Unsupervised Machine Learning

Dimensionality Reduction

Mikhail Evchenko

Rob Romijnders

About us

Mike

Rob

Why Unsupervised Learning?

- You don’t have the information about data.
- Labels for data are missing.

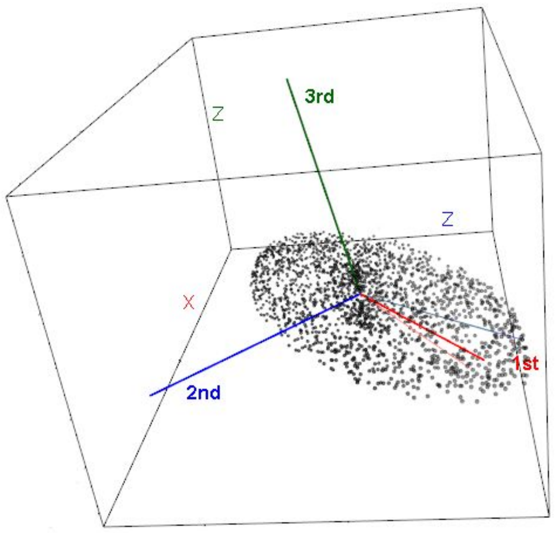
Why Dimensionality Reduction?

- You have multi-dimensional data.
- Visual representation of data can provide insight into the structure and properties of data.

PCA

Principal Component Analysis

PCA applied to an ellipsoidically shaped point cloud



more information: www.joyofdata.de/blog/illustration-of-principal-component-analysis-pca

Principal Components Analysis

- Principal component analysis (PCA) is a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components.
- The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

- Principal Component Analysis (PCA) is a dimension-reduction tool that can be used to reduce a large set of variables to a small set that still contains most of the information in the large set.

More information is availabe at ftp://statgen.ncsu.edu/pub/thorne/molevoclass/AtchleyOct19.pdf

What is explained variance?

Original data
with 3 dimensions.

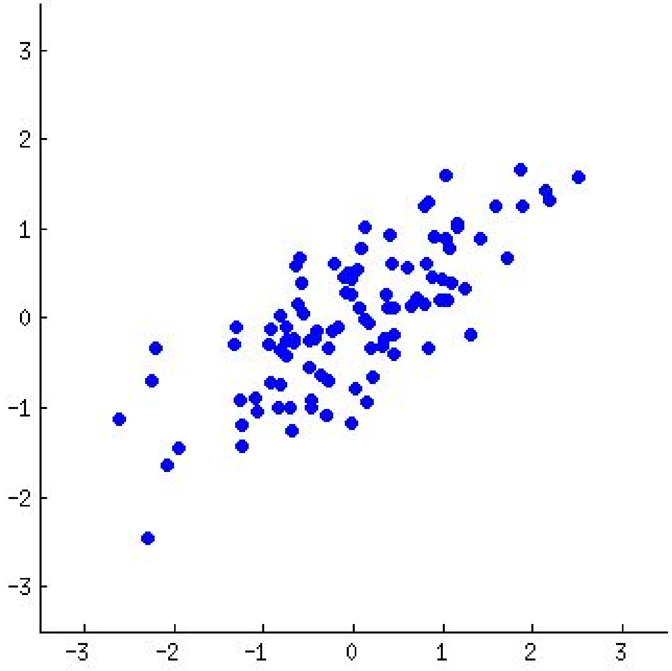
Data with two main PC that hold
the most of the variance in data.

PCA in details

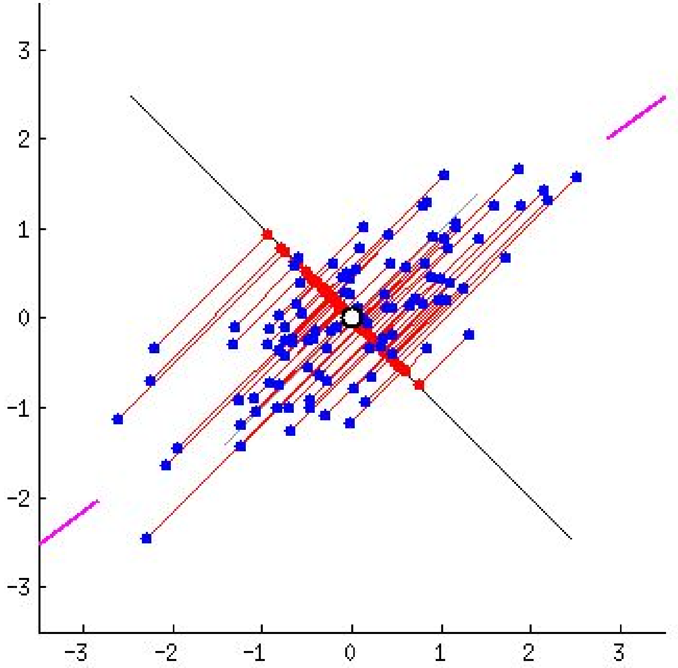
1. Take the whole dataset consisting of d-dimensional samples ignoring the class labels.
2. Compute the d-dimensional mean vector (i.e., the means for every dimension of the whole dataset).
3. Compute the scatter matrix (alternatively, the covariance matrix) of the whole data set.
4. Compute eigenvectors (e1,e2,...,ed) and corresponding eigenvalues (λ1,λ2,...,λd).
5. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a d×k dimensional matrix W (where every column represents an eigenvector).
6. Use this d×k eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the mathematical equation: $y = W^T * x$ (where x is a d×1-dimensional vector representing one sample, and y is the transformed k×1-dimensional sample in the new subspace).

