machine learning in climate science

AN OVERVIEW OF REDUCED REPRESENTATIONS

(AKA DIMENSIONALITY REDUCTION TECHNIQUES)

Bedartha Goswami

Journal Club

23 Feb 2021

What & Why

1

- What are reduced representations (RR)?
- Why do we need RR?
- Existing methods

PCA and Co

2

- PCA
- LLE
- LEM
- MDS
- Isomap
- kPCA

Others

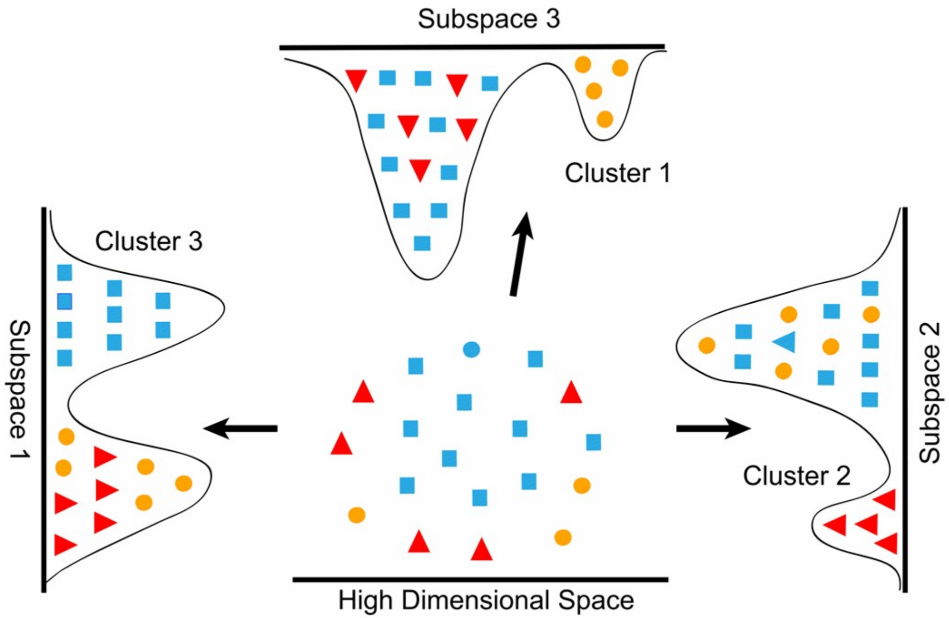
3

- Graph Clustering
- VAE
- SOM
- NMF

Reduced representations

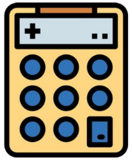
- Represent a smaller set of 'essential' features 'hidden' in high-dim data
- Parametric or non-parametric
- Linear (PCA) or nonlinear (kPCA)
- Convex (LLE) or non-convex (VAE)
- Sparse (graphs) or non-sparse (NMF)

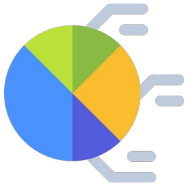
https://datascience.stackexchange.com/a/23087



Reduced representations are useful as

- They reduce computational complexity
- They reduce informational complexity
  → increase interpretability
- They help in visualising and
  understanding essential features
- They help remove ‘noise’ or
  unnecessary components of data
- They help in predictions





A few important methods –

- Principal Component Analysis (PCA)
- Local Linear Embedding (LLE)
- Laplacian Eigenmap (LEM)
- Metric Multidimensional Scaling (MDS)
- Isomap
- Kernel PCA
- Graph clustering
- Self-organising Map (SOM)
- Variational Autoencoder (VAE)
- Non-negative Matrix Factorisation (NMF)
- …

