Machine learning primer

JOANNA BYSZUK & JEREMI OCHAB

DHSI 2024, "DIY COMPUTATIONAL TEXT ANALYSIS WITH R"

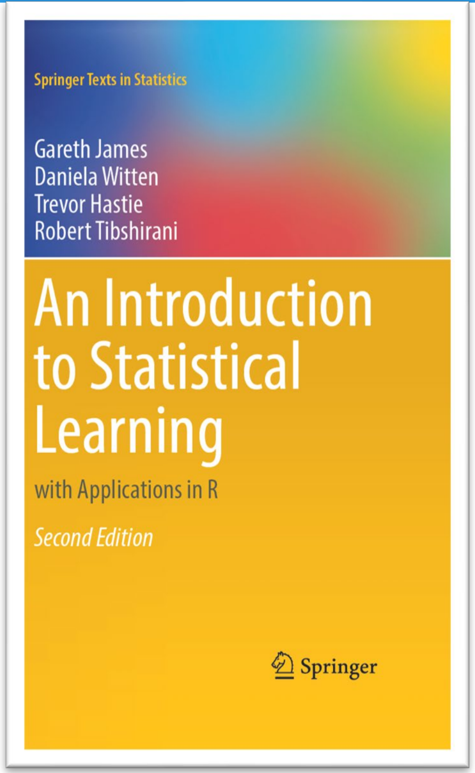
Literature

Literature

AN INTRODUCTION TO STATISTICAL LEARNING

Gareth M. James, Daniela Witten, Trevor Hastie, Robert Tibshirani. 2nd edSpringer (2021). https://www.statlearning.com/

-- (2023).



Springer Texts in Statistics

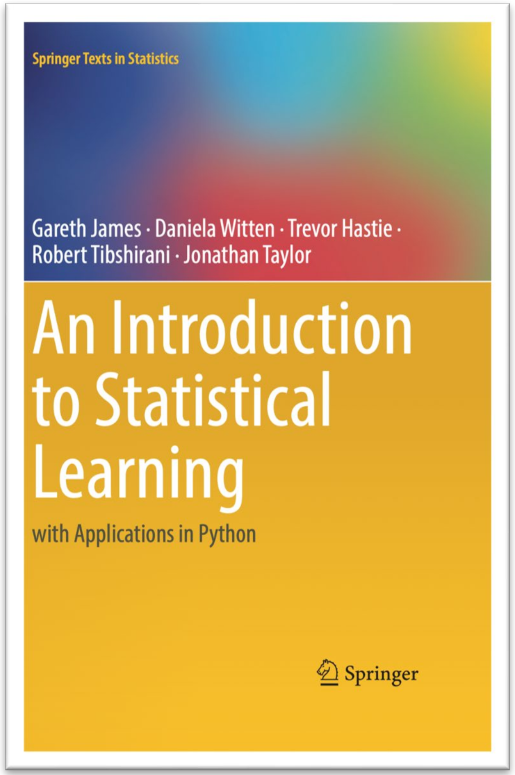
Gareth James
Daniela Witten
Trevor Hastie
Robert Tibshirani

An Introduction to Statistical Learning

with Applications in R

Second Edition

Springer



Springer Texts in Statistics

Gareth James · Daniela Witten · Trevor Hastie · Robert Tibshirani · Jonathan Taylor

