session 3:

Unsupervised learning

M. Kundegorski

13th December 2019

IAFIG-RMS - Bioimage Analysis With Python

Cambridge Bioinformatics Training Centre

Unsupervised learning

What can you do when you have no answers?

Uses of unsupervised learning

- pre-processing (sparse features clustering)
- simplifying data for supervised algorithms
- clustering for data exploration:
  - sparse data
  - big data
  - data for which we lack intuition to choose a statistical model
- label-free segmentation

Principles

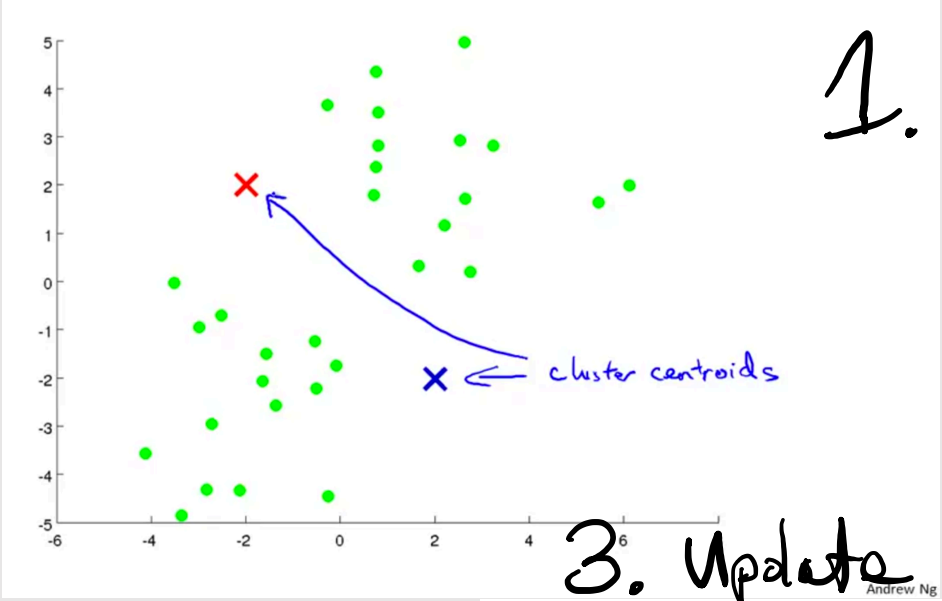
- exploration of variable space in looking for emerging patterns of similarity
- there is no provided ground truth, and in fact often it is not available at all.
- often used with huge amounts of data
- provided: knowledge of the feature space - for instance a definition of distance in feature space - that allows to draw some outcomes from provided data.