What & Why

1

- What are reduced representations (RR)?
- Why do we need RR?
- Existing methods

PCA and Co

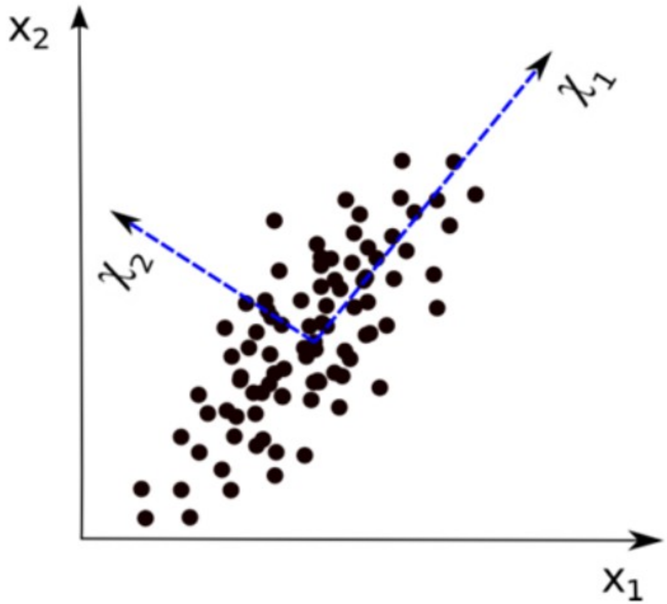
2

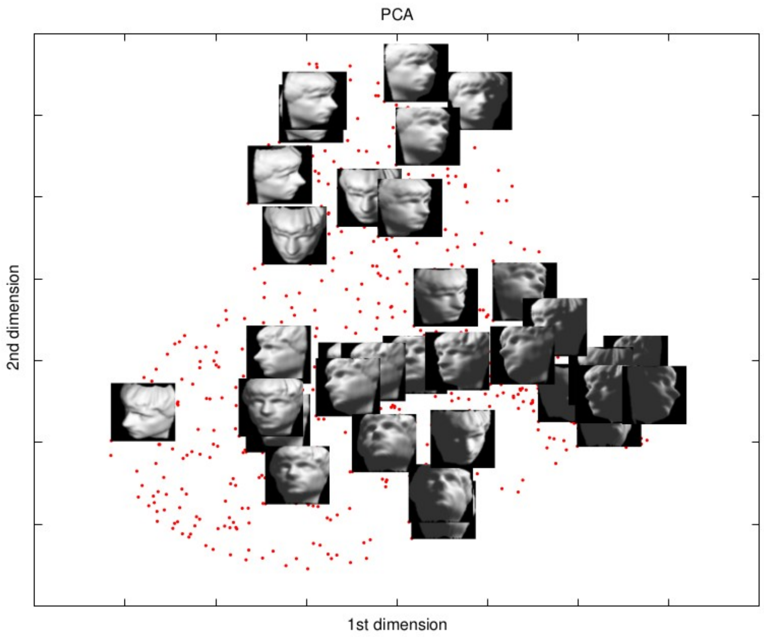
- PCA
- LLE
- LEM
- MDS
- Isomap
- kPCA

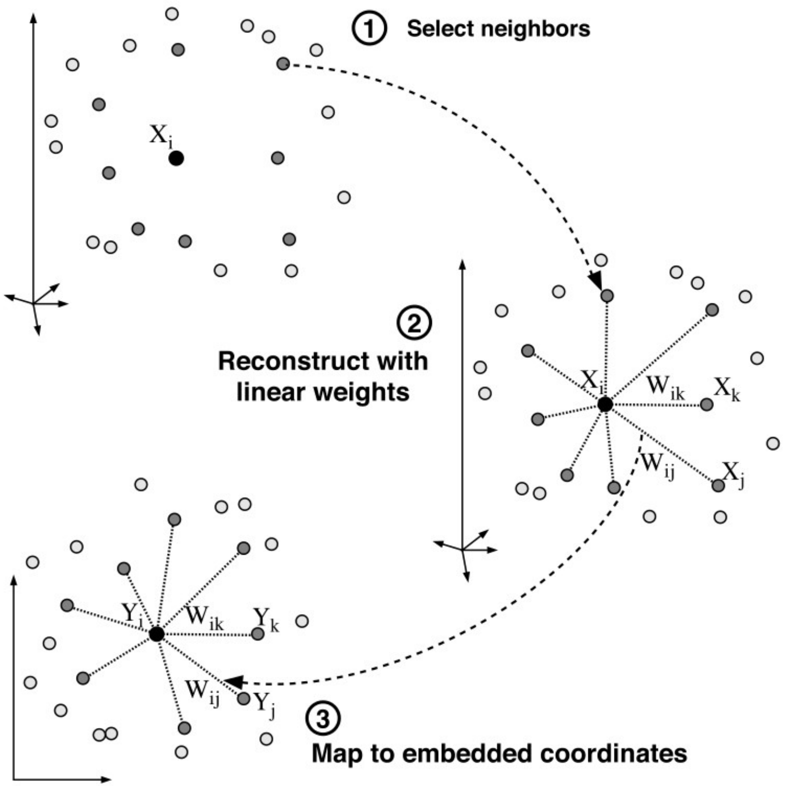
Others