http://sebastianraschka.com/Articles/2014_pca_step_by_step.html

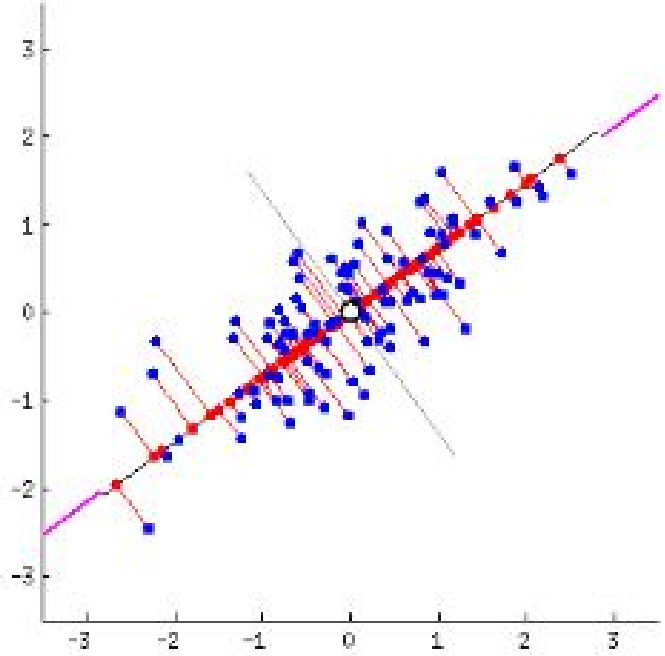
PCA visualization



PCA visualization

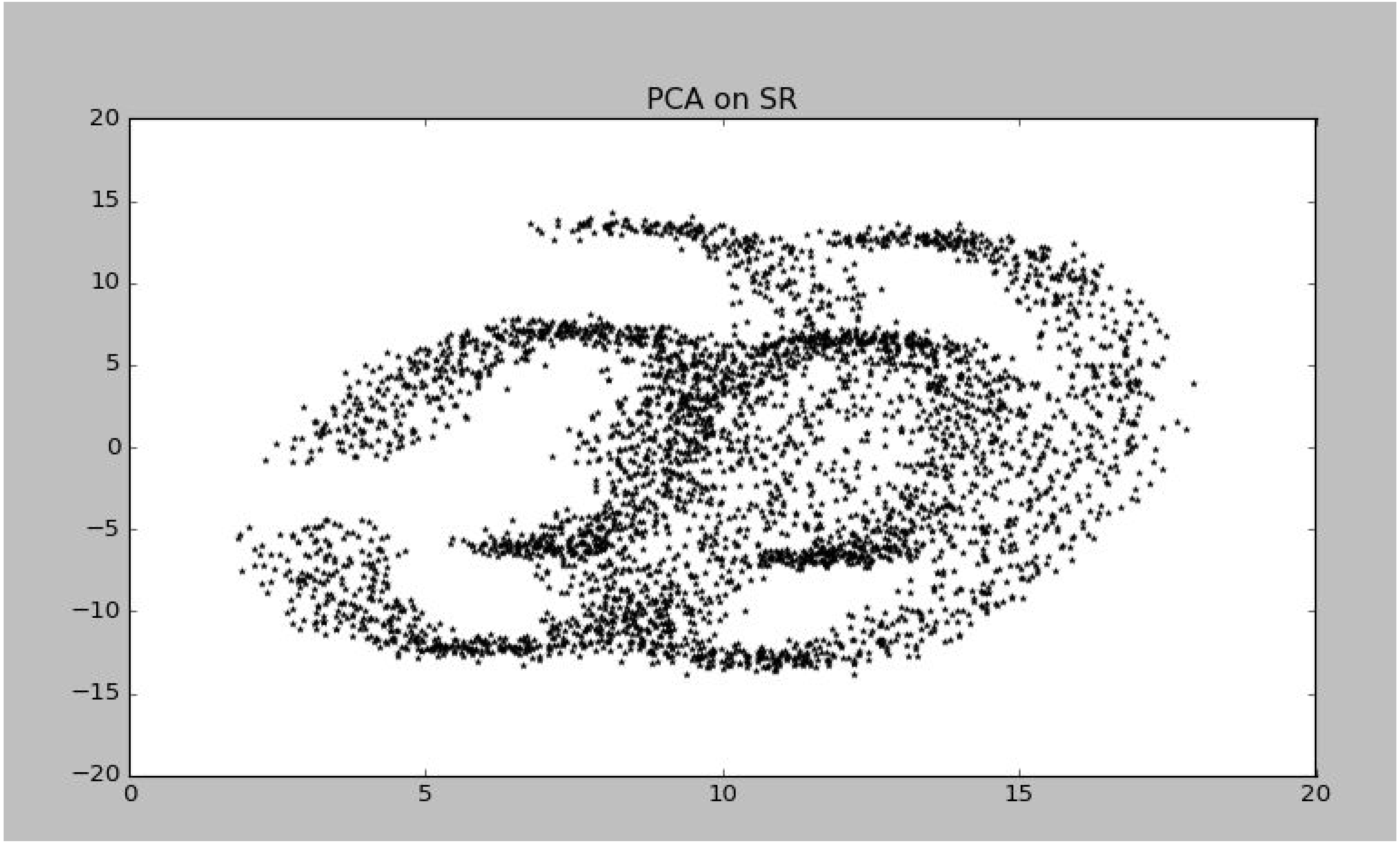


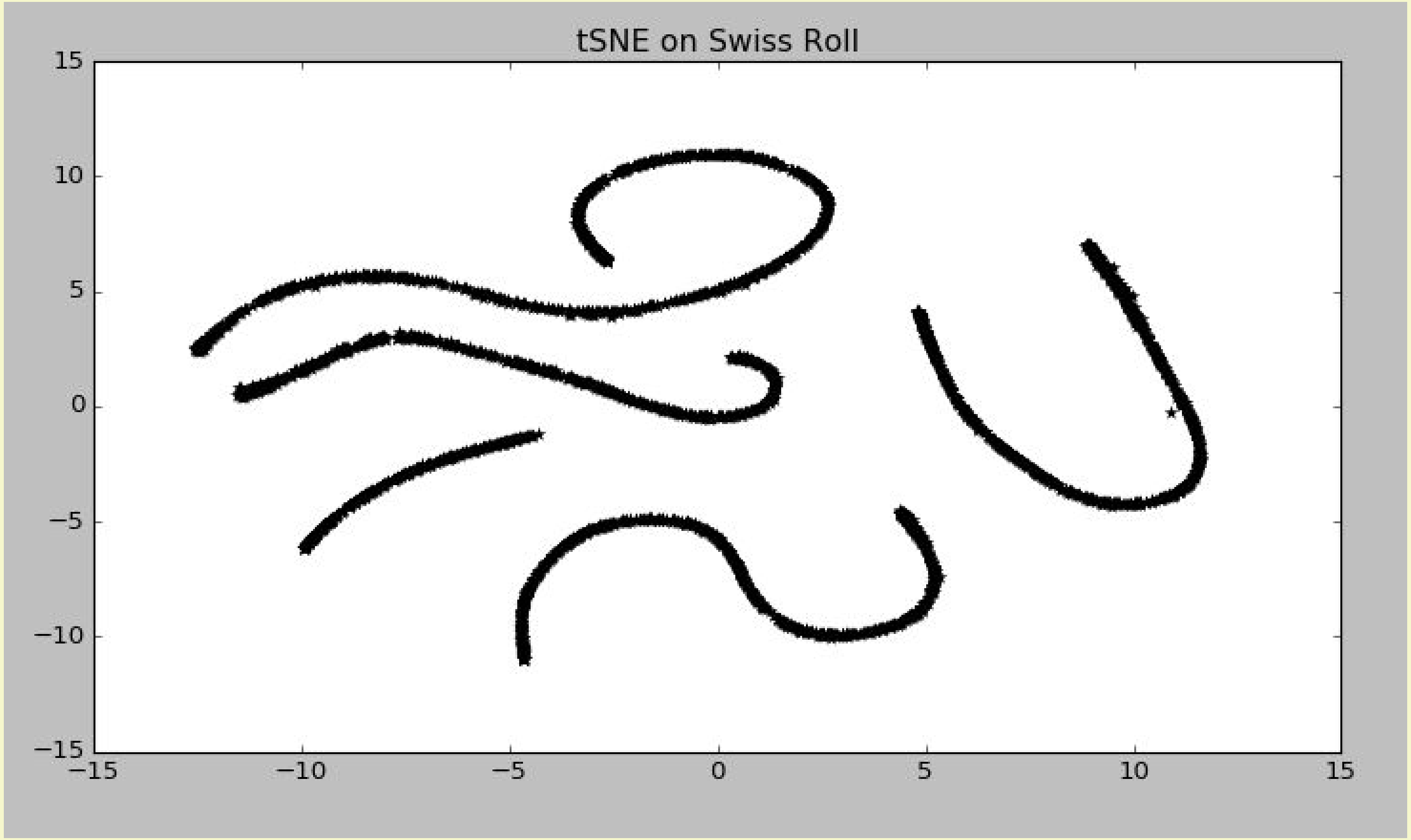
PCA visualization



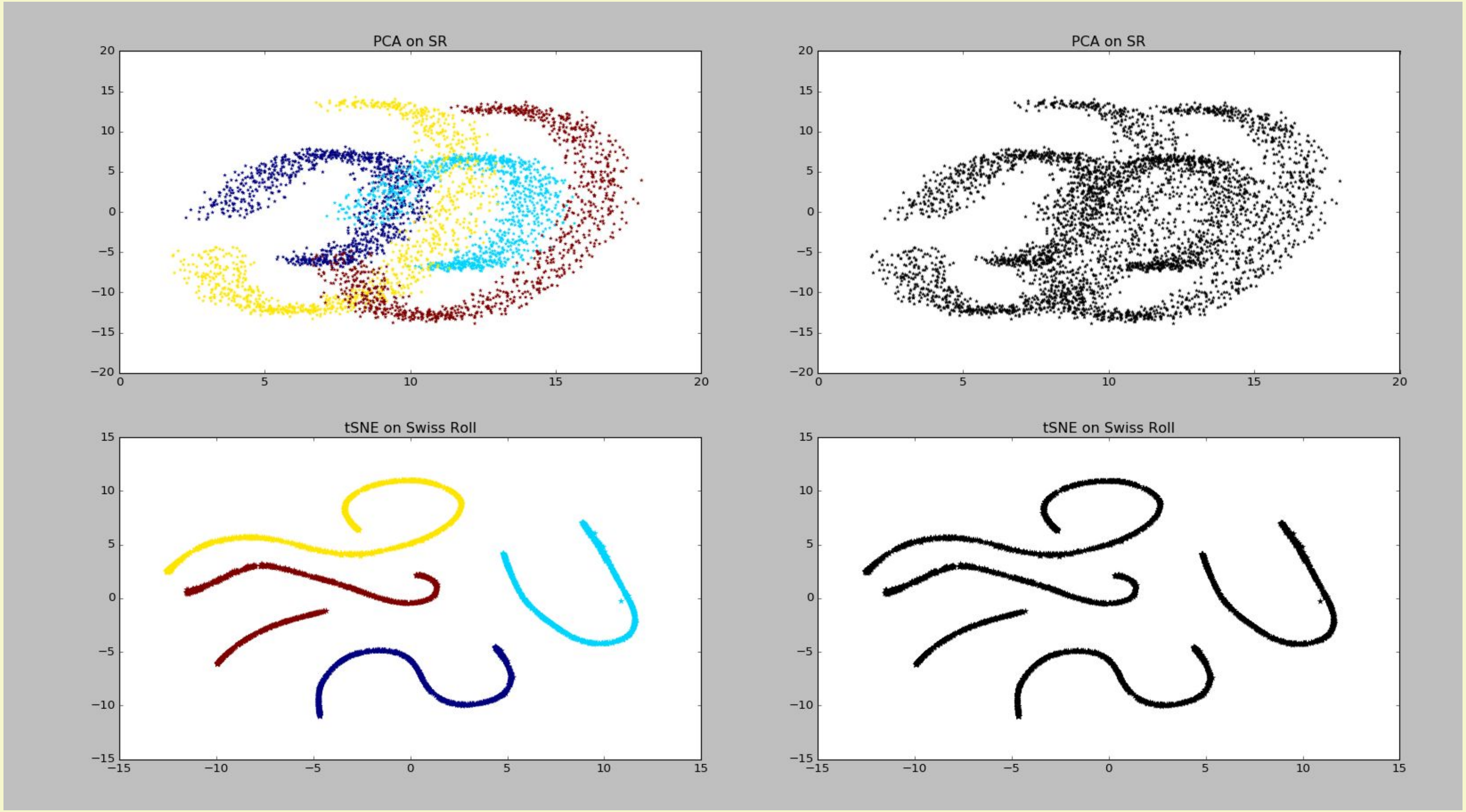
t-SNE