Clustering: finding data similarity

source: https://www.coursera.org/learn/machine-learning

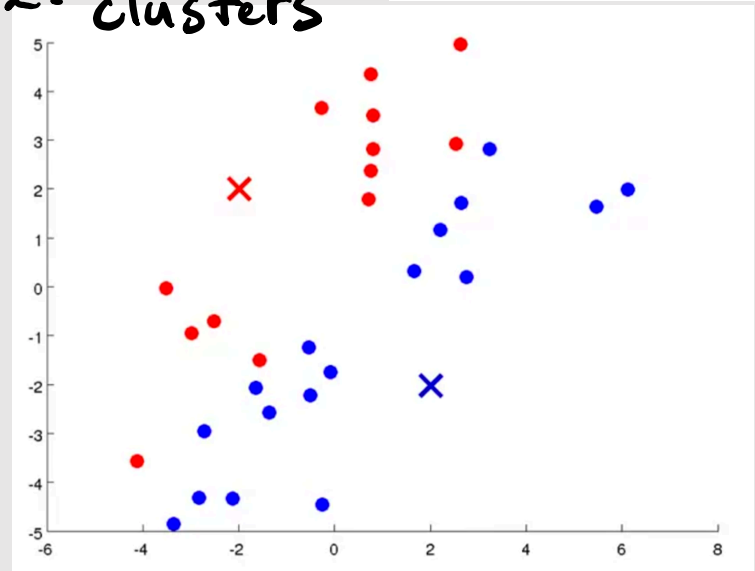
k-means

K = 2

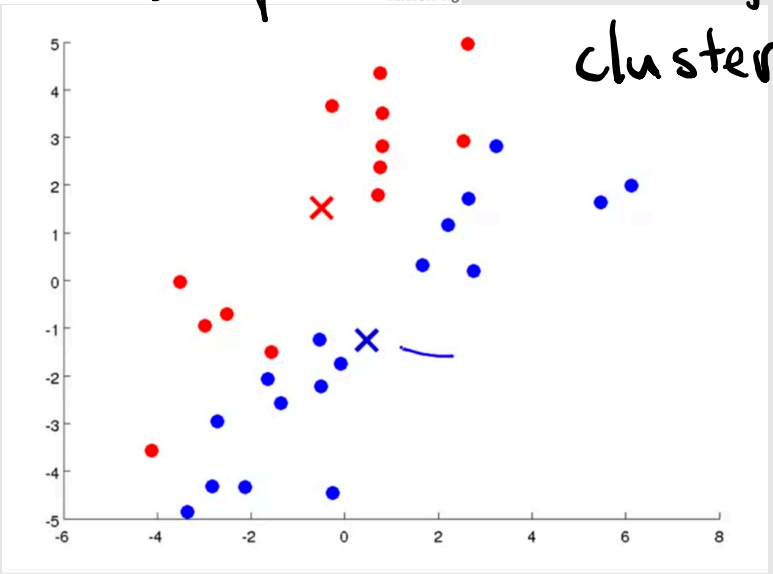


1. random initialisation

2. find clusters



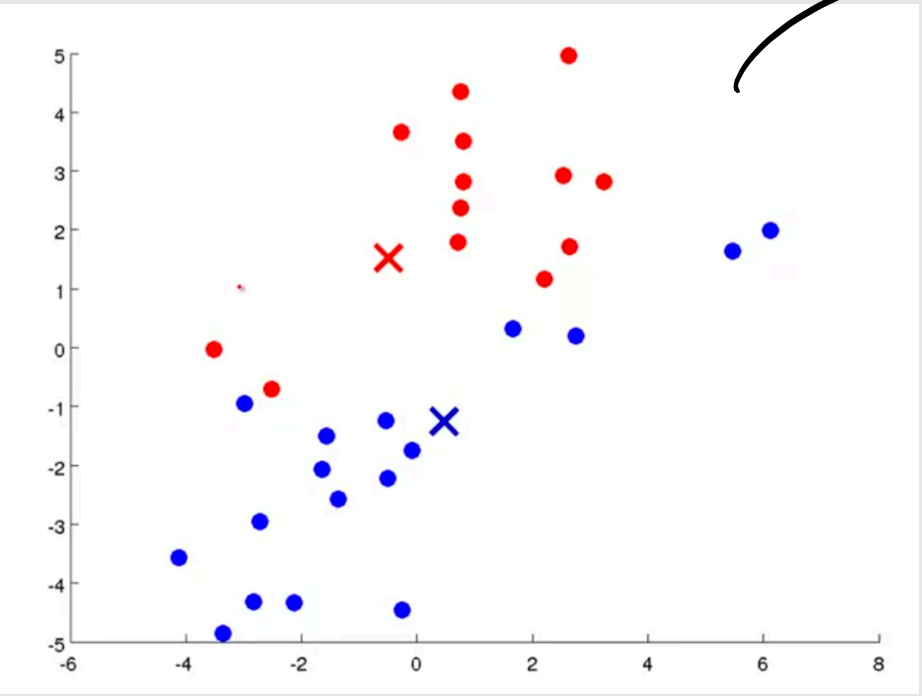
3. Update mean of clusters

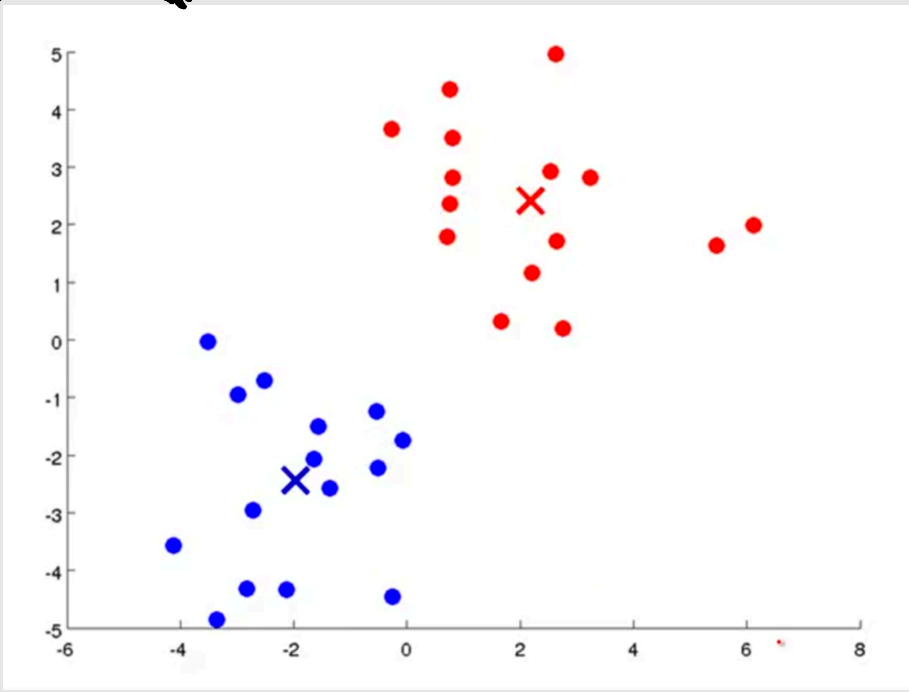


source: https://www.coursera.org/learn/machine-learning

k means

convergence

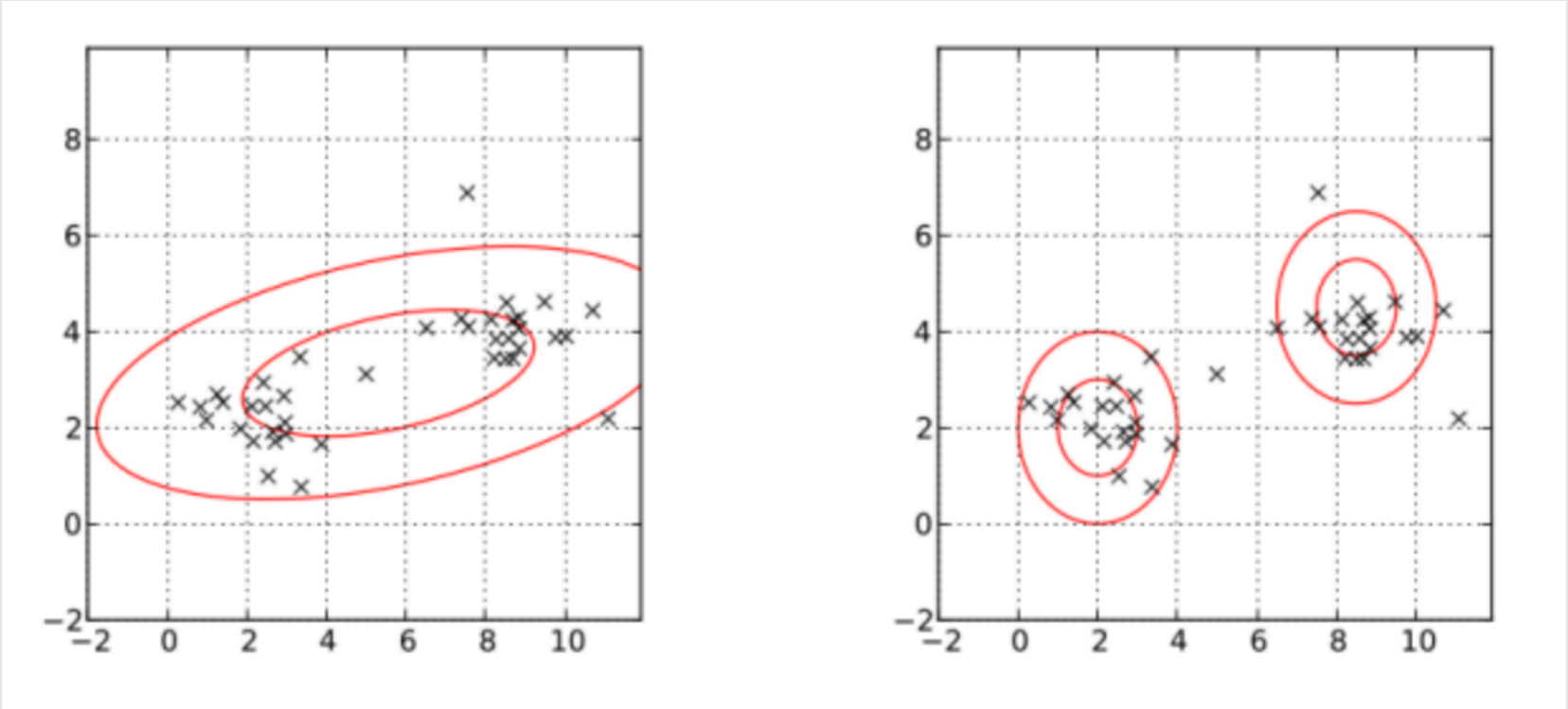




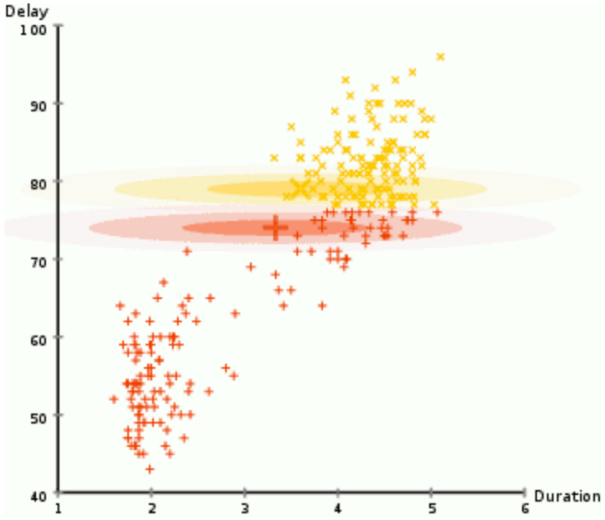
k-medoids

- Just like k-means but choosing median data point as centre

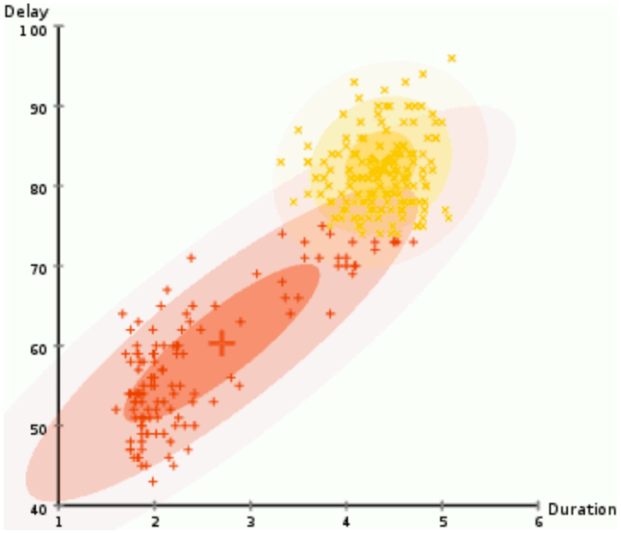
Mixture of Gaussians



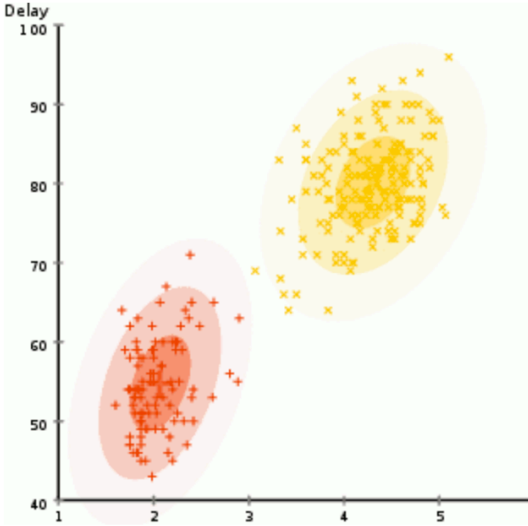
Expectation-Maximisation



The EM algorithm updating the parameters of a two-component bivariate Gaussian mixture model.[2]



The EM algorithm updating the parameters of a two-component bivariate Gaussian mixture model.[2]