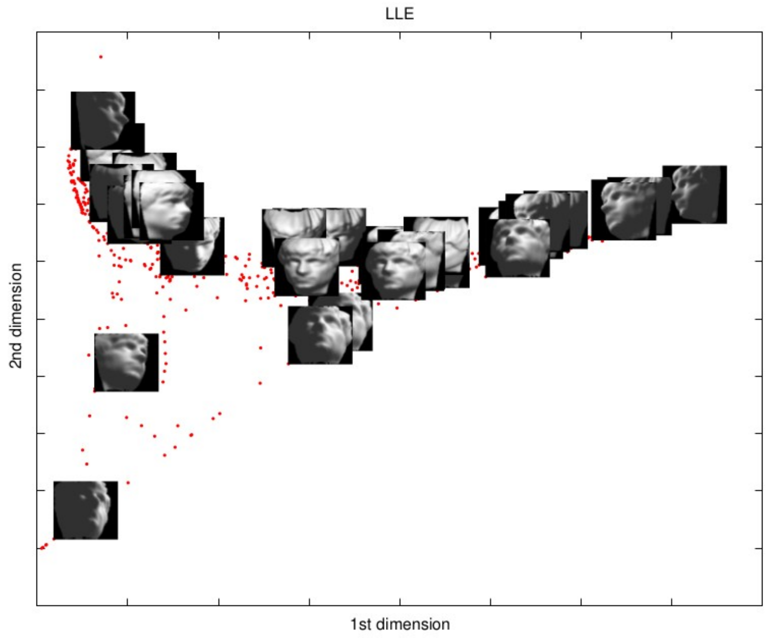
3

- Graph Clustering
- VAE
- SOM
- NMF


x_2
χ_1
χ_2
x_1


PCA
2nd dimension
1st dimension



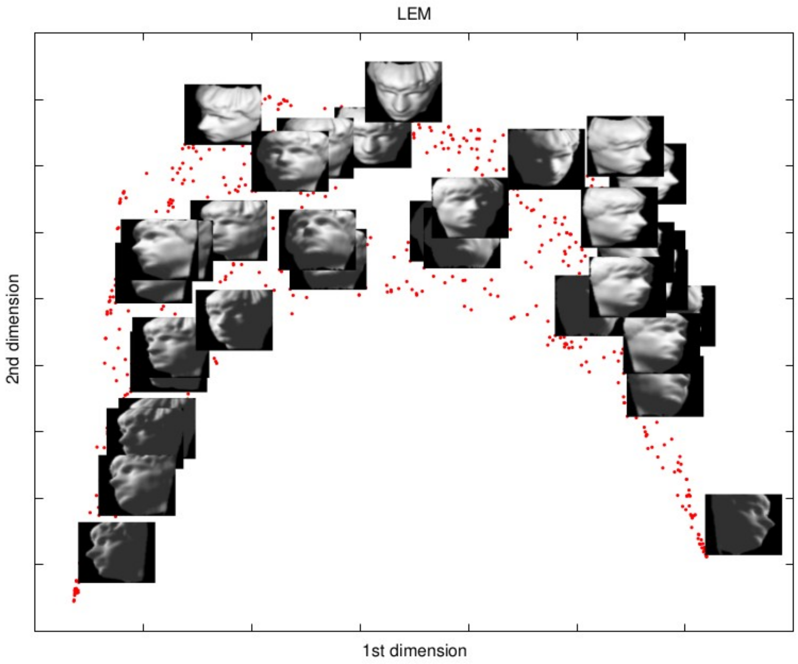


2. PCA and Co → LLE

$$\min_{Y} \sum_{i=1}^{t} \sum_{j=1}^{t} (\mathbf{y}_i - \mathbf{y}_j)^2 W_{ij}$$