Image: Laurens van der Maaten
https://lvdmaaten.github.io/


PCA on SR
20
15
10
5
0
−5
−10
−15
−20
0
5
10
15
20



tSNE on Swiss Roll
15
10
5
0
−5
−10
−15
−15
−10
−5
0
5
10
15



PCA on SR
PCA on SR
tSNE on Swiss Roll
tSNE on Swiss Roll

t-SNE

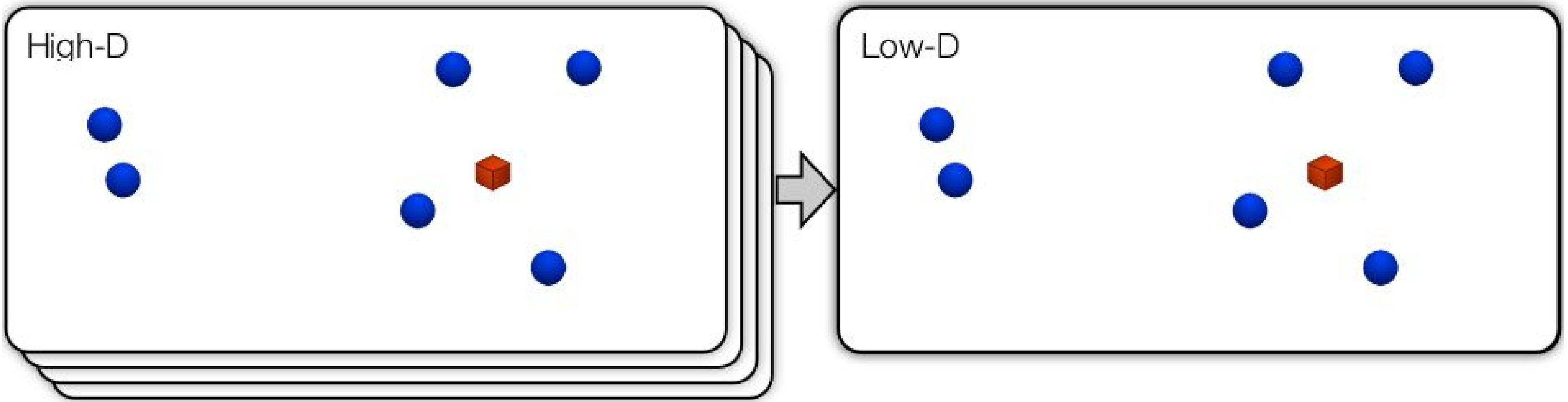
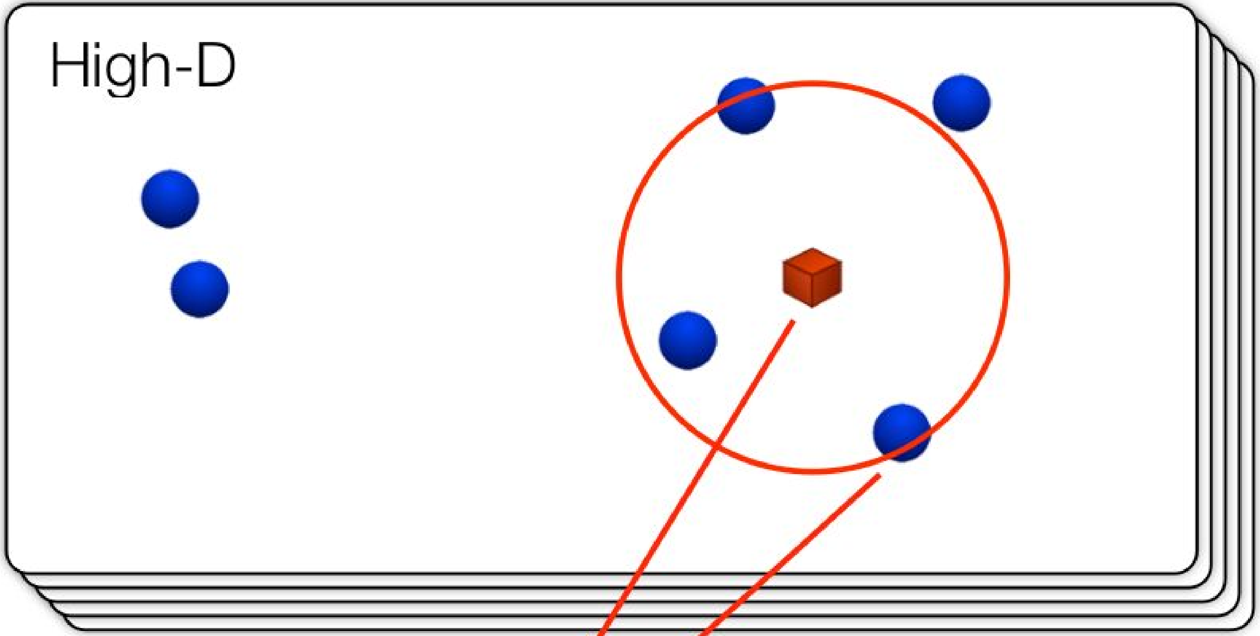


