STEVENS INSTITUTE of TECHNOLOGY
School of Business

Multivariate Data Analytics

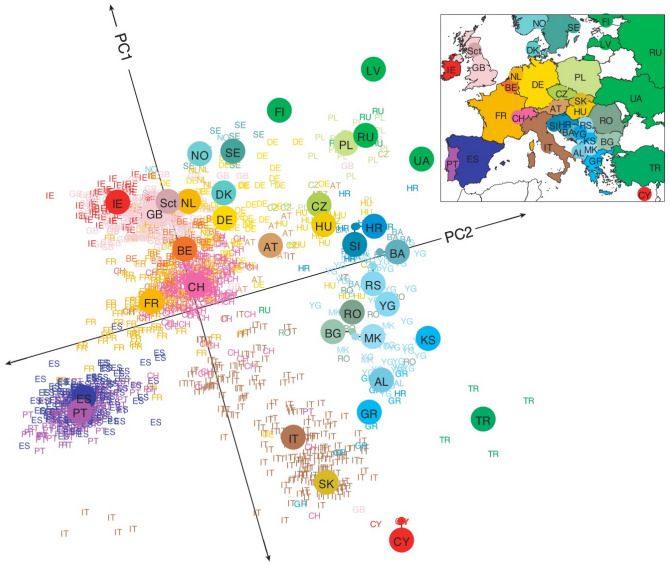
Dimension Reduction

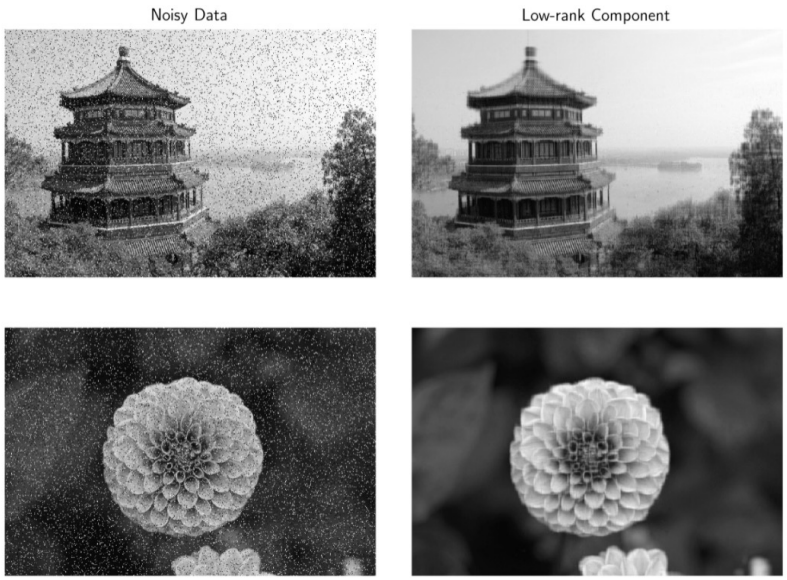
Prof. Feng Mai
School of Business

For academic use only.

Motivation

- Unsupervised learning techniques
- Extract hidden (lower dimensional) structure from high dimensional datasets
- Useful for
  - Visualization: Project high-dimensional data to 2-D
  - Noise removal
  - Pre-process data for building better statistical models, fewer dimensions → better generalization