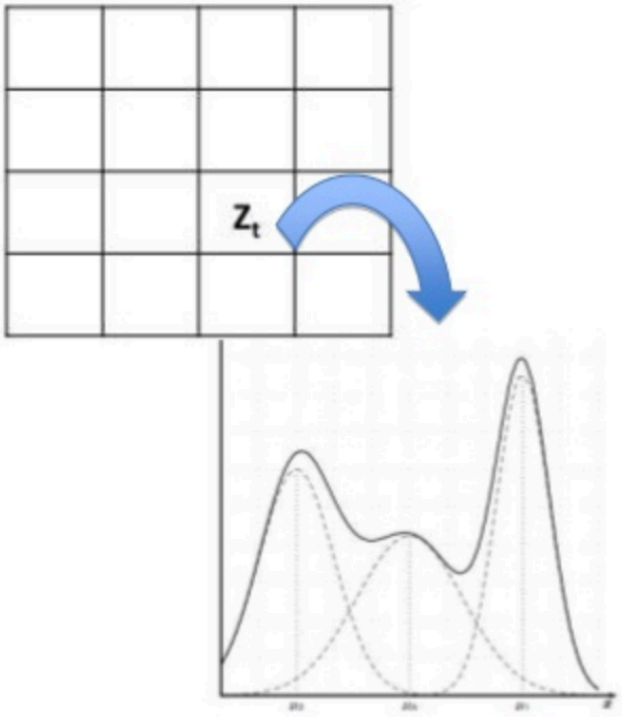


The EM algorithm updating the parameters of a two-con bivariate Gaussian mixture model.[2]

https://commons.wikimedia.org/wiki/File:EM_Clustering_of_Old_Faithful_data.gif

Gaussian Mixture Model

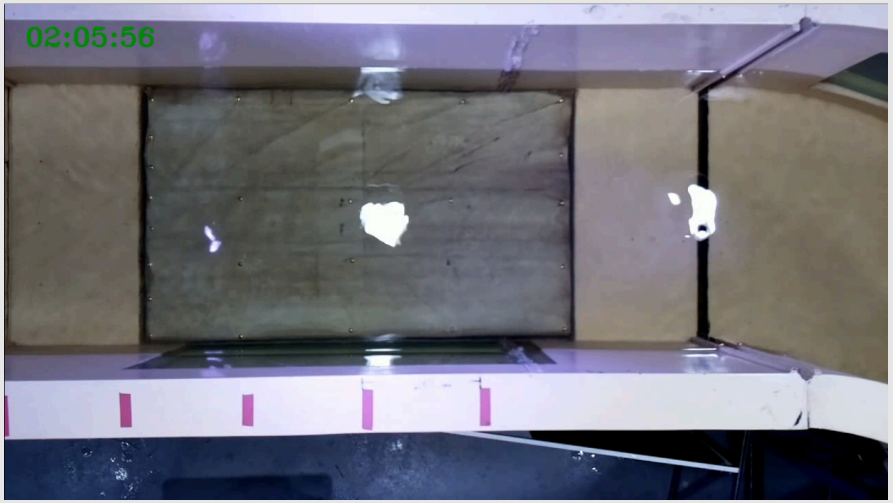
Gaussian Mixture Model (GMM)



- A GMM is a mixture pdf which is a linear combination of K Gaussian pdfs.
- $\Sigma w_i = 1$
- each pixel is given one GMM

source: Zivkovic , 2004. Improved adaptive gaussian mixture model for background subtraction. Figure by Raviraj singh shekhawat

Example:



Mean-shift segmentation

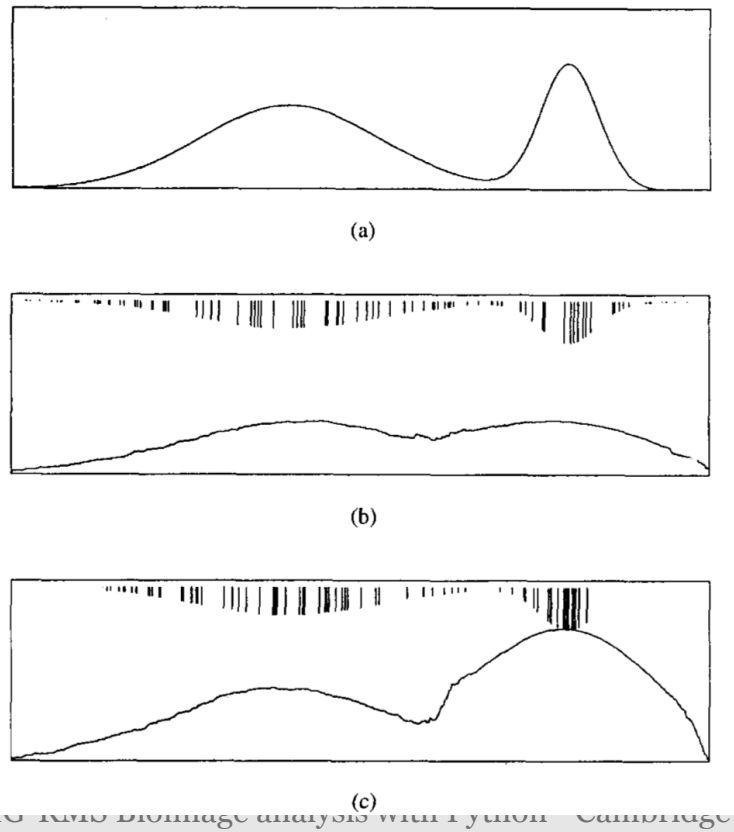
- KDE + gradient ascent

Mean Shift, Mode Seeking, and Clustering

Yizong Cheng

***Abstract*—Mean shift, a simple iterative procedure that shifts each data point to the average of data points in its neighborhood, is generalized and analyzed in this paper. This generalization makes some *k*-means like clustering algorithms its special cases. It**

A relation among kernels called "shadow" will be defined in Section III. It will be proved that mean shift on any kernel is equivalent to gradient ascent on the density estimated with a shadow of its. Convergence and its rate is the subject of Sec-



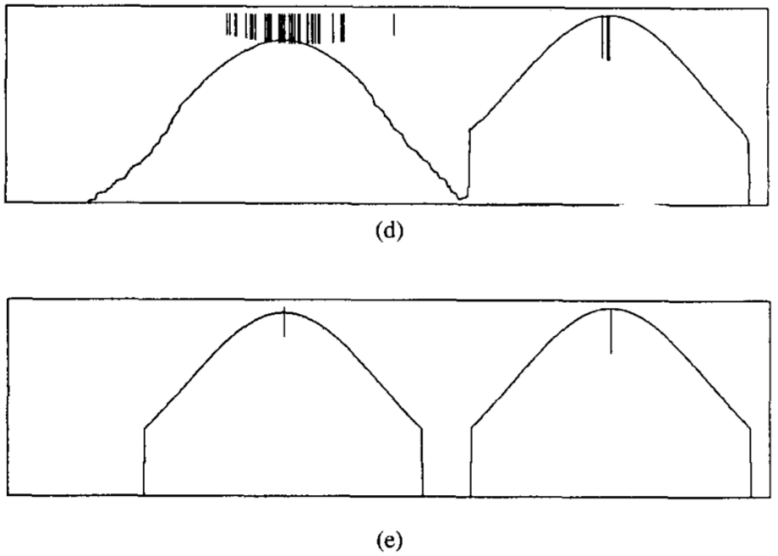
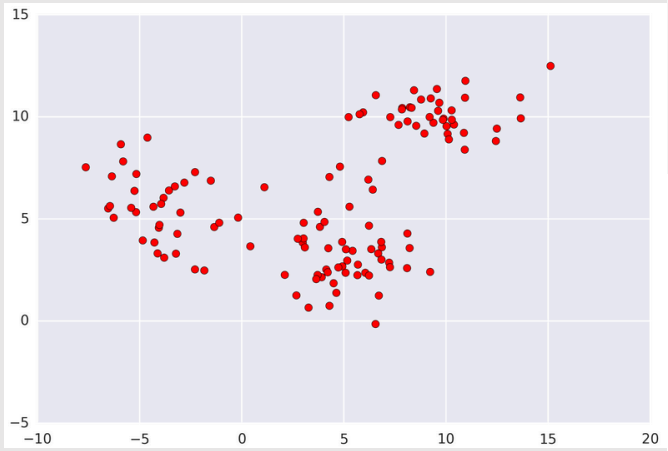
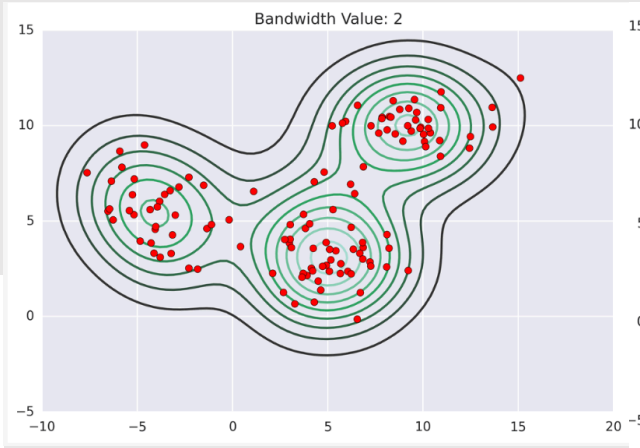


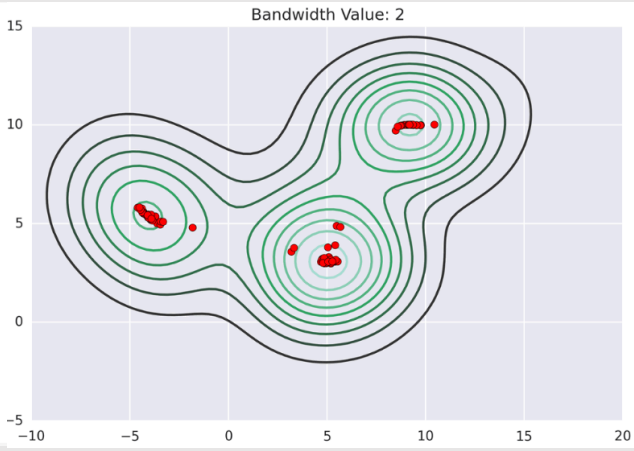
Fig. 8. Multistart global optimization using blurring. (a) shows the function f, whose global maximum is to be found. The next four figures show the mean shift of S, at the (b) initial, (c) first, (d) third, and (e) fifth iterations of a blurring process when f is used as the weight function. In each of these four figures, the vertical bars show the positions and f values of the S points, and the curve shows the q function, whose local maxima locations approximate those of f.