Image: Laurens van der Maaten
https://lvdmaaten.github.io/

High-D



$$p_{ij} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)}{\sum_k \sum_{l \neq k} \exp(-\|\mathbf{x}_k - \mathbf{x}_l\|^2/2\sigma^2)}$$

Image: Laurens van der Maaten
https://lvdmaaten.github.io/

Low-D

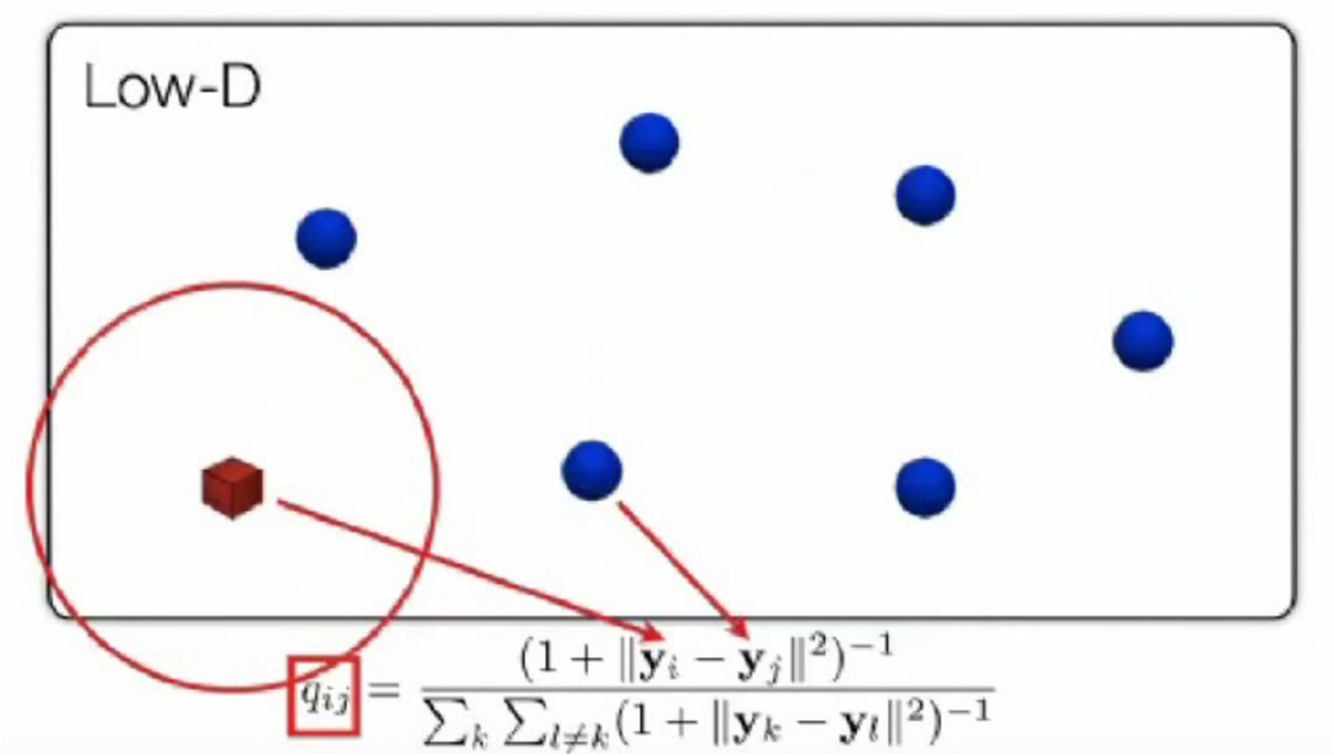


Image: Laurens van der Maaten https://lvdmaaten.github.io/

What distances to preserve?

