMA2380 — Matrix computations

December 10, 2020

Summary

Applications - Image processing

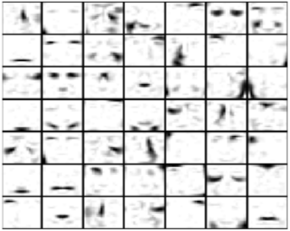
Goal : Facial Feature Extraction

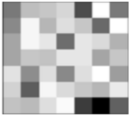
Data matrix : $X \in \mathbb{R}_+^{p \times n}$

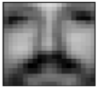
- p : total number of pixels
- n : number of faces
- $X(i,j)$: the gray-level of the i-th pixel in the j-th face

Applications - Image processing

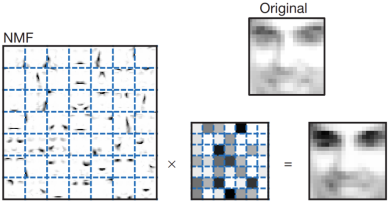
$$\underbrace{X(:,j)}_{j\text{th facial image}} \approx \sum_{k=1}^{r} \underbrace{W(:,k)}_{\text{facial features}} \underbrace{H(k,j)}_{\substack{\text{importance of features} \\ \text{in } j\text{th image}}} = \underbrace{WH(:,j)}_{\substack{\text{approximation} \\ \text{of } j\text{th image}}}$$



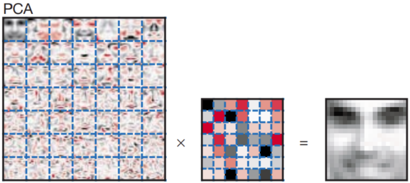




Applications - Image processing



NMF decomposition



PCA decompostion

Applications - Text mining

Goal : Topic Recovery and Document Classification

Data matrix : $X \in \mathbb{R}^{n \times m}$

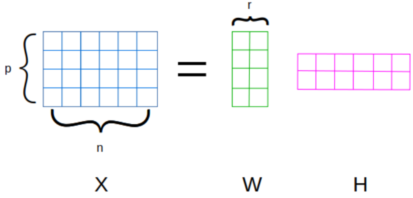
- each column : a document
- each line : a word
- $X(i,j)$: number of times the i-th word appears in the j-th document

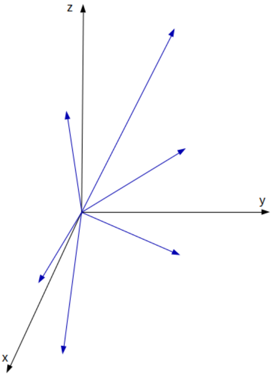
$$\underbrace{X(:,j)}_{j\text{th document}} \approx \sum_{k=1}^{r} \underbrace{W(:,k)}_{k\text{th topic}} \underbrace{H(k,j)}_{\substack{\text{importance of } k\text{th topic} \\ \text{in } j\text{th document}}}, \quad \text{with } W \geq 0 \text{ and } H \geq 0.$$

Nonnegative rank

Definition (Nonnegative rank)

Given $X \in \mathbb{R}_+^{p \times n}$, the nonnegative rank of X, denoted rank$_+(X)$ is the minimum r s.t. $\exists W \in \mathbb{R}_+^{p \times r}, H \in \mathbb{R}_+^{r \times n}$ with $X = WH$.





Computational Geometry : Nested polytopes problem

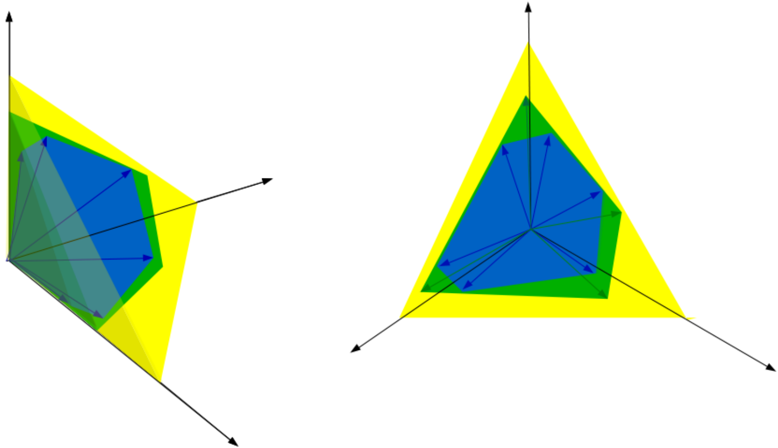


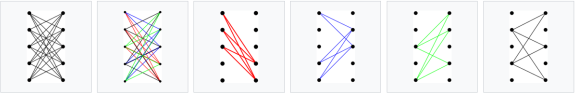
Figure: Finding a polytope with minimum nb of vertices nested between 2 polytopes

Graph Theory : Bipartite dimension

Let $G(X) = (V_1 \cup V_2, E)$ be a bipartite graph induced by X (i.e. $(i,j) \in E \Leftrightarrow X_{ij} \neq 0$).

Definition (Biclique and bipartite dimension)

- *A biclique (or a complete bipartite graph) is a bipartite graph s.t. every vertex in V_1 is connected to every vertex in V_2.*
- *The bipartite dimension (or the minimum biclique cover) bc$(G(X))$ is the minimum number of bicliques needed to cover all edges in E.*



A bipartite graph... | ...and a covering with four bicliques | the red biclique from the cover | the blue biclique from the cover | the green biclique from the cover | the black biclique from the cover

Figure: Example for biclique edge cover **[biclique]**

For any $(W, H) \geq 0$ s.t. $X = WH = \sum_{k=1}^{r} W_{:k} H_{k:}$, we have

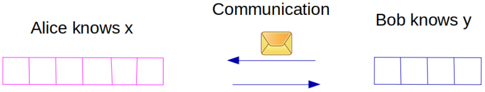
$$G(X) = \cup_{k=1}^{r} G(W_{:k} H_{k:})$$

where $G(W_{:k} H_{k:})$ are complete bipartite subgraphs (bc$(G(W_{:k} H_{k:} = 1 \forall k)$.

Theorem (Rectangle covering bound)

$$bc(G(X)) \leq rank_{+}(X)$$

Communication complexity



Alice and Bob want to compute :

$$f : \{0,1\}^m \times \{0,1\}^n \to \{0,1\} : (x,y) \to f(x,y)$$

While minimizing the nb of bits exchanged (i.e. communication complexity (CC)).
Nondeterministic comm. complex. of f (NCC): CC of f with oracle/ message before starting the communication.
The communication matrix $X \in \{0,1\}^{2^n \times 2^m}$ is equal to the function f for all possible combinations of inputs.

Theorem (Yannakis)

$$\textit{NCC of } f \leq \log_2(\textit{rank}_+(X))$$