Machine Learning

Dimension Reduction

Jan Platoš

November 22, 2023

Department of Computer Science
Faculty of Electrical Engineering and Computer Science
VŠB - Technical University of Ostrava

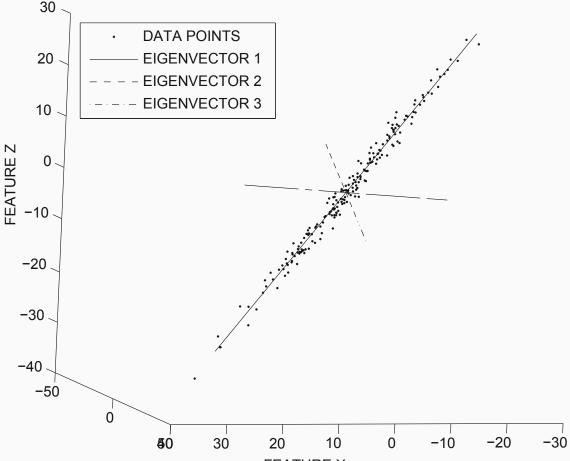
Dimension Reduction

Dimension Reduction

- Highly dimension data brings problems with clustering/classification.
- Many features are noisy or noise itself.
- Many features correlates with another features.
- Feature selection:
  - Select features according a measure and removes is from the dataset.
  - Measure is based on a mathematical principle (Variance, Entropy, etc.)
- Dimension Reduction:
  - Search for optimal mapping between original dimension into defined amount of dimensions.
  - Each new dimension is a linear/non-linear combination of original features.

Principal Component Analysis (PCA)

- The goal of PCA is to rotate the data into an axis-system where the greatest amount of variance is captured in a small number of dimensions.



Principal Component Analysis (PCA)

- The PCA for the input matrix D is computed as:

$$C = \frac{D^T D}{n} - \overline{\mu}^T \overline{\mu}$$

- C is a covariance matrix of D, n is the number of points of the D, μ is the mean vector.

$$C = P \Delta P^T$$

- P contains orthonormal eigenvectors and Δ contain eigenvalues.

$$D' = DP$$

- D' is transformed matrix in the terms of new axis P.

Singular Value Decomposition (SVD)

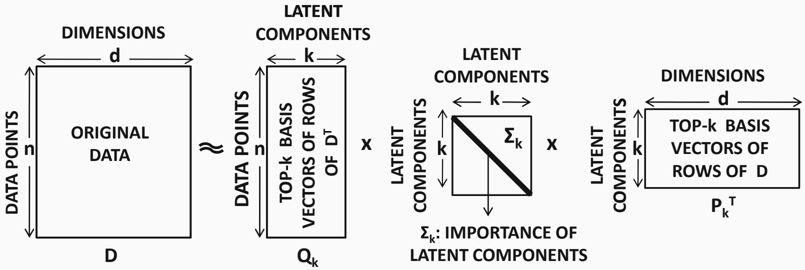
- Generalization of the PCA.

$$D = U\Sigma V^T$$

- where U contains left singular vectors, Σ contains singular values and V^T contains right singular vectors.
- The presented decomposition is proven to be optimal.
- Reducing the Σ to k coefficients leads to best approximation of the matrix D

$$D \approx U_k \Sigma_k V_k^T$$

Singular Value Decomposition



Non-negative Matrix Factorization (NMF or NNMF)

- A factorization methods which works and produces only non-negative elements.