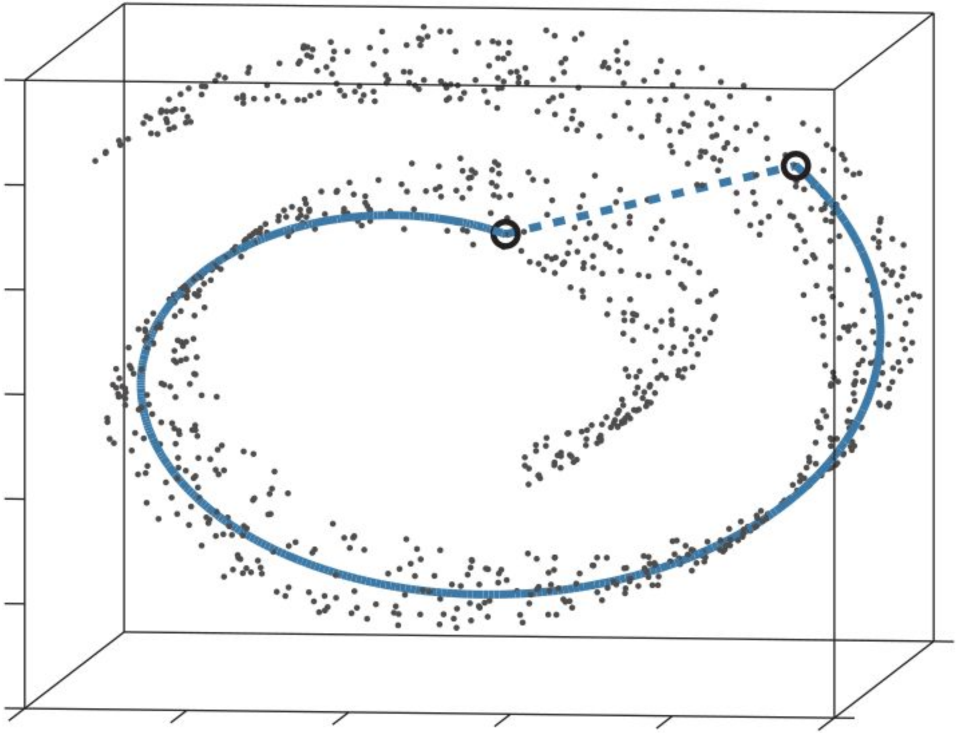


Image: Laurens van der Maaten https://lvdmaaten.github.io/

$$KL(P||Q) = \sum_{i} \sum_{j \neq i} p_{ij} log(\frac{p_{ij}}{q_{ij}})$$

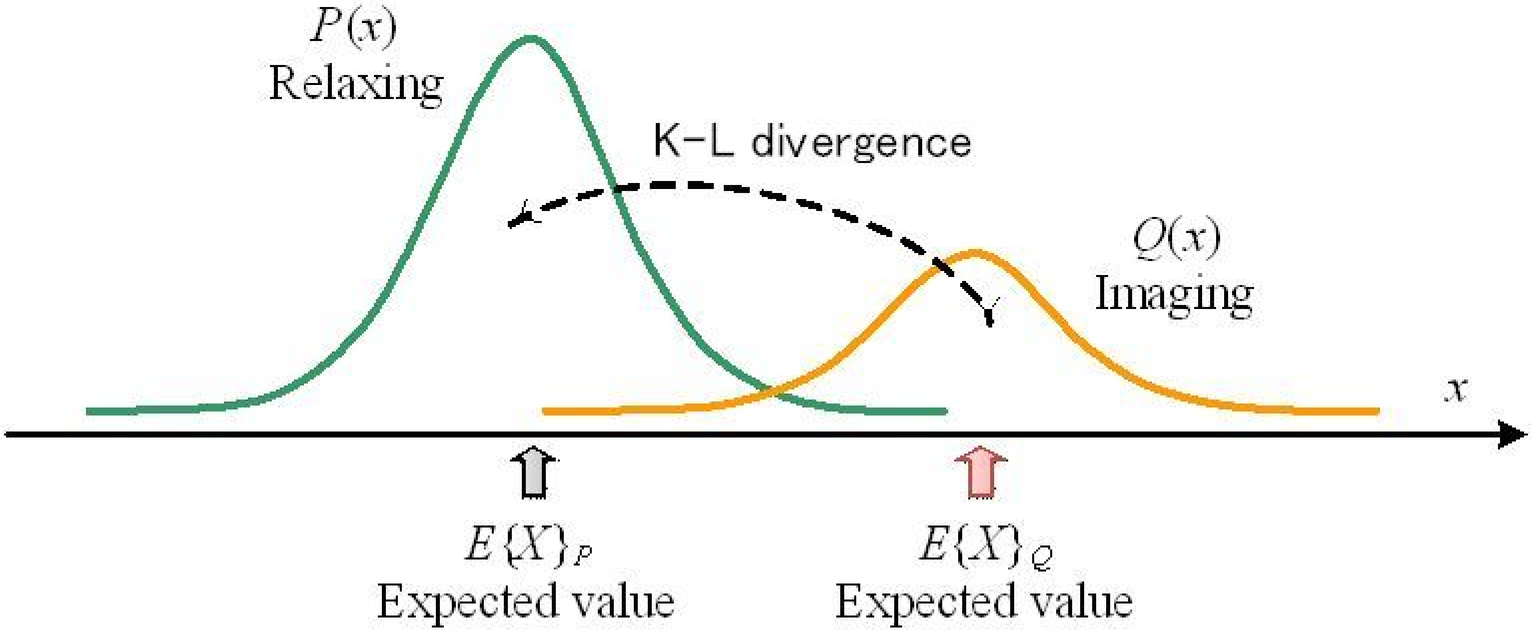
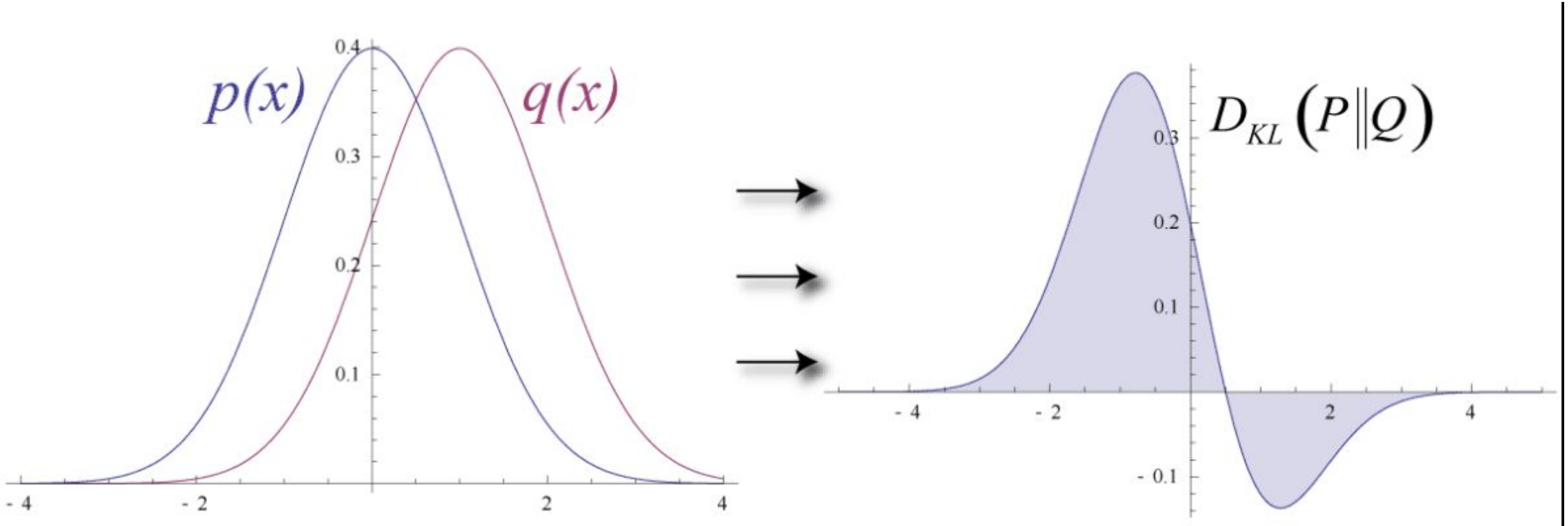