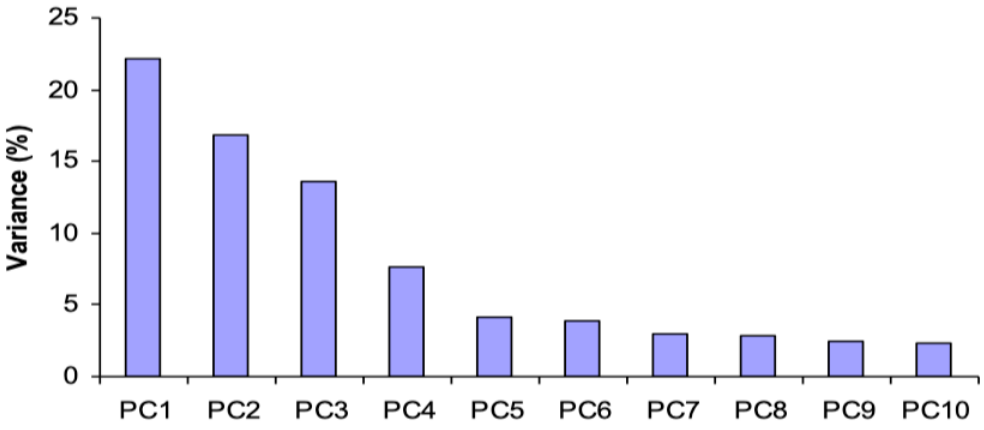




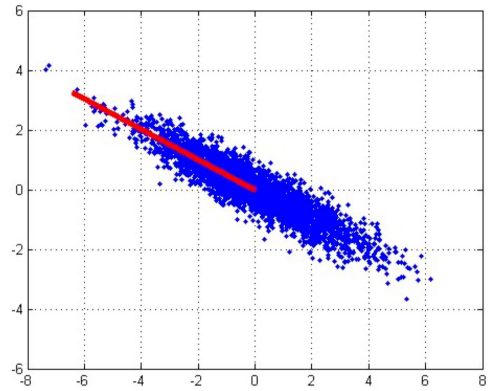
Source: Novembre et al. (2008). Genes mirror geography within Europe. *Nature*, *456*(7218), 98-101
https://tangbinh.github.io/

Principal Component Analysis (PCA)

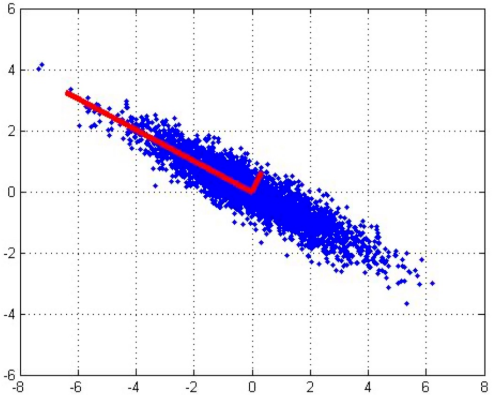
- Reduces the dimension of a dataset
- Project data into a (possibly lower dimensional) subspace so that the variance of the projected data is maximized.
- Can be computed by performing Singular Value Decomposition (SVD) on centered (demeaned) data matrix
- If the original data has *k* dimensions, we can find *k* PCs. The 1st PC explains the most variance, followed by the 2nd PC...How many should PCs should we retain?



1st PCA axis



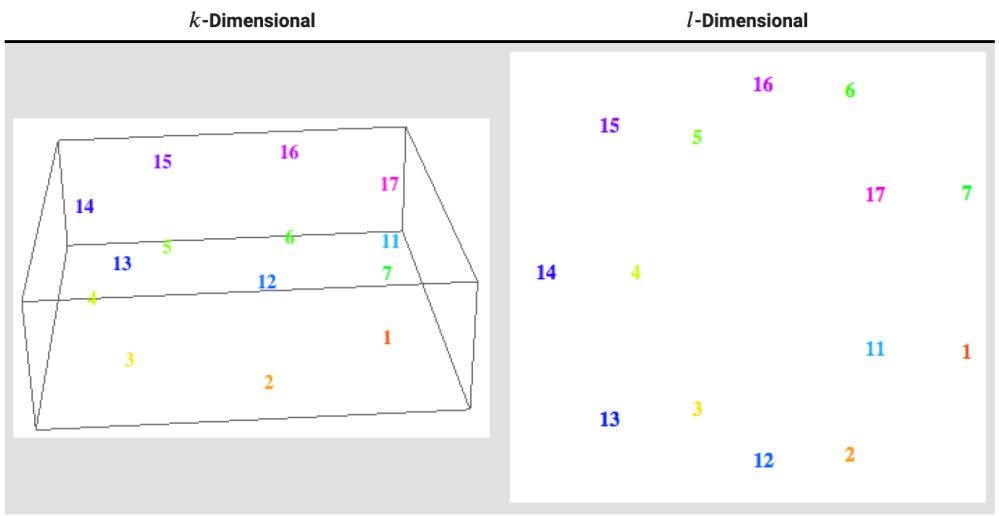
2nd PCA axis

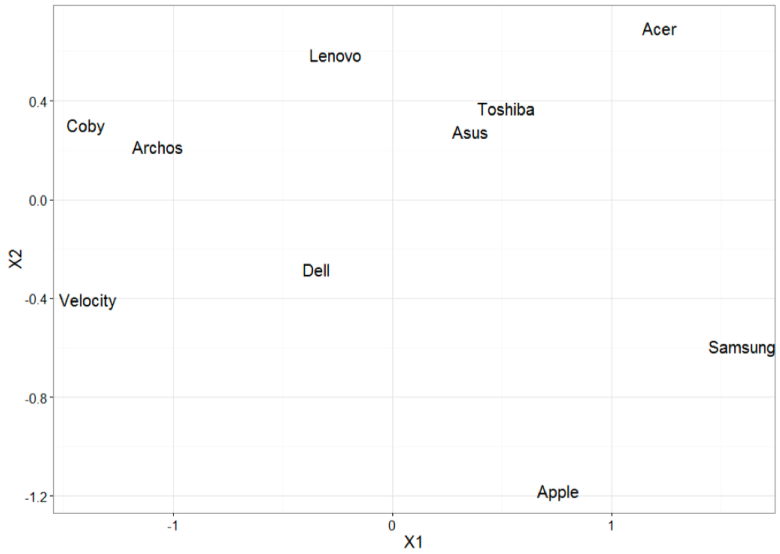


Source: Exmples from Nina Balcan and Eric Xing

Multidimensional Scaling (MDS)