An Introduction to Statistical Learning

with Applications in Python

Springer

Online materials

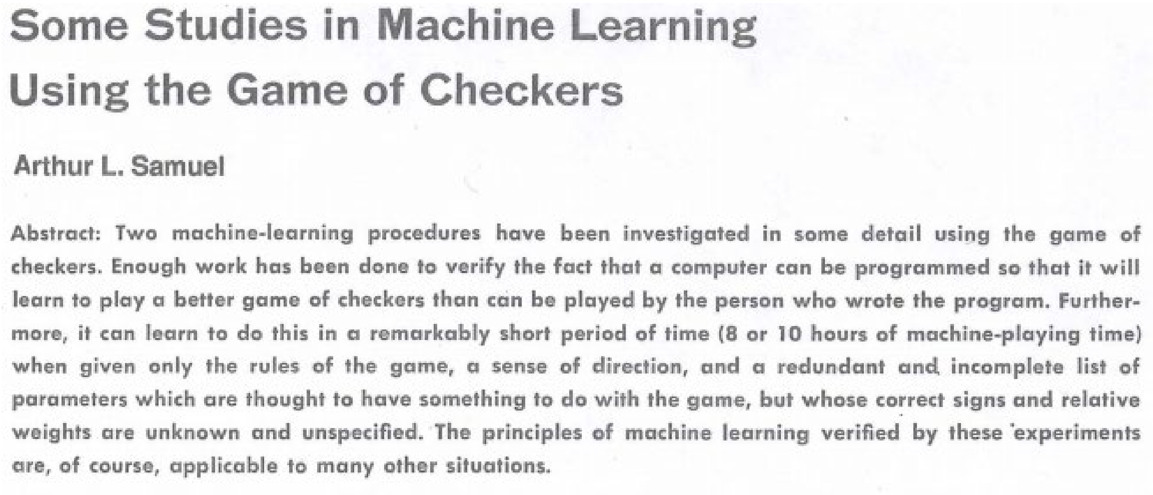
IN ENGLISH

- *scikit-learn* User guide
https://scikit-learn.org/stable/user_guide.html

Machine learning

WHAT IS IT?

What is machine learning?



Some Studies in Machine Learning Using the Game of Checkers

Arthur L. Samuel

Abstract: Two machine-learning procedures have been investigated in some detail using the game of checkers. Enough work has been done to verify the fact that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program. Furthermore, it can learn to do this in a remarkably short period of time (8 or 10 hours of machine-playing time) when given only the rules of the game, a sense of direction, and a redundant and incomplete list of parameters which are thought to have something to do with the game, but whose correct signs and relative weights are unknown and unspecified. The principles of machine learning verified by these experiments are, of course, applicable to many other situations.

- Born from the research on artificial intelligence (1956)
- Considered a separate discipline in the 90s (focus on the statistical approach and practical problems)

IBM 704

What is machine learning?

DEFINITIONS

- “Field of study that gives computers the ability to learn with explicitly being programmed.”
  Arthur Samuel

- The definition of algorithms studied in the field of machine learning: "A computer program learns […] if its performance at tasks in ***T***, as measured by ***P***, improves with experience ***E***.”
  Tom M. Mitchell

- “Machine learning explores the study and construction of algorithms that can learn from and make predictions on data – such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs.”
  Wikipedia

General aims of ML

ALGORITHMS

- more efficient and more accurate algorithms.
- deal with large-scale problems.
- handle a variety of different learning problems.

General aims of ML

THEORETICAL QUESTIONS

- what can be learned, under what conditions?
- are there learning guarantees?
- analysis of learning algorithms.

ALGORITHMS

- more efficient and more accurate algorithms.
- deal with large-scale problems.
- handle a variety of different learning problems.

Types of ML

SCENARIOS

- **batch**: learner receives full (training) sample, which he uses to make predictions for unseen points.
- **on-line**: learner receives one sample at a time and makes a prediction for that sample.

QUERIES

- **active**: the learner can request the label of a point.
- **passive**: the learner receives labeled points.

Types of ML

SCENARIOS

- **batch**: learner receives full (training) sample, which he uses to make predictions for unseen points.
- **on-line**: learner receives one sample at a time and makes a prediction for that sample.

PREDOMINANT BATCH SCENARIOS

- **Unsupervised learning**: no labeled data.
- **Supervised learning**: uses labeled data for prediction on unseen points.
- **Semi-supervised learning**: uses labeled and unlabeled data for prediction on unseen points.
- **Transduction**: uses labeled and unlabeled data for prediction on seen points.
- **Reinforcement learning**: training data (in form of rewards and punishments) is given only as feedback to the program's actions in a dynamic environment, such as driving a vehicle or playing a game against an opponent.

Types of ML

EXAMPLES OF SPECIFIC TASKS

- Text: document classification, spam detection.
- Language: NLP tasks (e.g., morphological analysis, POS tagging, context-free parsing, dependency parsing).
- Speech: recognition, synthesis, verification.
- Image: annotation, face recognition, OCR, handwriting recognition.
- Games (e.g., chess, backgammon, go).
- Unassisted control of vehicles (robots, car).
- Medical diagnosis, fraud detection, network intrusion.