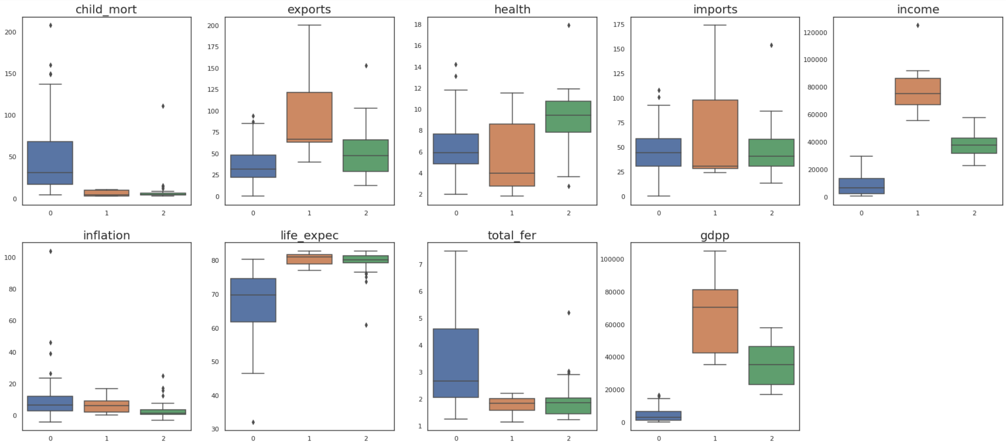
$$D = WH$$

- W contains weights and H contains basis vectors.
- Due to non-negativity the basis vectors as well as weights may be easily interpreted.
- The NMF inherits clustering property, where close vectors are clustered together.
- The cost function is defined usually as a Frobenius norm:

$$E = \|D - WH\|_F$$

$$\|A\|_F = \sqrt{\sum_i \sum_j |a_{ij}|^2}$$

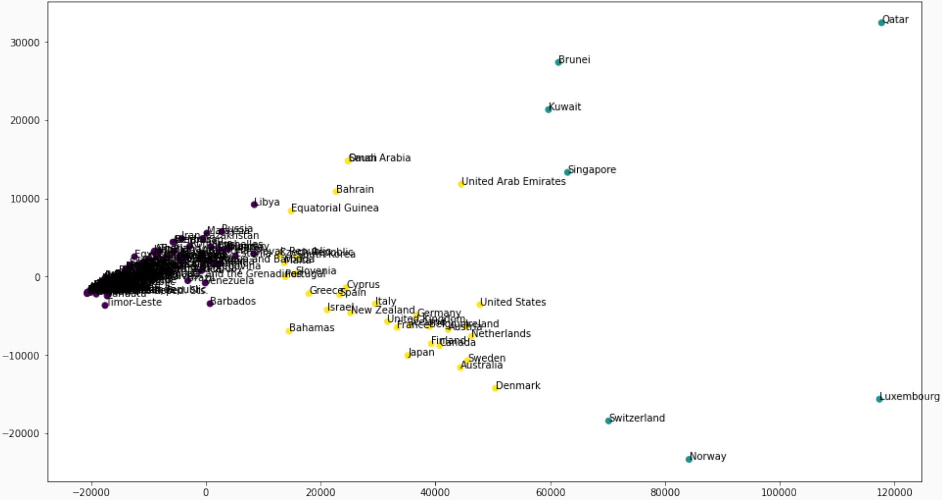
Dimension Reduction - Example



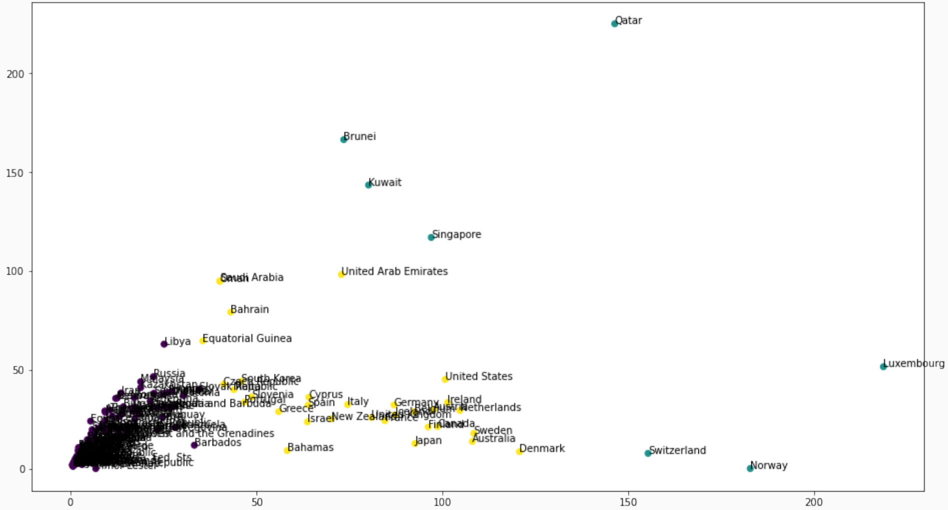
Dimension Reduction - Example

Cluster 0	Afghanistan, Albania, Algeria, Angola, Antigua and Barbuda, Argentina, Armenia, Azerbaijan, Bangladesh, Barbados, Belarus, Belize, Benin, Bhutan, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Cape Verde, Central African Republic, Chad, Chile, China, Colombia, Comoros, Congo, Dem. Rep., Congo, Rep., ...
Cluster 1	Brunei, Kuwait, Luxembourg, Norway, Qatar, Singapore, Switzerland
Cluster 2	Australia, Austria, Bahamas, Bahrain, Belgium, Canada, Cyprus, Czech Republic, Denmark, Equatorial Guinea, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Japan, Malta, Netherlands, New Zealand, Oman, Portugal, Saudi Arabia, Slovenia, South Korea, Spain, Sweden, United Arab Emirates, United Kingdom, United States

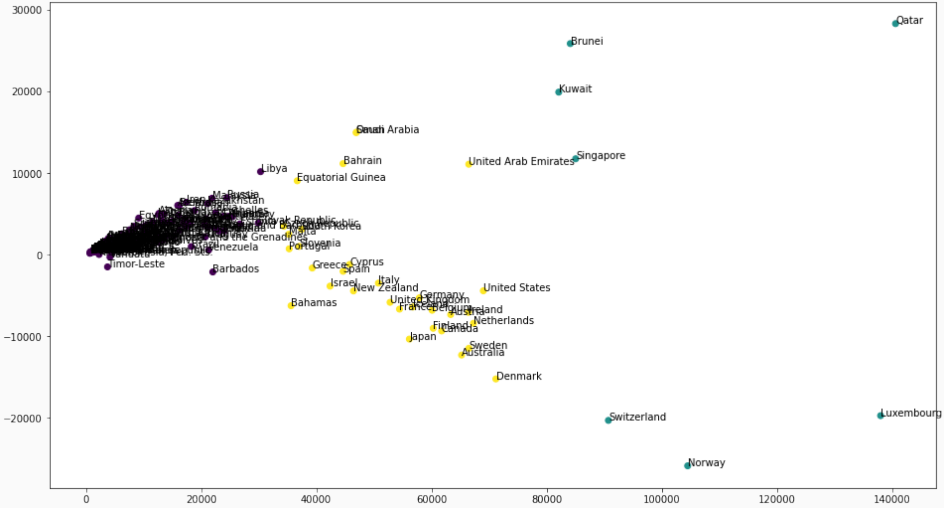
Dimension Reduction - Example - PCA



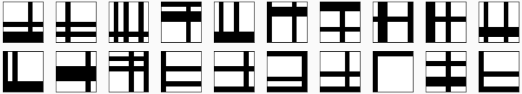
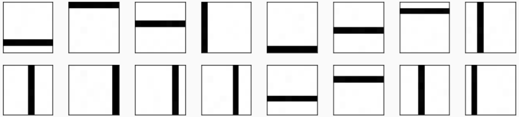
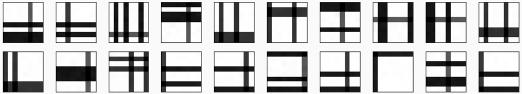
Dimension Reduction - Example - NMF