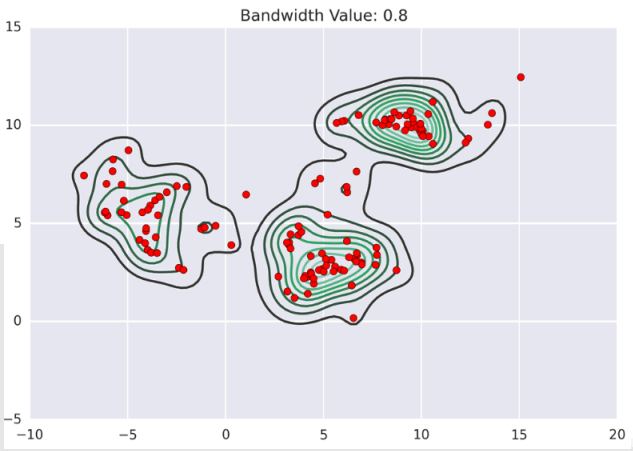
source: https://spin.atomicobject.com/2015/05/26/mean-shift-clustering/

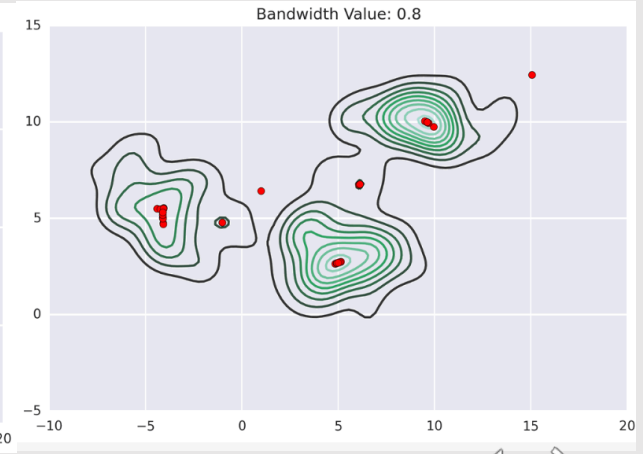
Mean-shift segmentation











DBSCAN – industry standard

Published in Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)

A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise

Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu

Institute for Computer Science, University of Munich
Oettingenstr. 67, D-80538 München, Germany
{ester | kriegel | sander | xwxu}@informatik.uni-muenchen.de

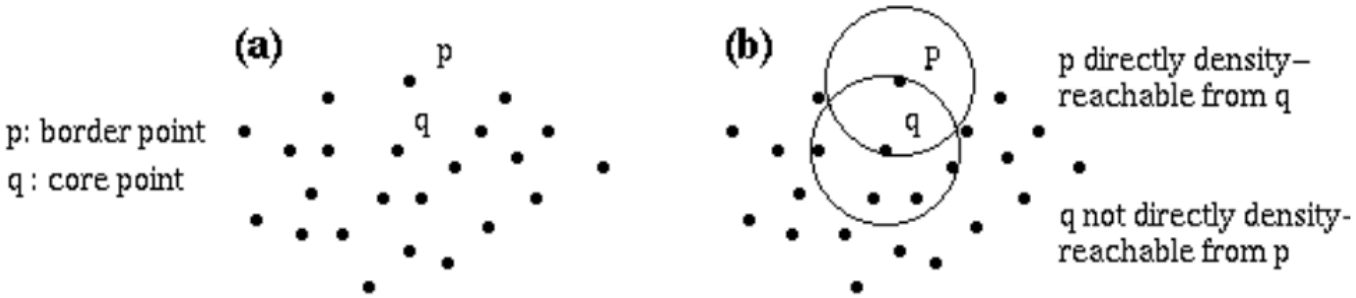


figure 2: core points and border points

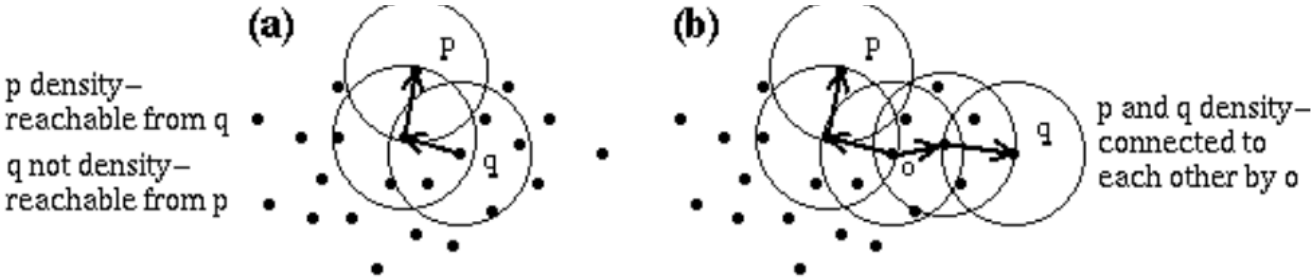


figure 3: density-reachability and density-connectivity

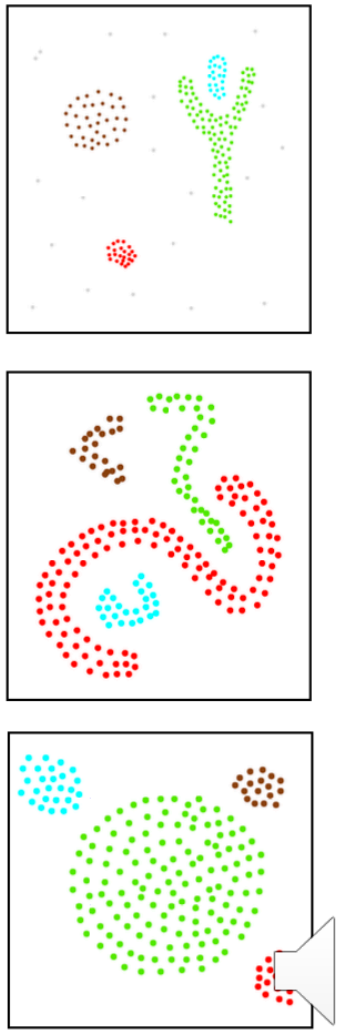
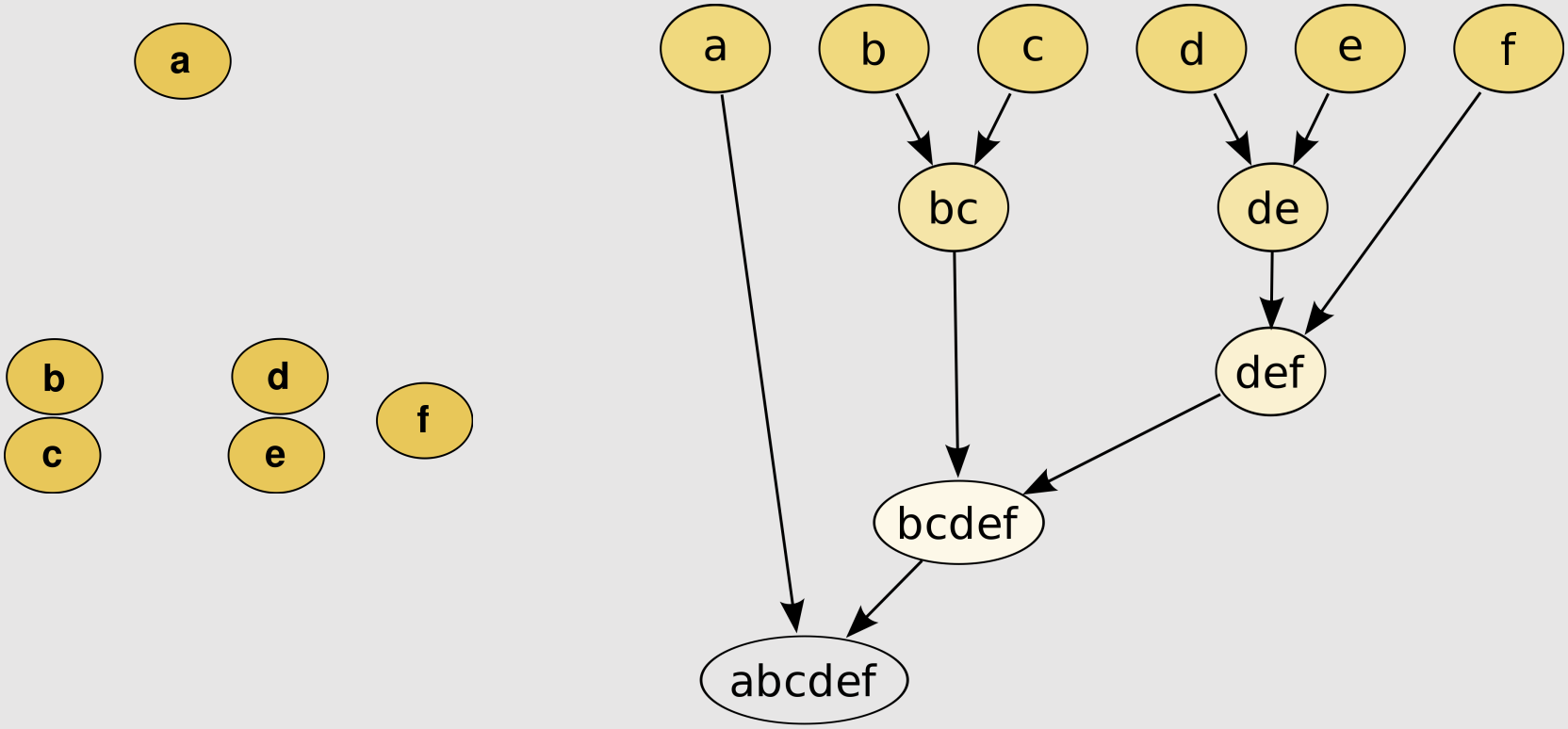