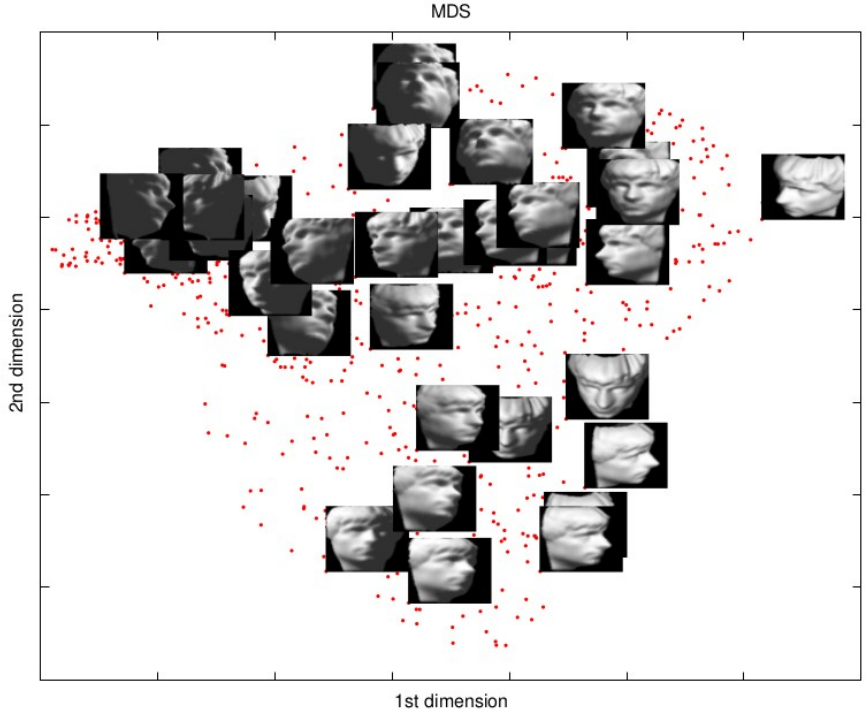
$$\Updownarrow$$

$$\min_{Y} \mathrm{Tr}(Y L Y^T)$$

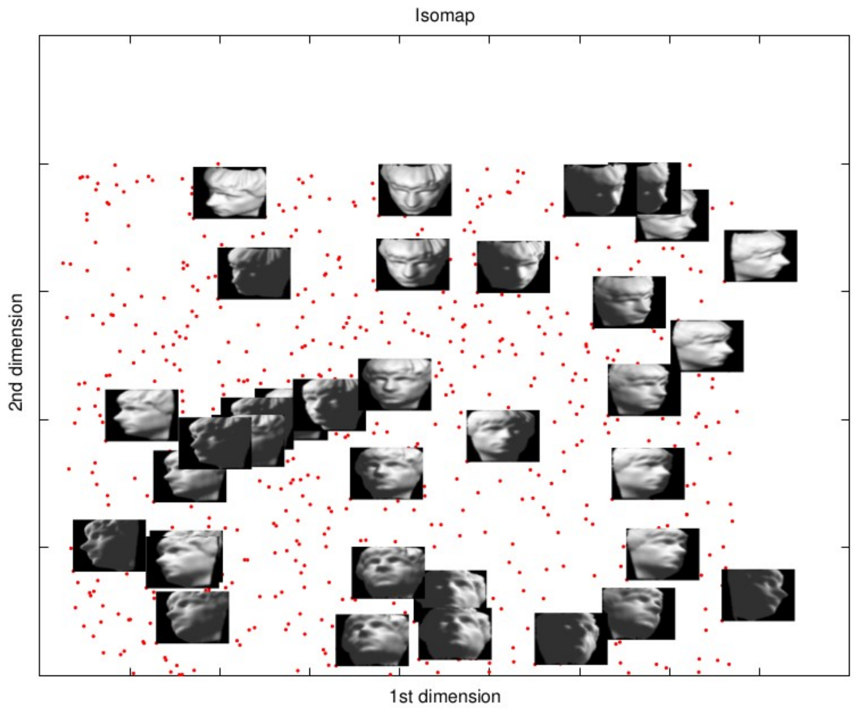


- Estimate pairwise distances from high-minsional data (e.g. for time series, it can be correlation-based distance)

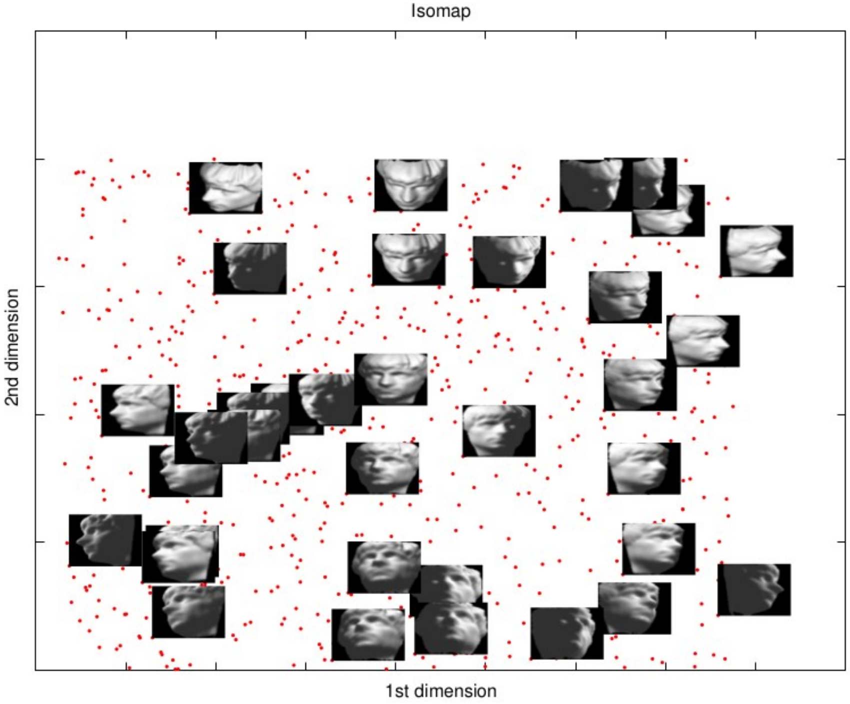
- Find low-dimensional points (after choosing a lower dimension 'd') that preserve the pairwise distances



- Find the neighbours of each data point in high-dimensional data space.
- Compute the geodesic pairwise distances between all points.
- Embed the data via MDS so as to preserve these distances.



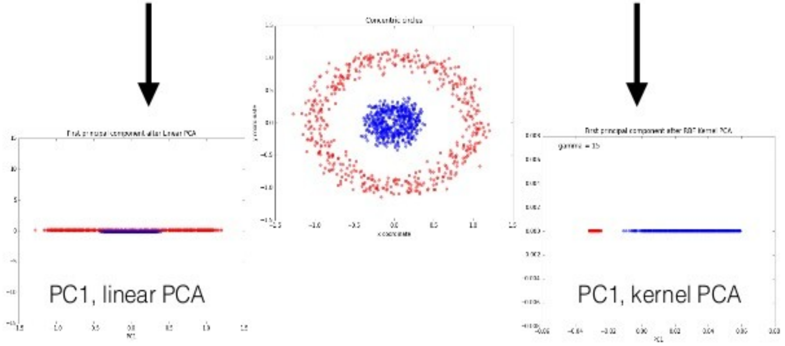
- Find the neighbours of each data point in high-dimensional data space.
- Compute the geodesic pairwise distances between all points.
- Embed the data via MDS so as to preserve these distances.