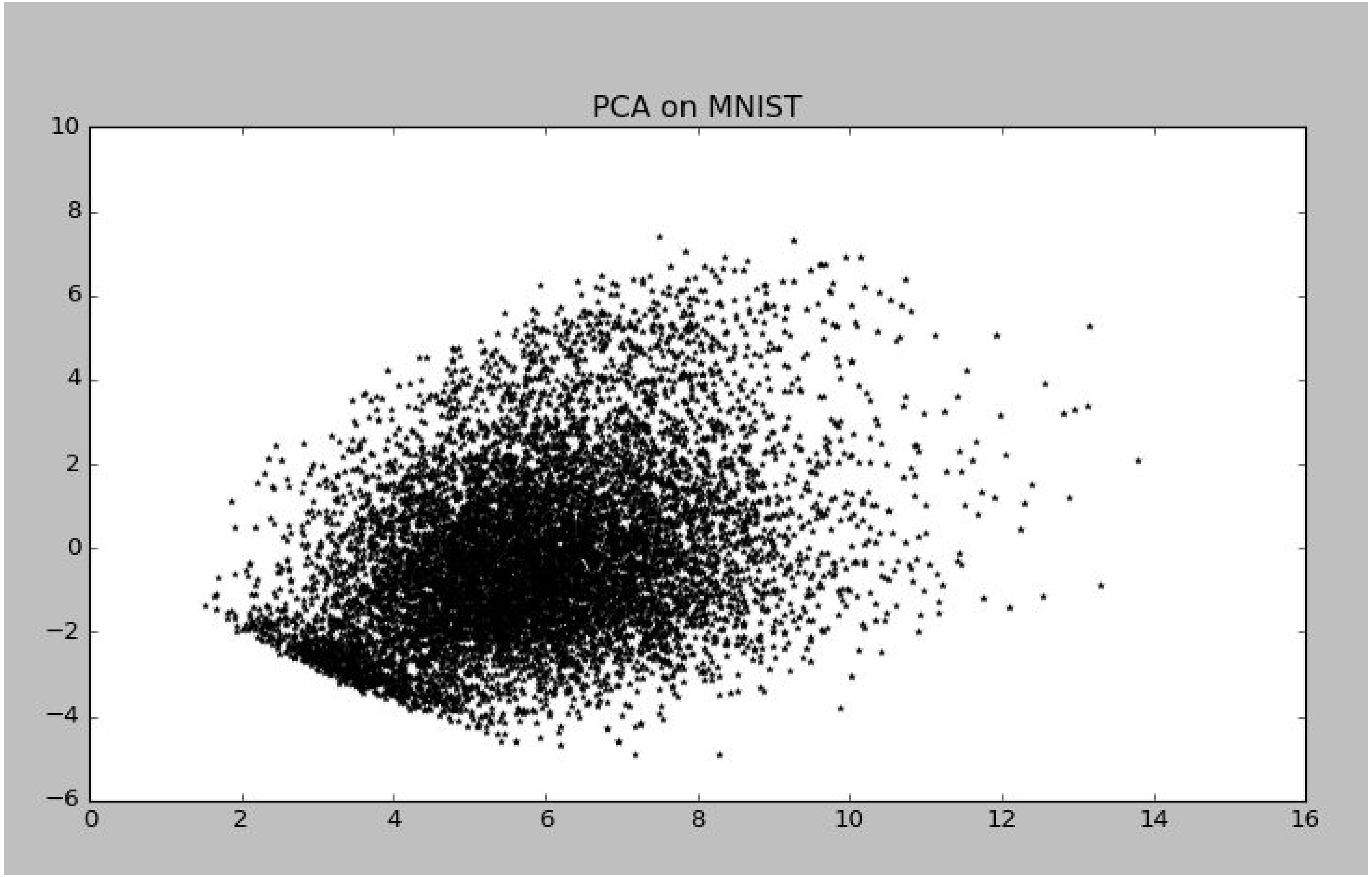
Image: Wikipedia - KL divergence

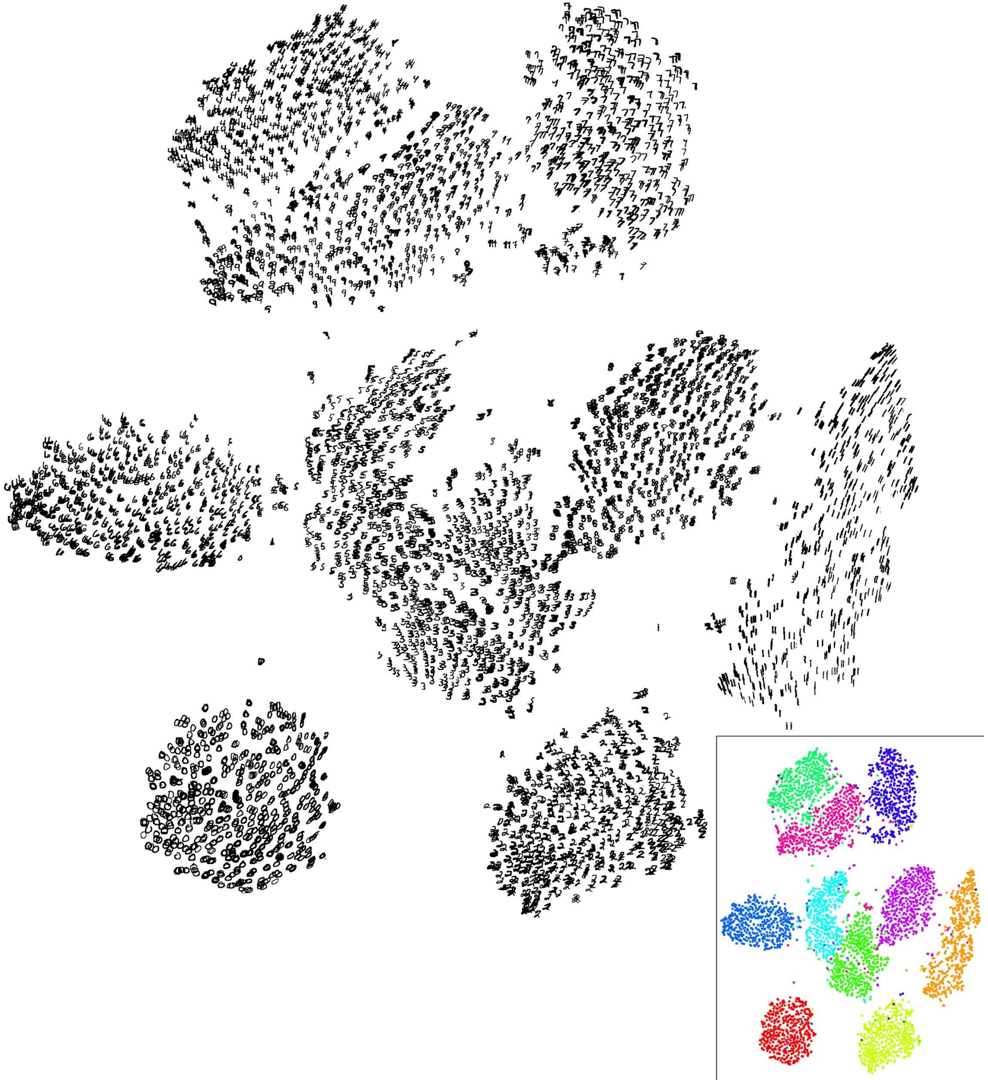


Original Gaussian PDF's

KL Area to be Integrated

Image: Wikipedia - KL divergence


PCA on MNIST
10
8
6
4
2
0
−2
−4
−6
0
2
4
6
8
10
12
14
16



WDBC

a) radius (mean of distances from center to points on the perimeter)

b) texture (standard deviation of gray-scale values)

c) perimeter

d) area

e) smoothness (local variation in radius lengths)

f) compactness (perimeter^2 / area - 1.0)

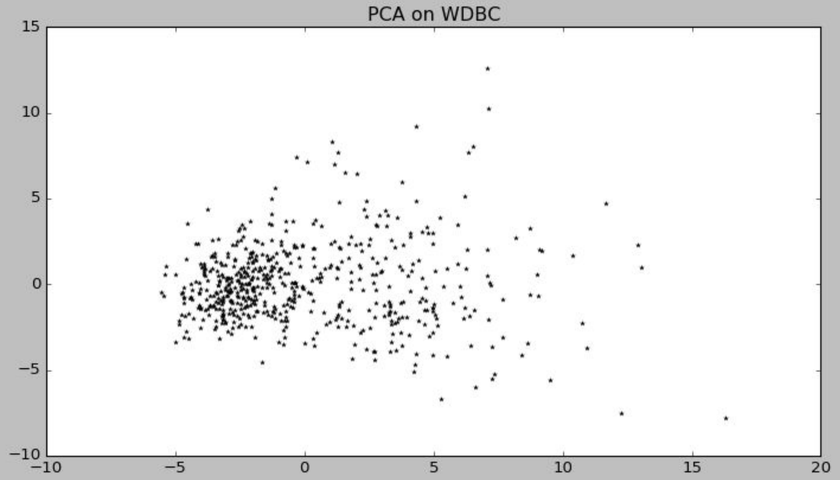
g) concavity (severity of concave portions of the contour)

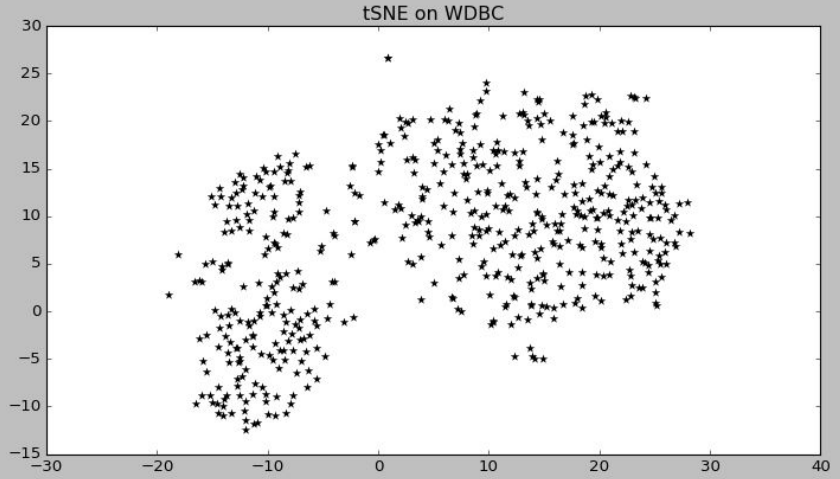
h) concave points (number of concave portions of the contour)