figure 6: Clusterings discovered by DBSCAN

Hierarchical clustering



https://en.wikipedia.org/wiki/Hierarchical_clustering

Component Analysis: feature selection, dimensionality reduction, encoding…

Priniciple Component Analysis

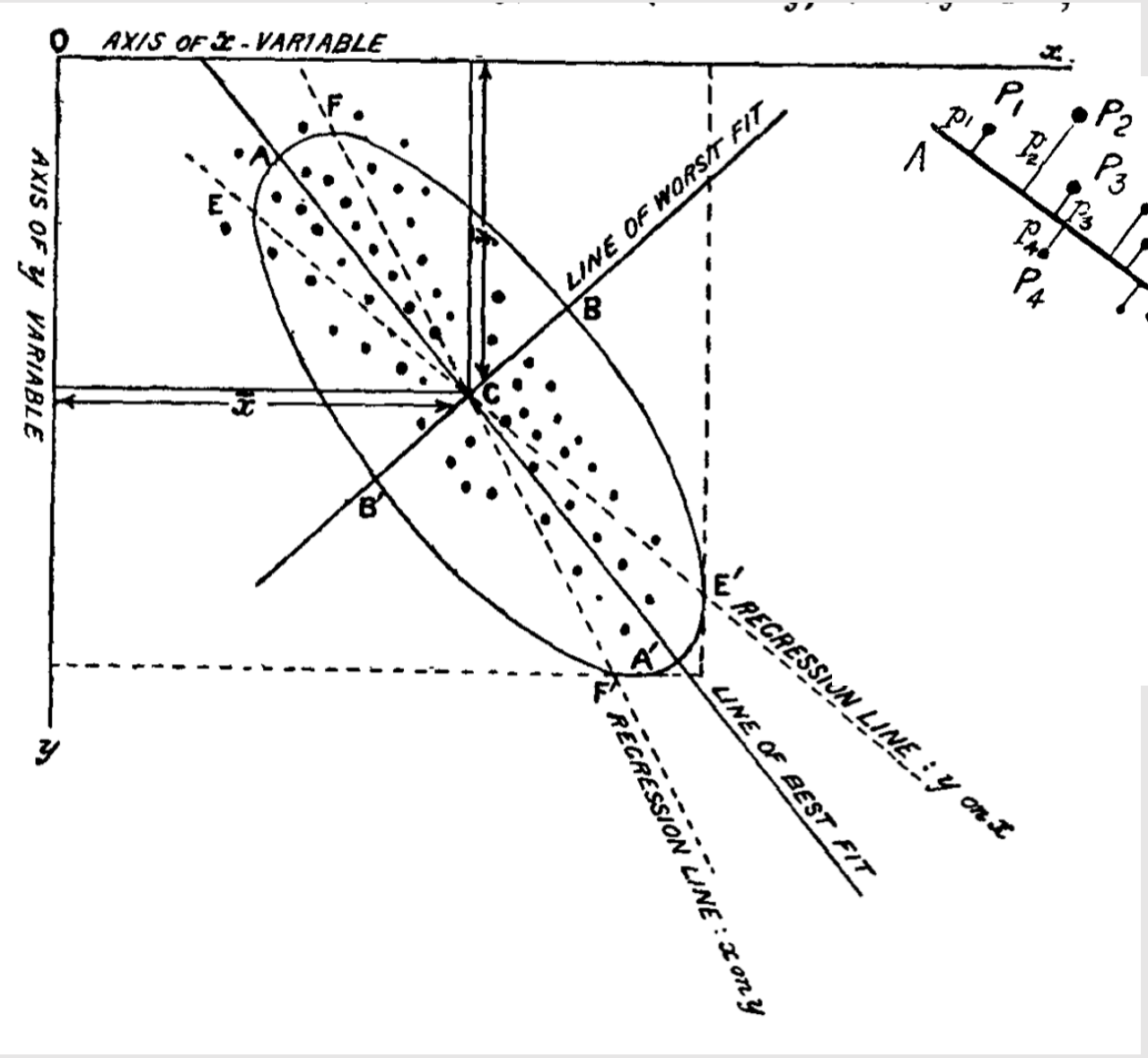
LIII. *On Lines and Planes of Closest Fit to Systems of Points in Space.* *By* KARL PEARSON, *F.R.S., University College, London* *.

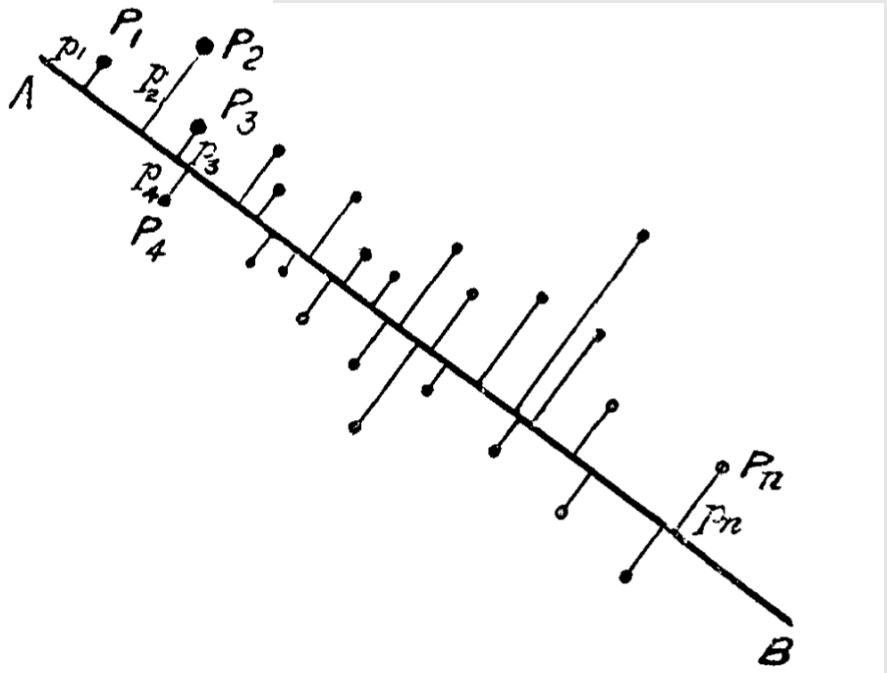
(1) IN many physical, statistical, and biological investigations it is desirable to represent a system of points in plane, three, or higher dimensioned space by the "best-fitting" straight line or plane. Analytically this consists in taking