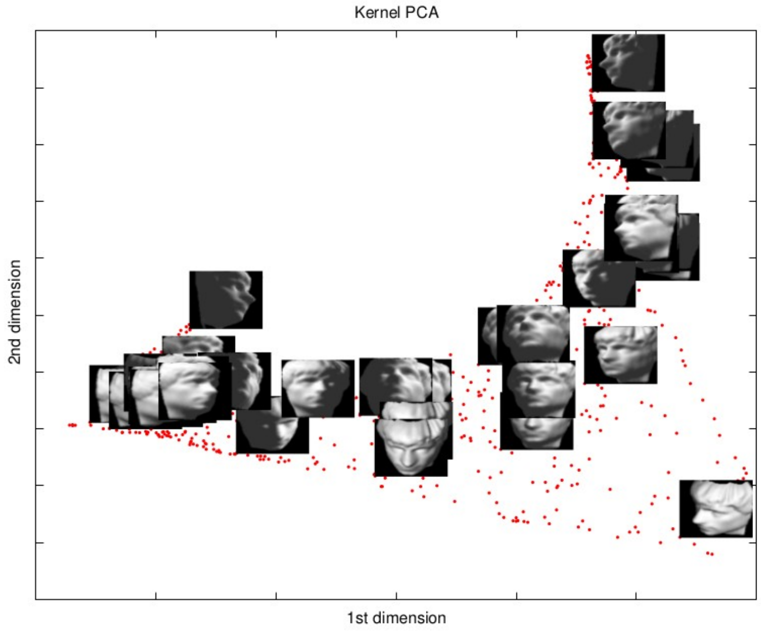


Bedartha Goswami

$$\mathrm{Cov} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i^T \boldsymbol{x}_i$$

$$\mathrm{Cov} = \frac{1}{N}\sum_{i=1}^{N} \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_i)$$





2. PCA and Co → kPCA

$$K_{LLE} = \lambda_{max} I - L$$

$$L = (I - W)^T (I - W)$$

$$K_{MDS} = -\frac{1}{2}(I - ee^T) D (I - ee^T)$$

e is a column vector of all ones

distance matrix D

$$K_{LEM} = L^{\dagger}$$

$$L = R - W$$

R is diagonal, and $R_{ii} = \sum_{j=1}^{t} W_{ij}$

$$K_{Isomap} = -\frac{1}{2}(I - ee^T) D^{(\mathcal{G})} (I - ee^T)$$

e is a column vector of all ones

geodesic distance $D^{(\mathcal{G})}$

What & Why

1
- What are reduced representations (RR)?
- Why do we need RR?
- Existing methods