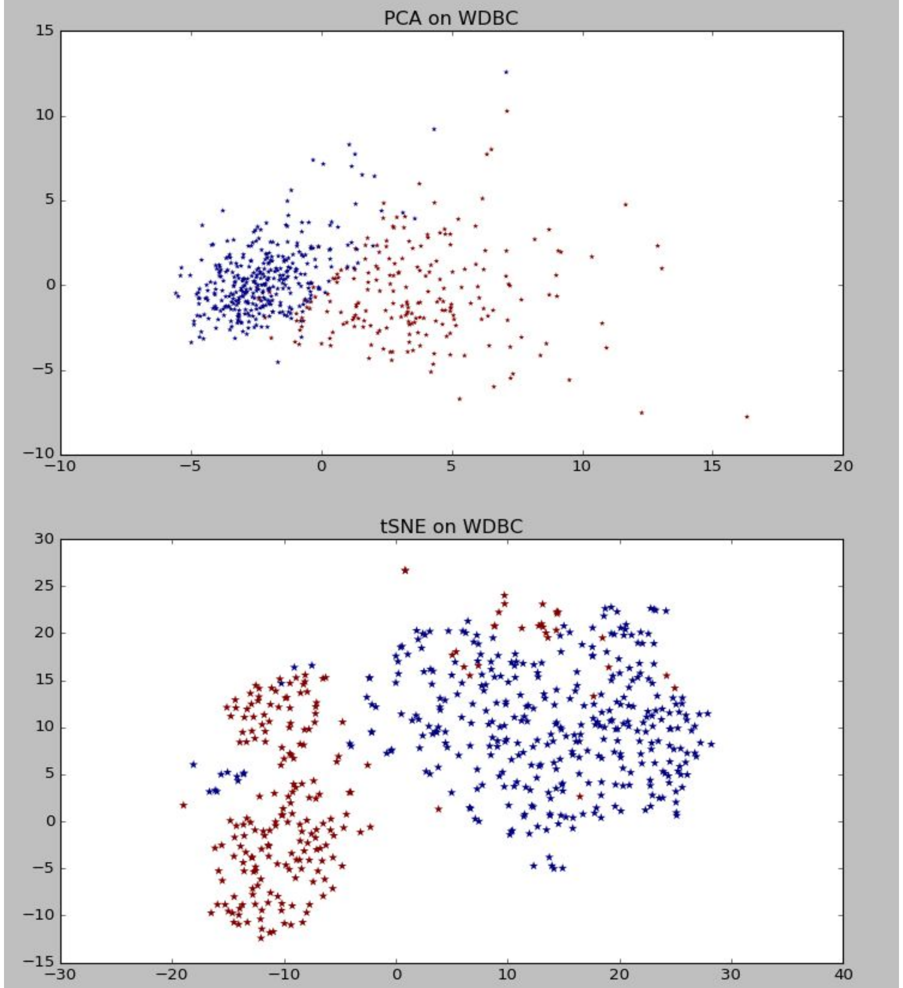
i) symmetry

j) fractal dimension ("coastline approximation" - 1)