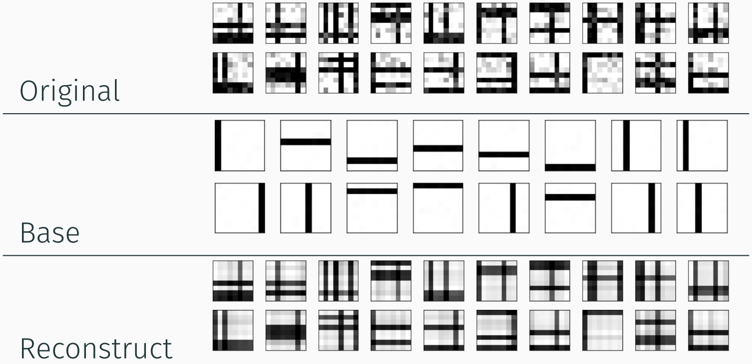
Dimension Reduction - Example - SVD



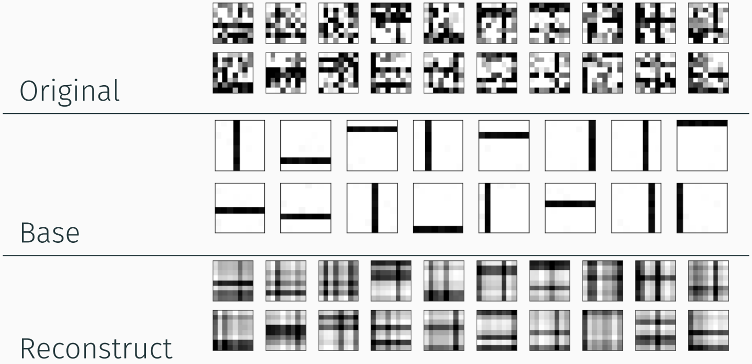
Dimension Reduction - Example - NMF - Bars

Original	
Base	
W[0]	3,0,0,0,4,3,0,0,0,0,0,3,0,0,0,0
H[0]	0,1,0,0,0,0,1,0,...,88,88,87,87,87,87,87,87,0,0,0,0,0,0,0,1
Reconstruct	

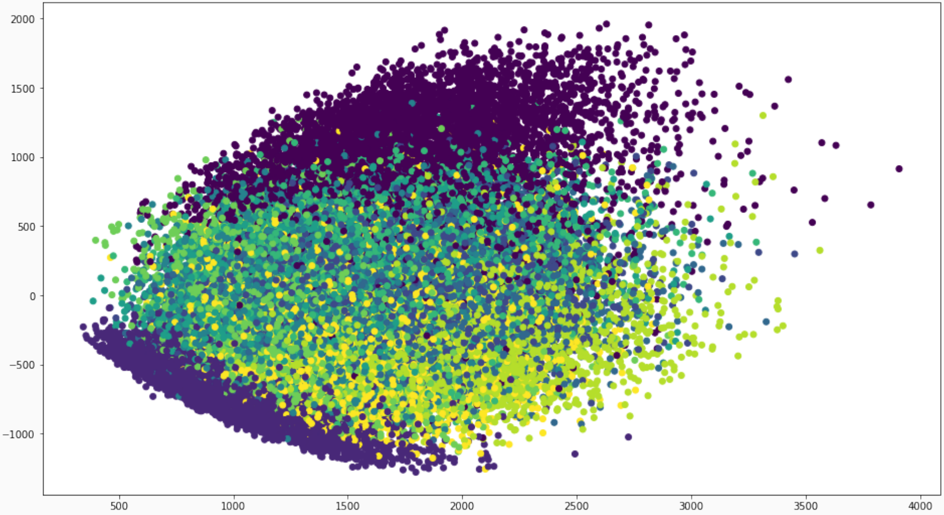
Dimension Reduction - Example - NMF - Bars



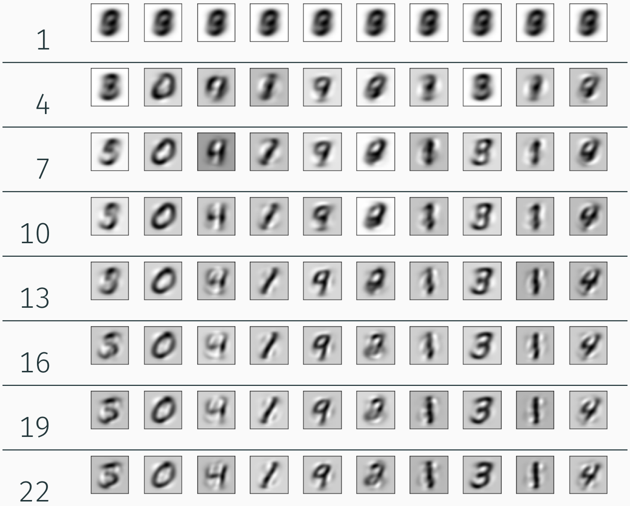
Dimension Reduction - Example - NMF - Bars



Dimension Reduction - Example - SVD - Mnist



Dimension Reduction - Example - SVD - Mnist



Questions?