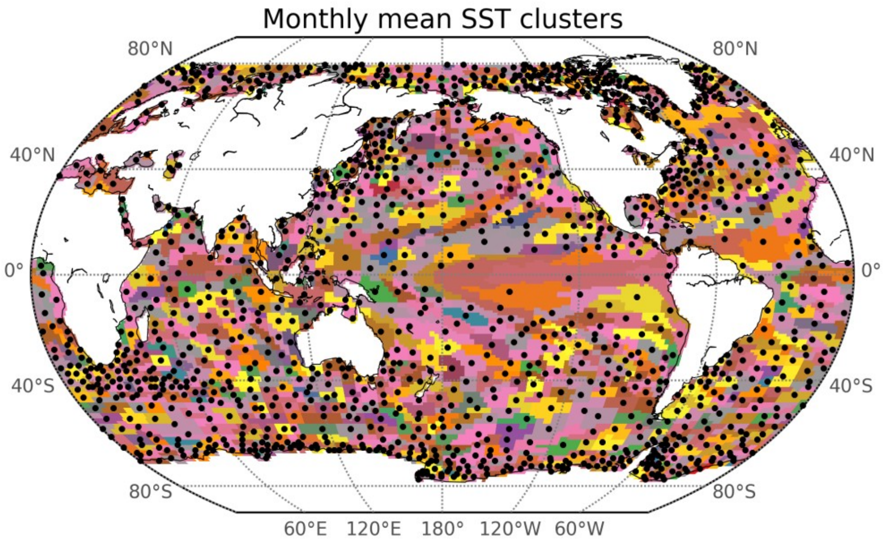
PCA and Co

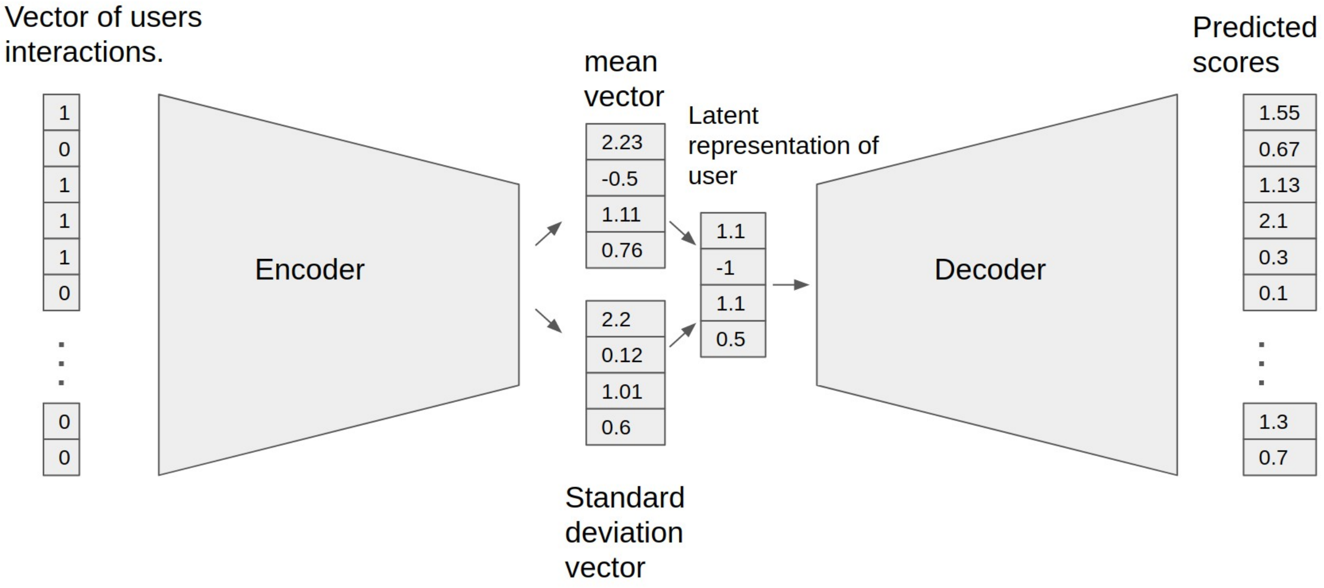
2
- PCA
- LLE
- LEM
- MDS
- Isomap
- kPCA