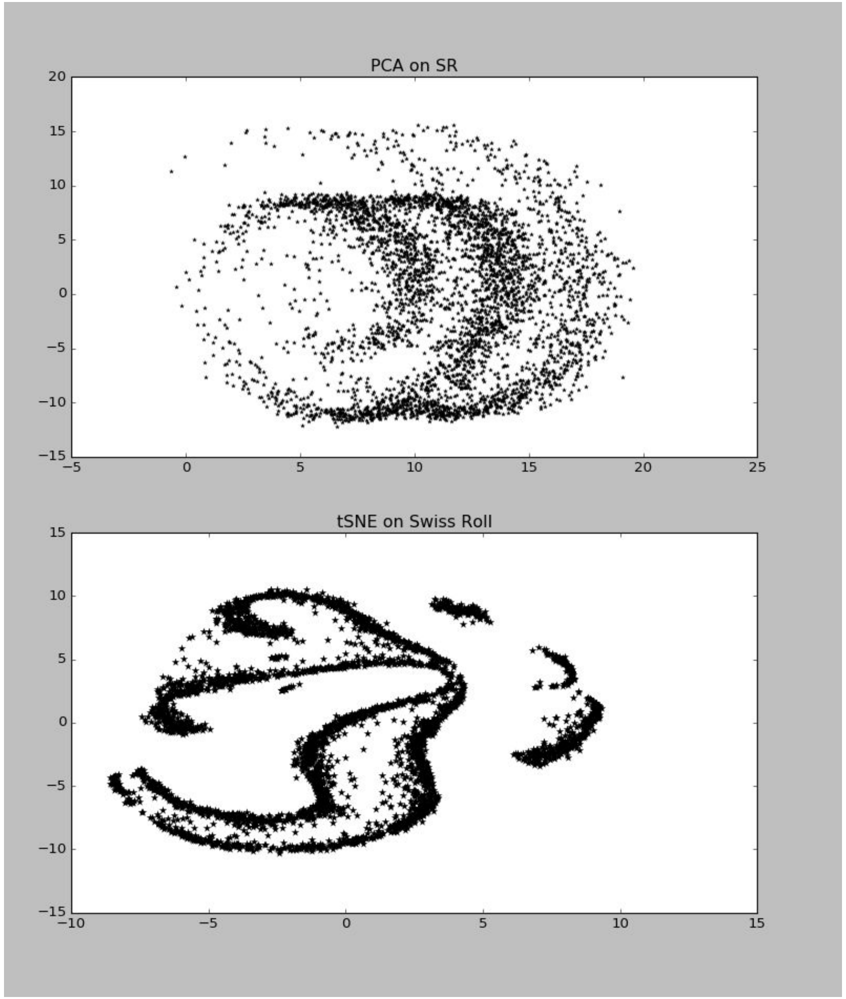
Comparison

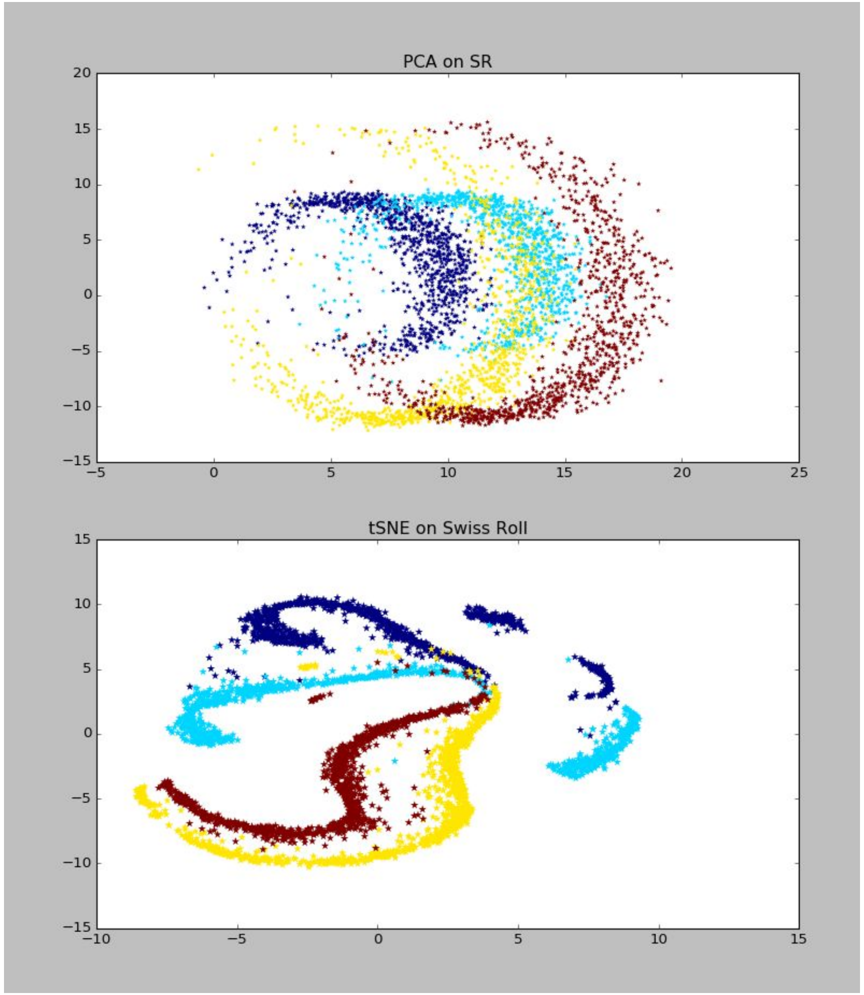

PCA on WDBC
15
10
5
0
−5
−10
−10
−5
0
5
10
15
20


tSNE on WDBC
30
25
20
15
10
5
0
−5
−10
−15
−30
−20
−10
0
10
20
30
40



PCA on WDBC
15
10
5
0
−5
−10
−10
−5
0
5
10
15
20
tSNE on WDBC
30
25
20
15
10
5
0
−5
−10
−15
−30
−20
−10
0
10
20
30
40


PCA on SR
tSNE on Swiss Roll


PCA on SR
tSNE on Swiss Roll

Assignment

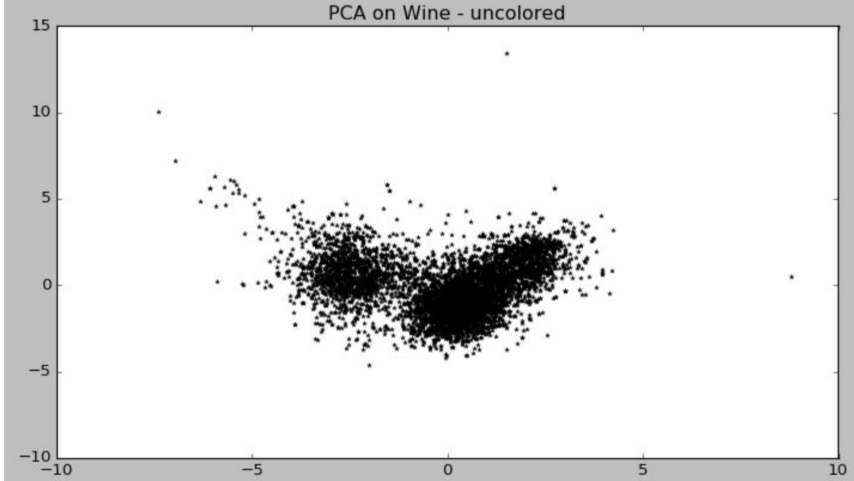
- Groups of 3
- Make best visualization

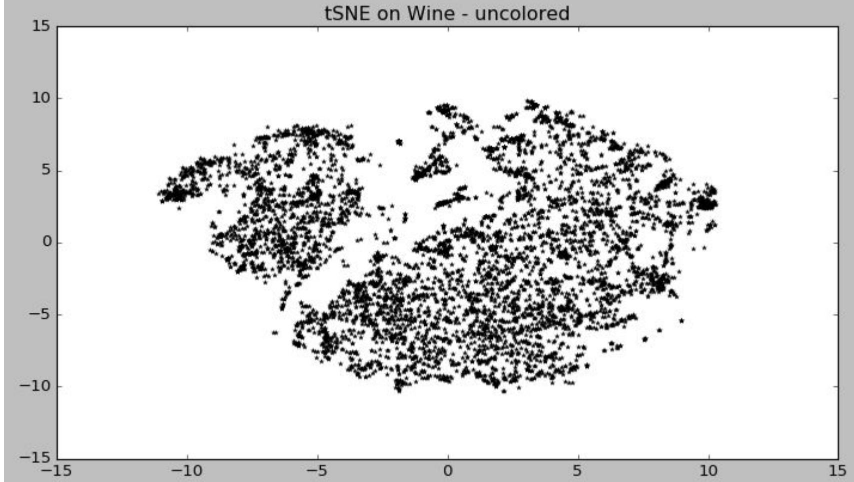
- 45min: hack
- 15min: insert labels
- 15min: presentations & discussions

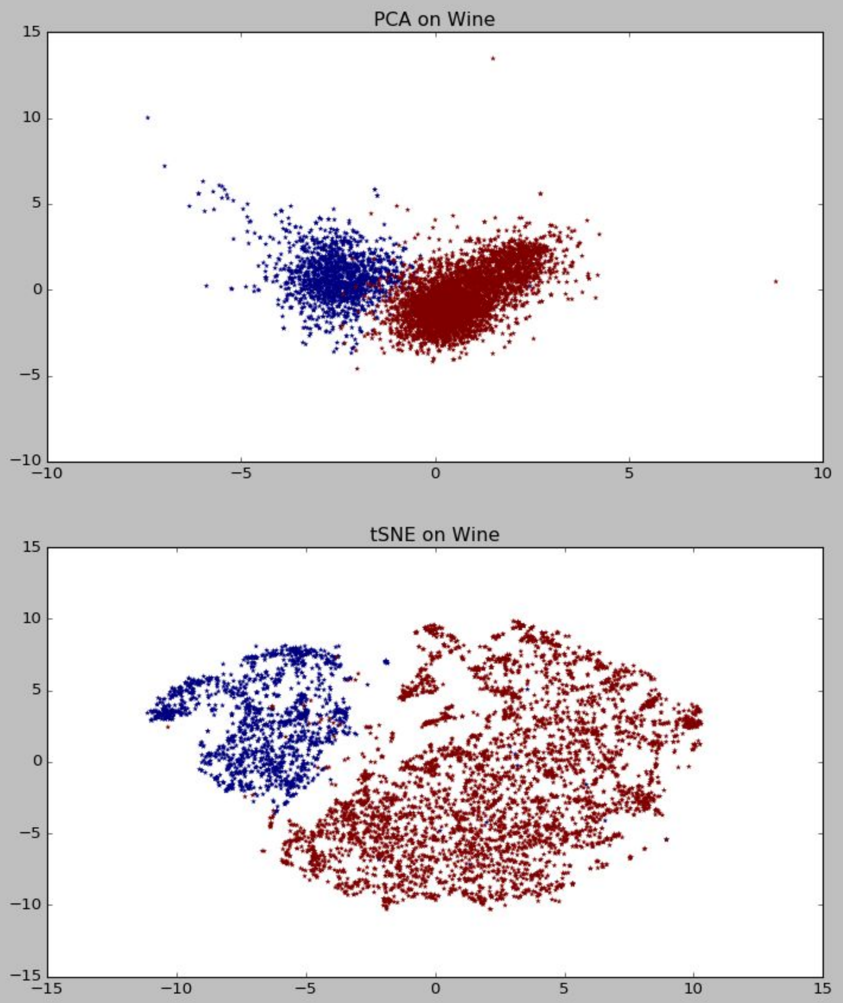
Details of the assignment

- github.com/RobRomijnders/EDS
    - Vis_unsuper.py to review the plots in the slides

Spoiler


PCA on Wine - uncolored
15
10
5
0
−5
−10
−10
−5
0
5
10


tSNE on Wine - uncolored
15
10
5
0
−5
−10
−15
−15
−10
−5
0
5
10
15


PCA on Wine
tSNE on Wine