- In PCA, the projection is linear, and is done to maximize the variation preserved.
- There is no guarantee that two data points (two rows of X) that are far away in the k dimensional space get projected to be very close
- Given pairwise dissimilarities at high dimensional space, MDS reconstructs a low dimensional map that preserves distances.
- Let $d_{ij}=\|x_i-x_j\|$ be the distance between points i and j in k-D. Let point i, j be projected to y_i, y_j in l-D, $l < k$
- MDS minimizes $\text{stress} = \sum_{i \neq j} w_{ij} \left(\|y_i - y_j\| - d_{ij} \right)^2$





Source: Mai, F. (2015). *Essays in Business Analytics*. University of Cincinnati.

t-distributed stochastic neighbor embedding (t-SNE)

- Similar to MDS, t-SNE preserves pairwise similarity at low dimensional space.
- Points i and j are similar → The conditional probability of i being j's neighbor is high
- Before projection

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|^2 / 2\sigma_i^2\right)},$$

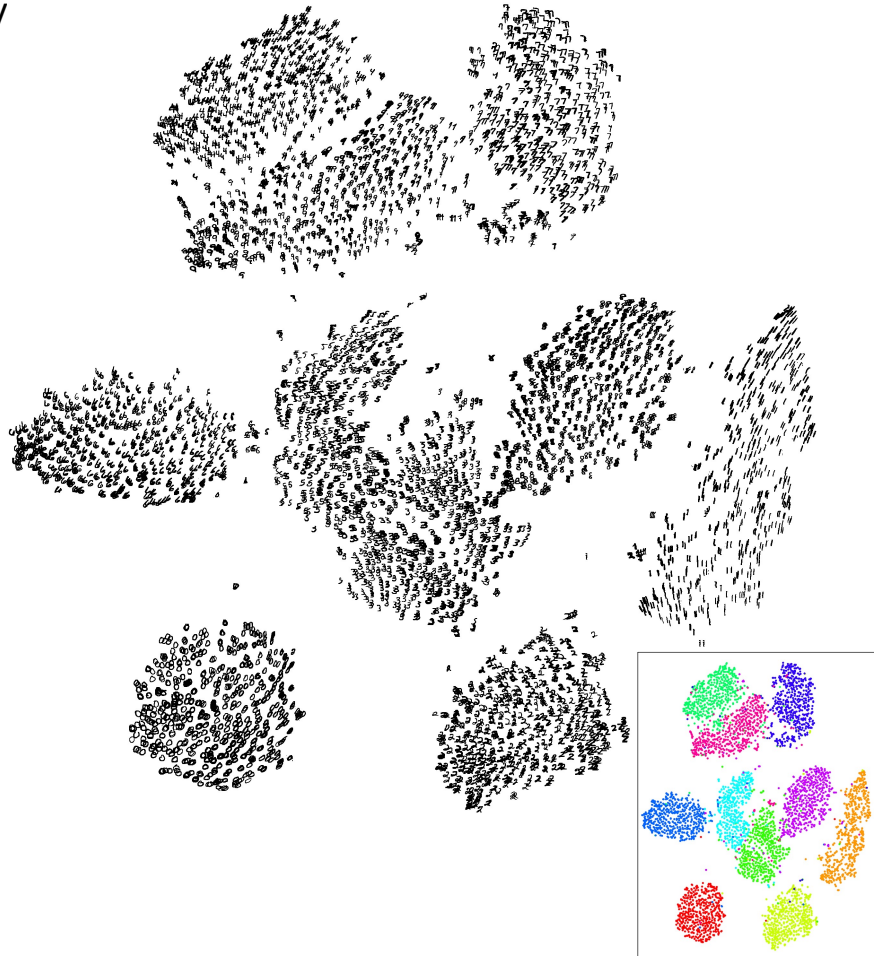
- After projection

$$q_{j|i} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k \neq i} \exp\left(-\|y_i - y_k\|^2\right)}.$$

- Objective

$$C = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}},$$

- Scalable to millions of points



Source:.

STEVENS
INSTITUTE of TECHNOLOGY
School of Business

Thank you!

Prof. Feng Mai
School of Business