Keunsu Kim

Topology-Aware Dimensionality Reduction

Institute of Mathematics for Industry
Kyushu University

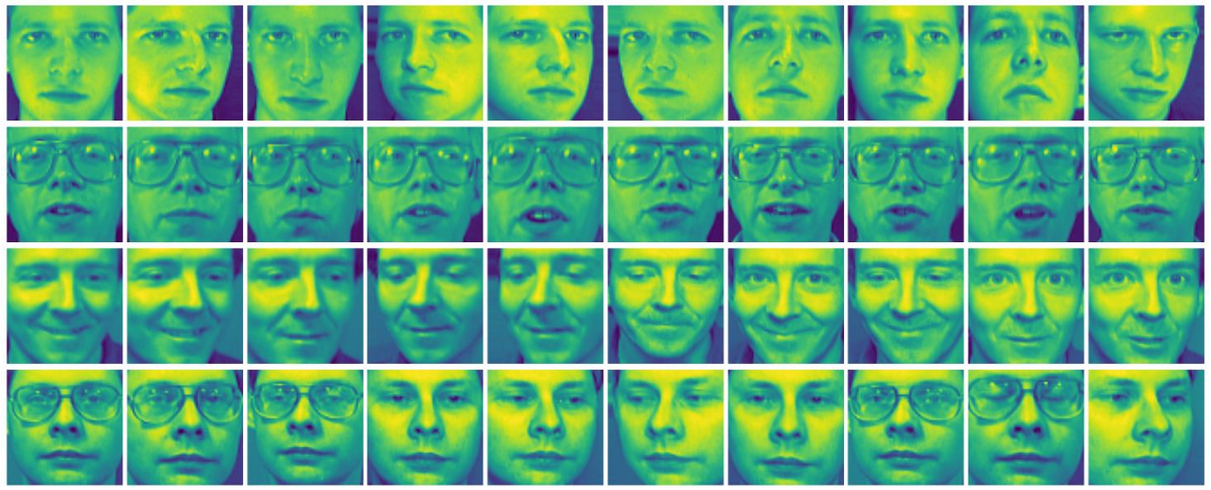
Background and Motivation

Extracting meaningful building blocks from data is crucial for reliable AI.

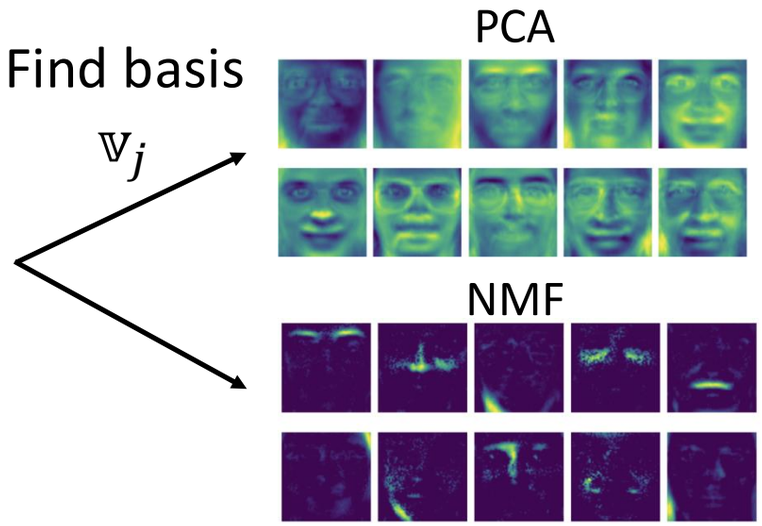
Given dataset $\{\mathbb{x}_1, \cdots, \mathbb{x}_n\}$ and $\mathbb{x}_i \in \mathbb{R}^d_{\geq 0}$,

Nonnegative Matrix Factorization (NMF): $\mathbb{x}_i \approx w_{i1}\mathbb{v}_1 + \cdots + w_{ir}\mathbb{v}_{\boldsymbol{r}}$ **,**$w_{ij}, \mathbb{v}_j \geq 0$

This additive nature ensures that basis vectors $\mathbb{v}_j$ (building blocks) do not cancel each other out.



Fetch olivetti faces dataset



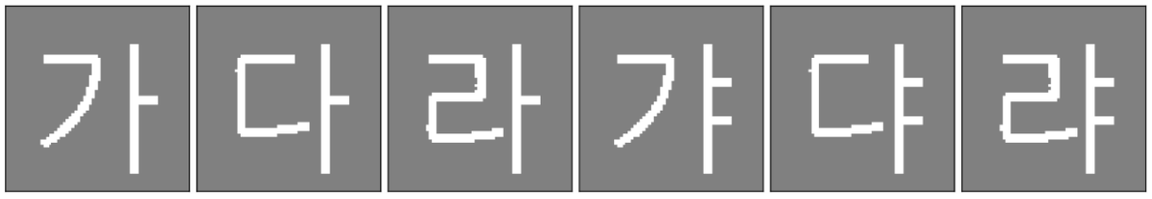
More interpretable!

Current Situation

- Nonnegative Matrix Factorization (NMF) provides more interpretable basis vectors than Principle Component Analysis (PCA) because it yields part-based representations.
- However, standard NMF lacks a direct mechanism to enforce these part-based structures.

Development

Topological Regularization: Enforces structural properties (e.g., connectivity) on basis vectors.



Ex: Korean character

Part-based representation