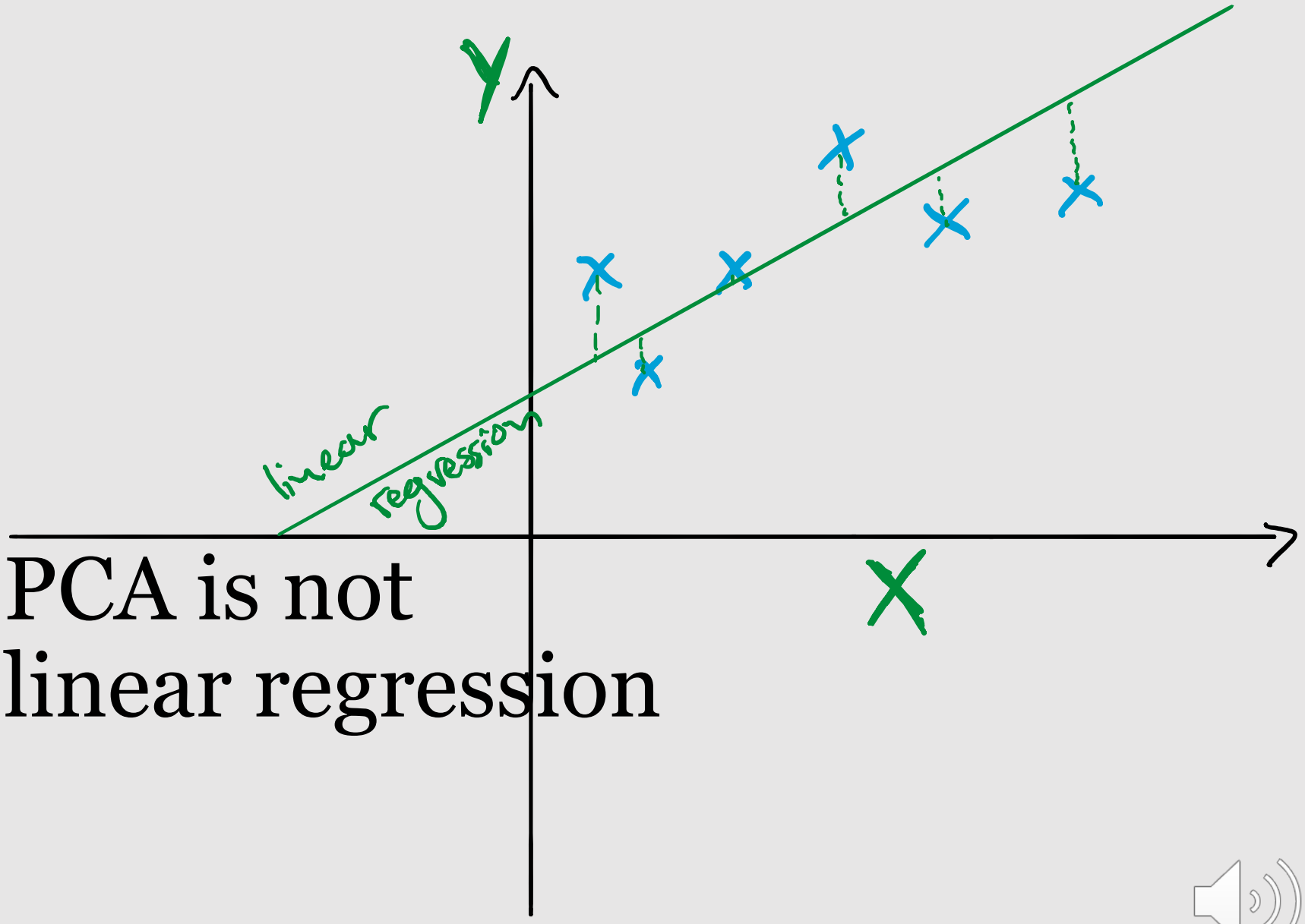
$$y = a_0 + a_1 x, \quad \text{or} \quad z = a_0 + a_1 x + b_1 y,$$

$$\text{or} \quad z = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \ldots + a_n x_n,$$

where y, x, z, x_1, x_2, $\ldots x_n$ are variables, and determining the "best" values for the constants a_0, a_1, b_1, a_0, a_1, a_2, a_3, $\ldots a_n$

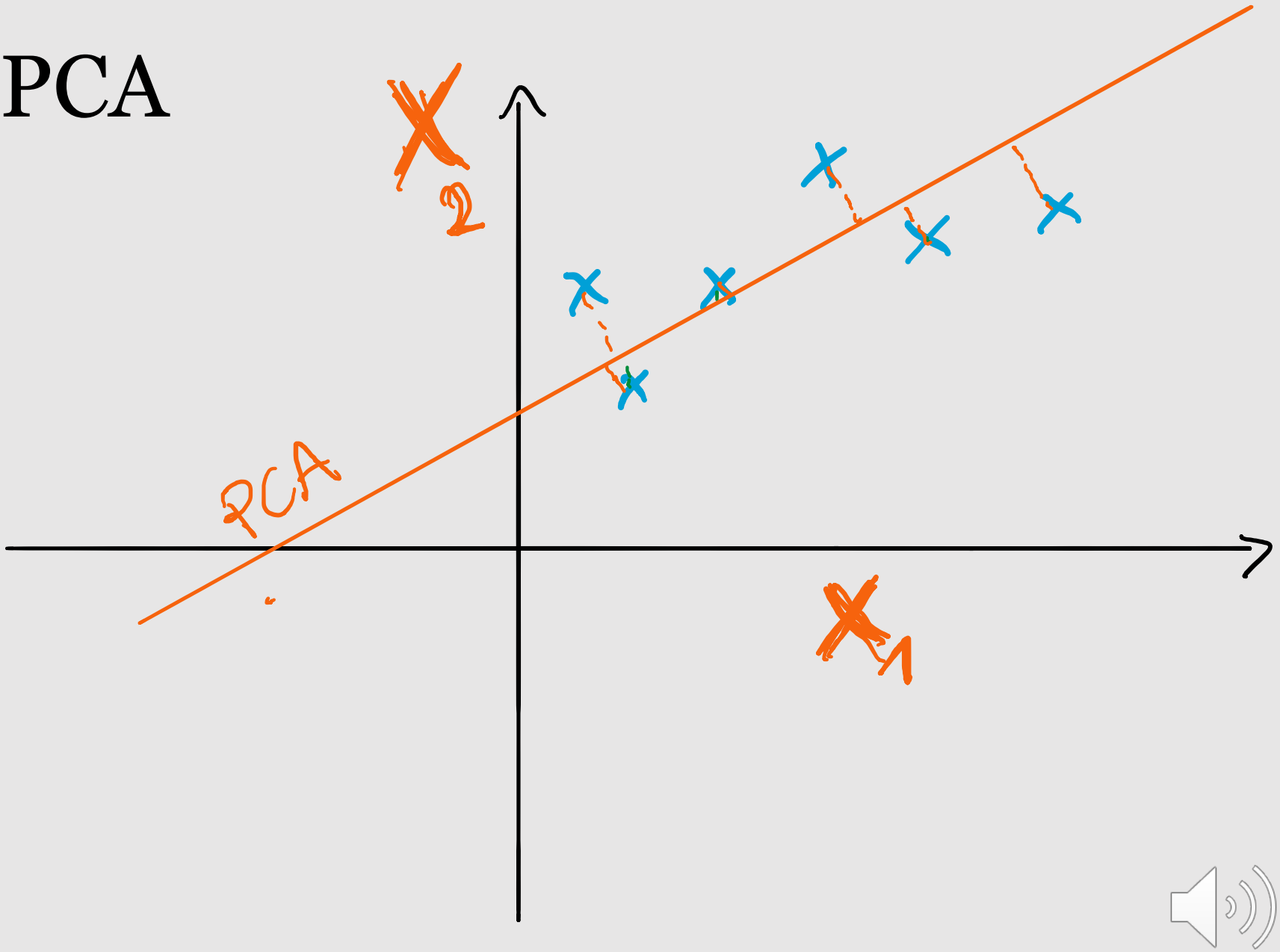

O
AXIS OF x - VARIABLE
x
AXIS OF y VARIABLE
y
F
A
E
LINE OF WORST FIT
B
C
x̄
B'
E'
REGRESSION LINE : y on x
A'
LINE OF BEST FIT
F'
REGRESSION LINE : x on y