Others

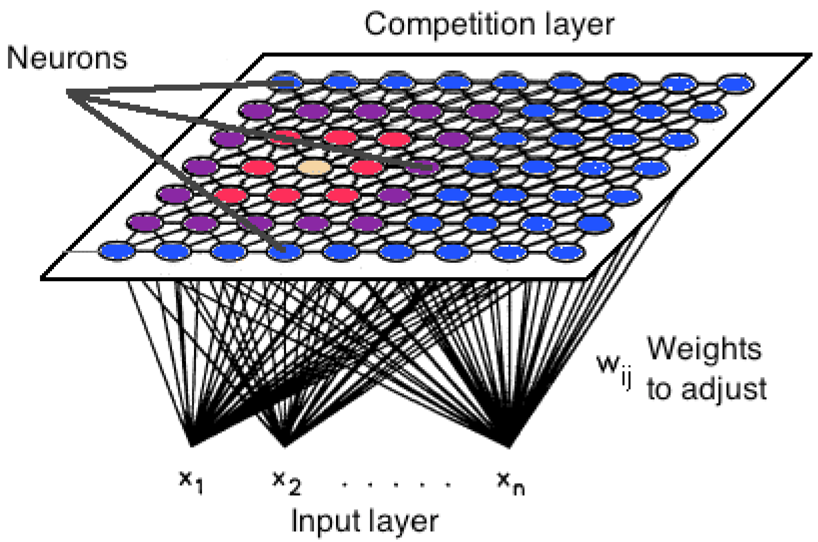
3
- Graph Clustering
- VAE
- SOM
- NMF

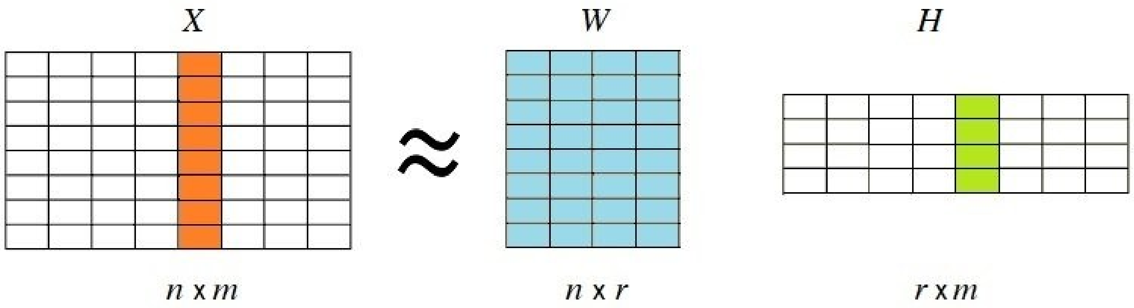
- Construct a correlation based graph from time series
- Use a clustering algorithm or perform community detection
- K-means clustering, hierarchical clustering
- Stochastic block model, modularity optimization

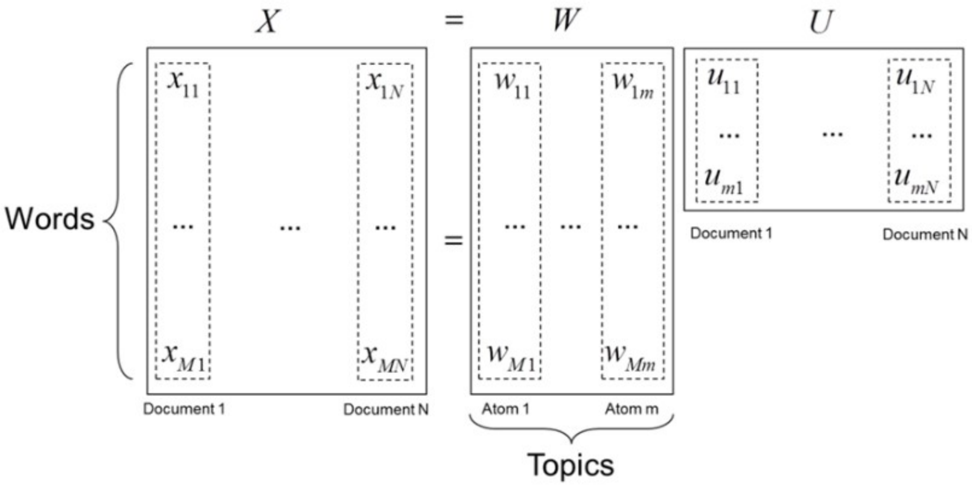




3.Others → Variational Autoencoders


Competition layer
Neurons
w_{ij}
Weights
to adjust
x_1
x_2
. . . . .
x_n
Input layer





3.Others → Non-negative Matrix Factorisation