P1
P2
P3
P4
Pn
A
B

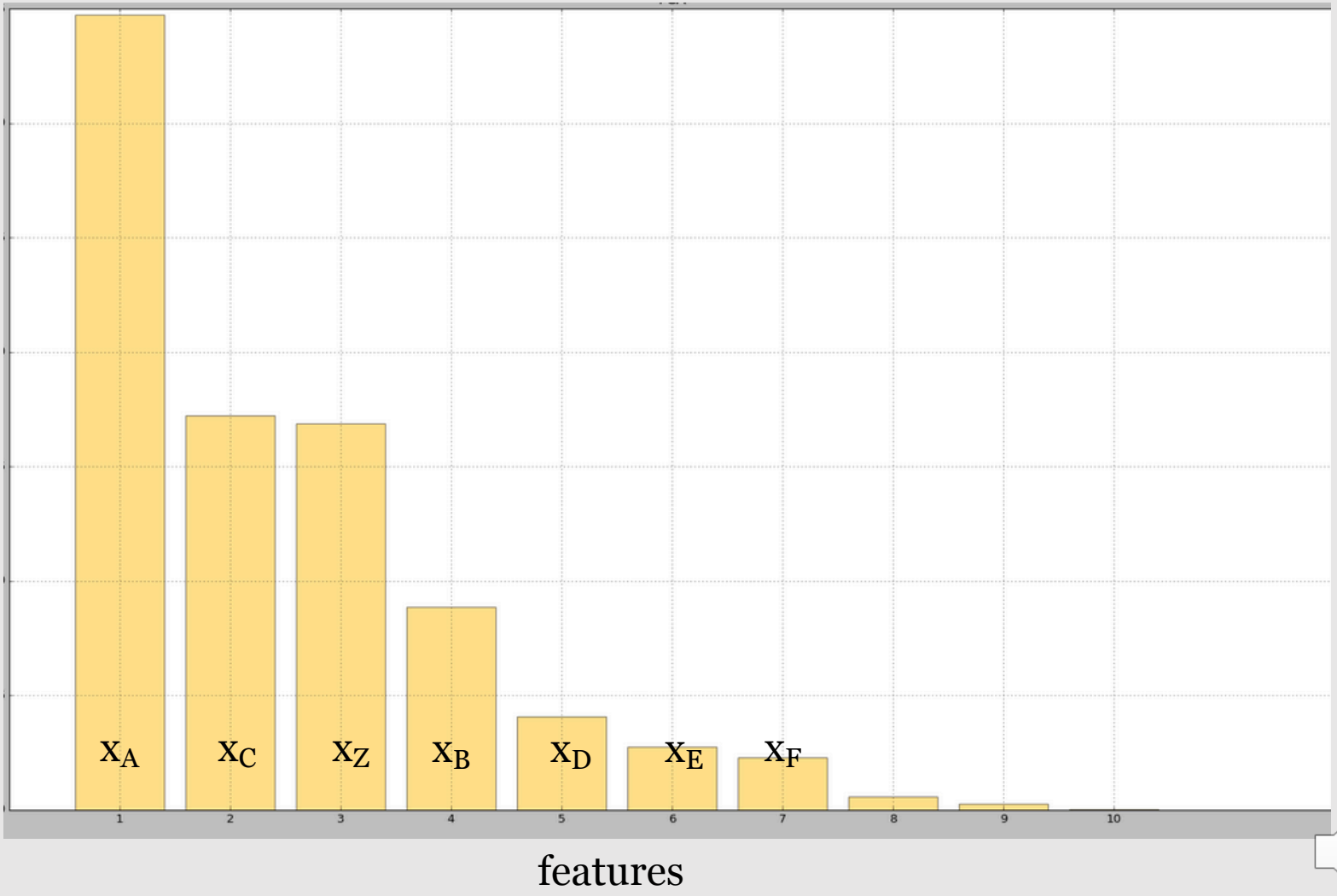


PCA is not linear regression

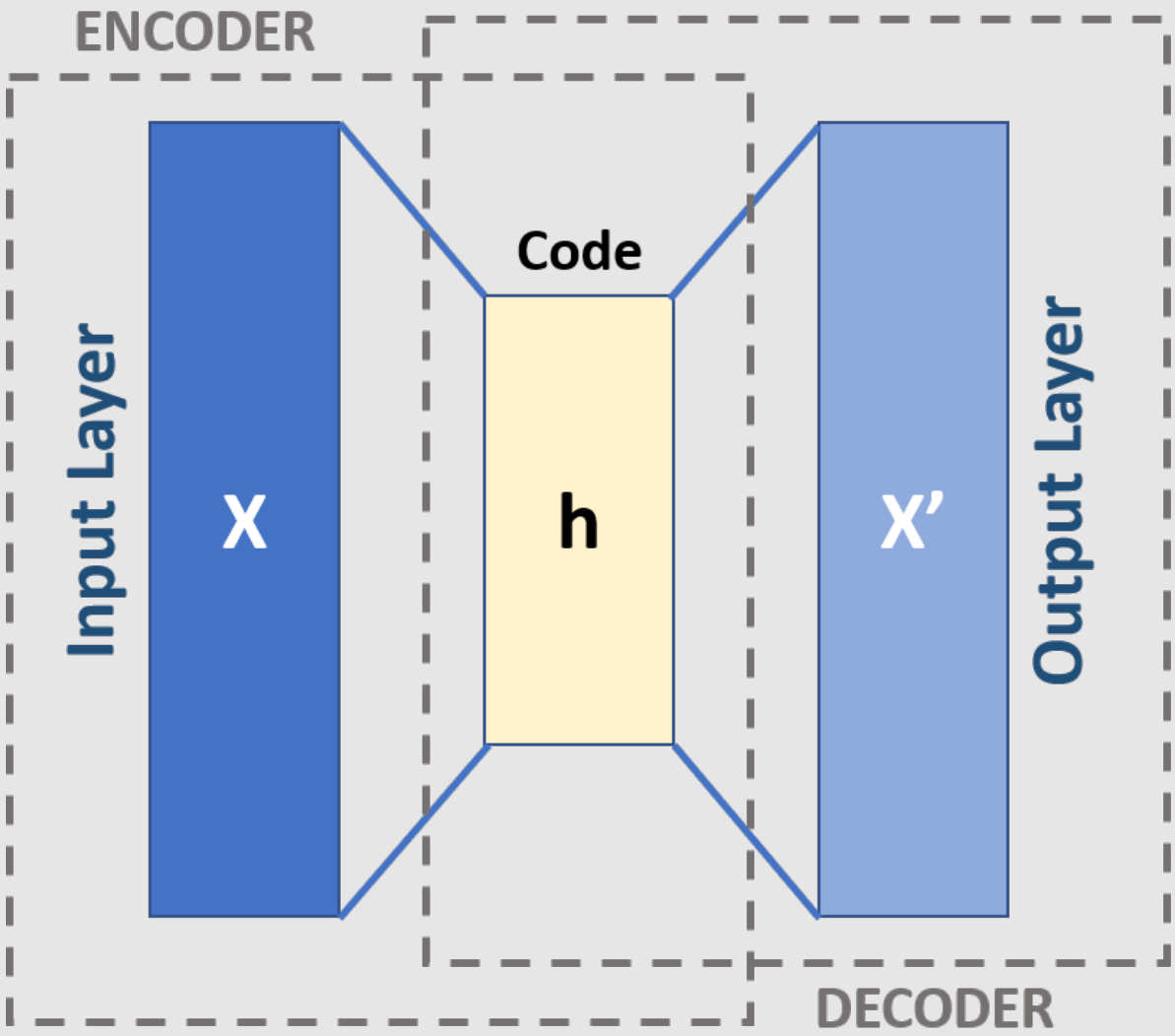
PCA



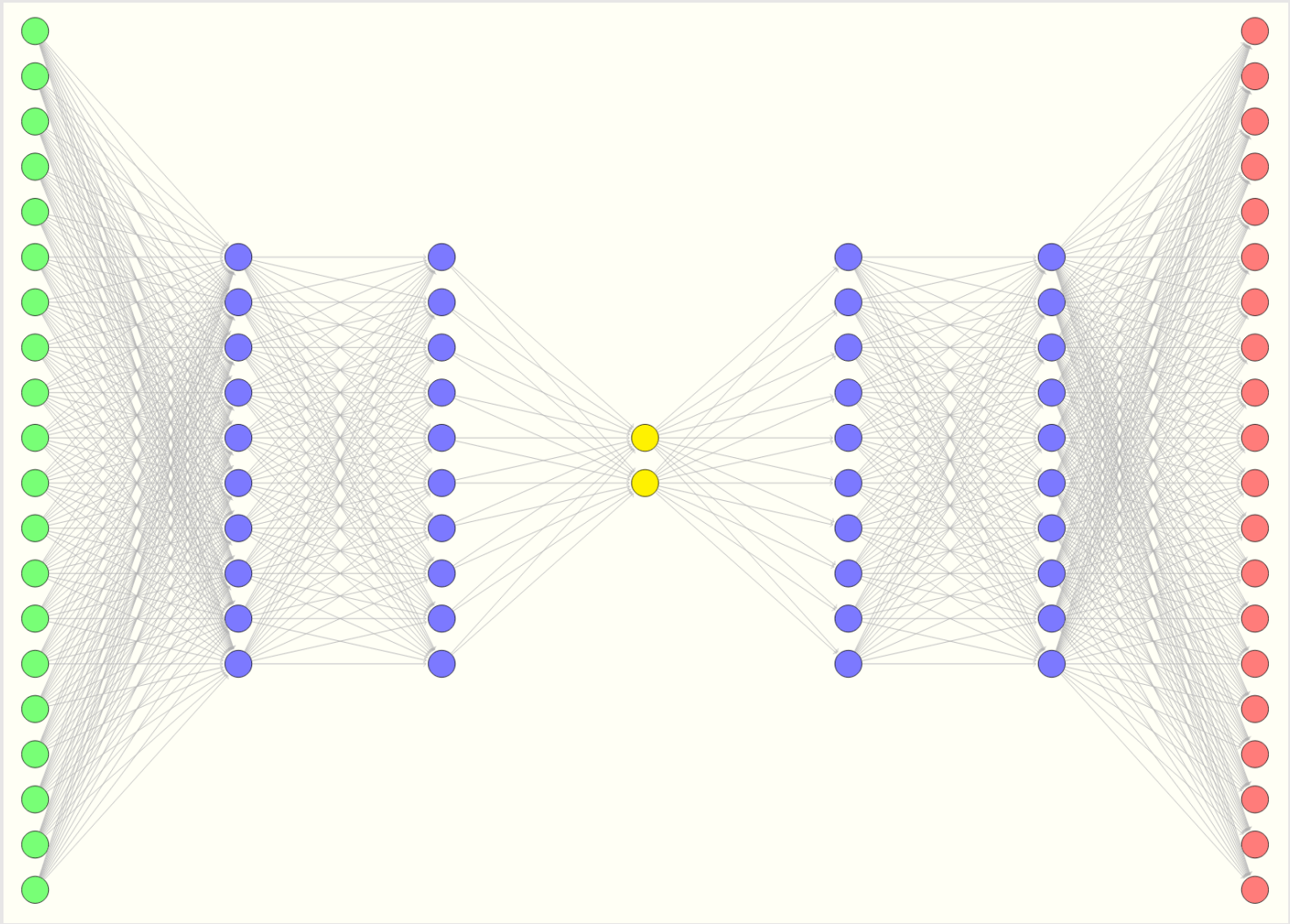
PCA – feature selection



Autoencoder



Autoencoder



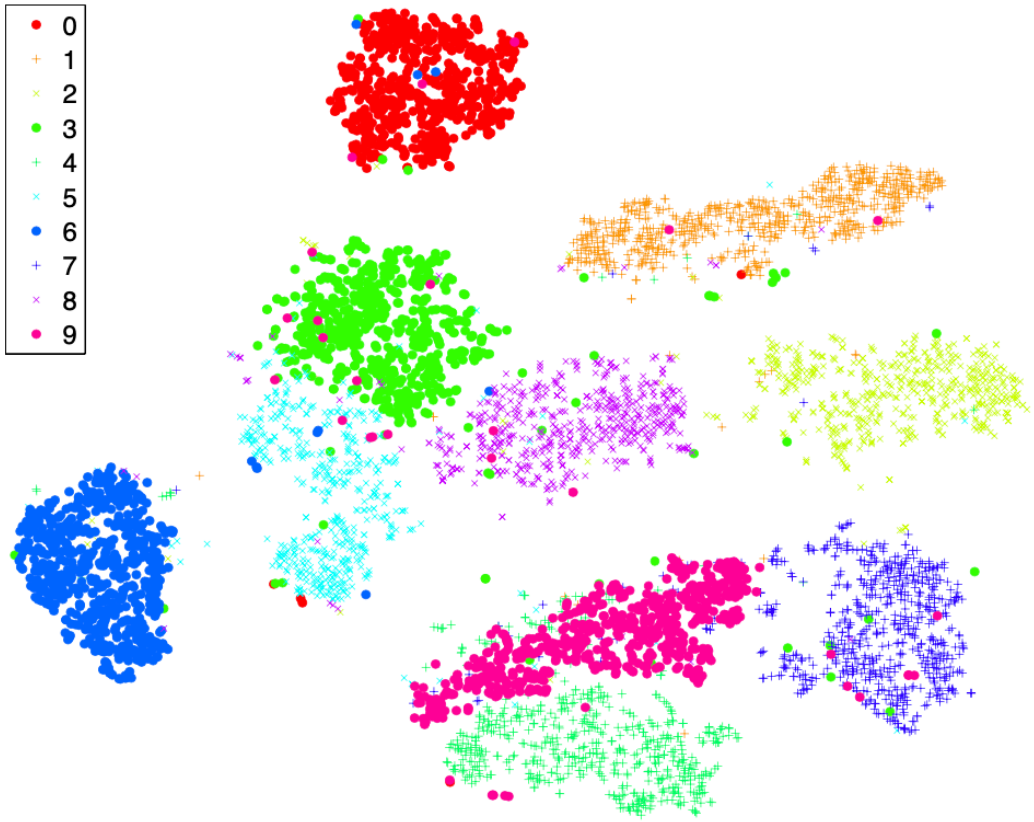
source: https://gertjanvandenburg.com/blog/autoencoder/

Visualizing Data using t-SNE

Laurens van der Maaten LVDMAATEN@GMAIL.COM
TiCC
Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Geoffrey Hinton HINTON@CS.TORONTO.EDU
Department of Computer Science
University of Toronto
6 King's College Road, M5S 3G4 Toronto, ON, Canada

t-SNE



Example

Phenotypic Profiling of High Throughput Imaging Screens with Generic Deep Convolutional Features

Philip T. Jackson[1], Yinhai Wang[2], Sinead Knight[2], Hongming Chen[2],
Thierry Dorval[2], Martin Brown[2], Claus Bendtsen[2], Boguslaw Obara[1]

[1]Department of Computer Science, Durham University [2]IMED Biotech Unit, AstraZeneca

claus.bendtsen@astrazeneca.com, boguslaw.obara@durham.ac.uk

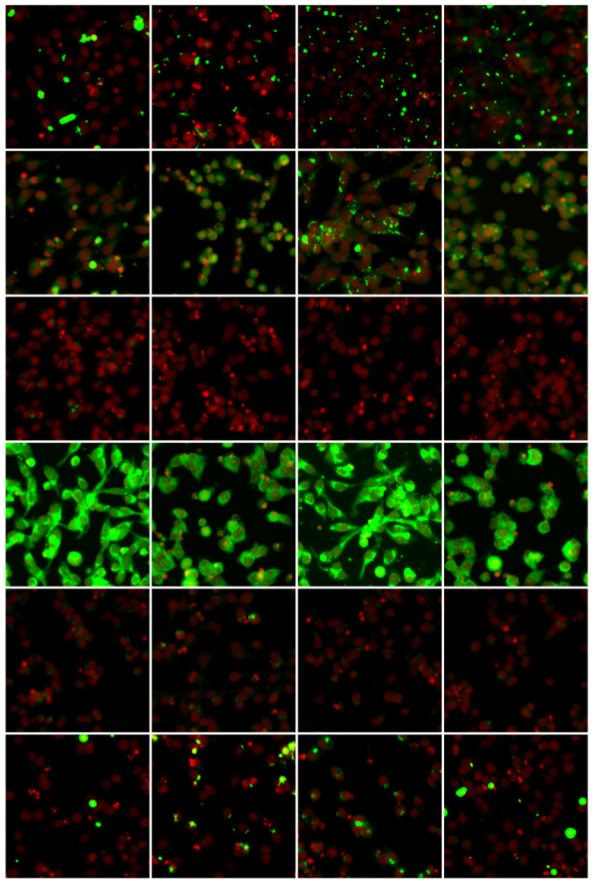


Figure 4. Samples from six of the 70 phenotypic clusters detected by k-means. Each row shows four example images from a single cluster. Rows 2 and 4 show genuine GFP expression.

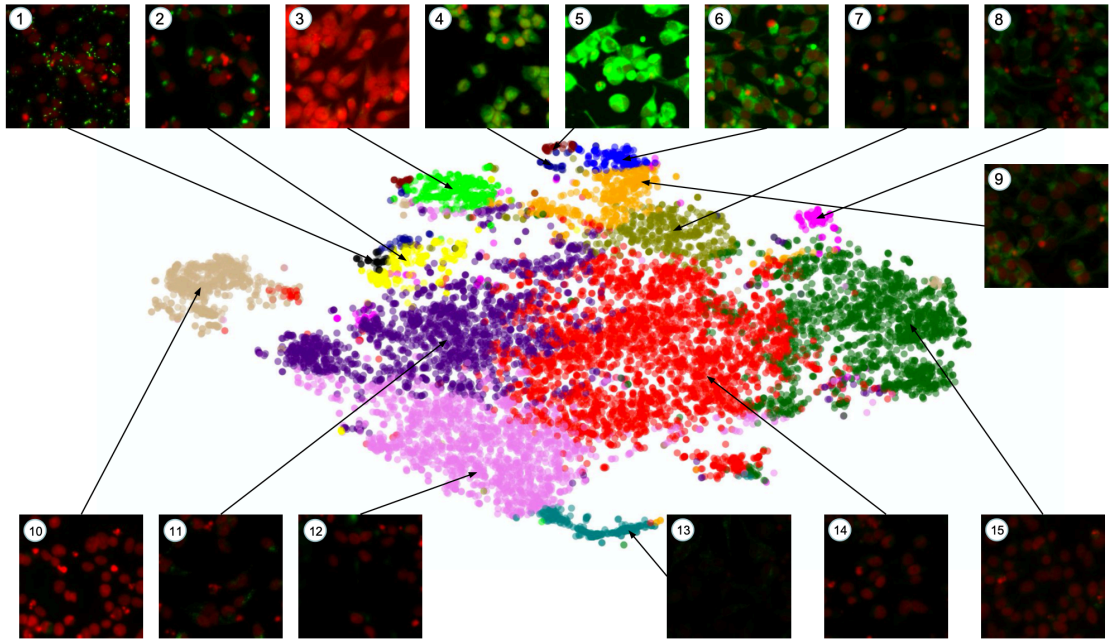


Figure 5. A t-sne embedding of our dataset, with colours showing phenotypic clusters discovered by k-means. For visualization purposes, we set $k = 15$ here.