EXAMPLES OF GENERAL TASKS

Types of ML

EXAMPLES OF SPECIFIC TASKS

- Text: document classification, spam detection.
- Language: NLP tasks (e.g., morphological analysis, POS tagging, context-free parsing, dependency parsing).
- Speech: recognition, synthesis, verification.
- Image: annotation, face recognition, OCR, handwriting recognition.
- Games (e.g., chess, backgammon, go).
- Unassisted control of vehicles (robots, car).
- Medical diagnosis, fraud detection, network intrusion.

EXAMPLES OF GENERAL TASKS

Classification: assign a category to each item (e.g., document classification).

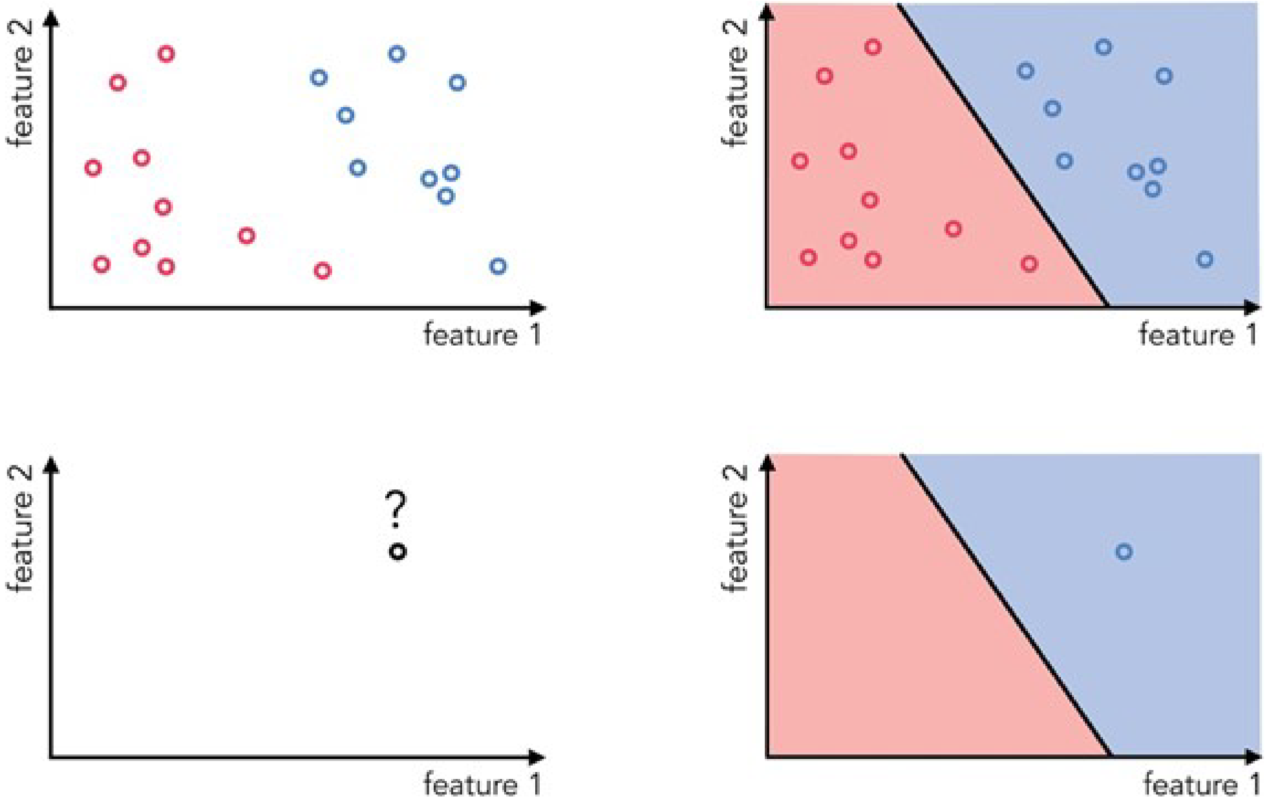
Regression: „predict” a real value for each item (e.g., prediction of stock values, economic variables).

Ranking: order items according to some criterion (e.g., relevant web pages returned by a search engine).

Clustering: group data into „homogenous” regions (analysis of very large data sets).

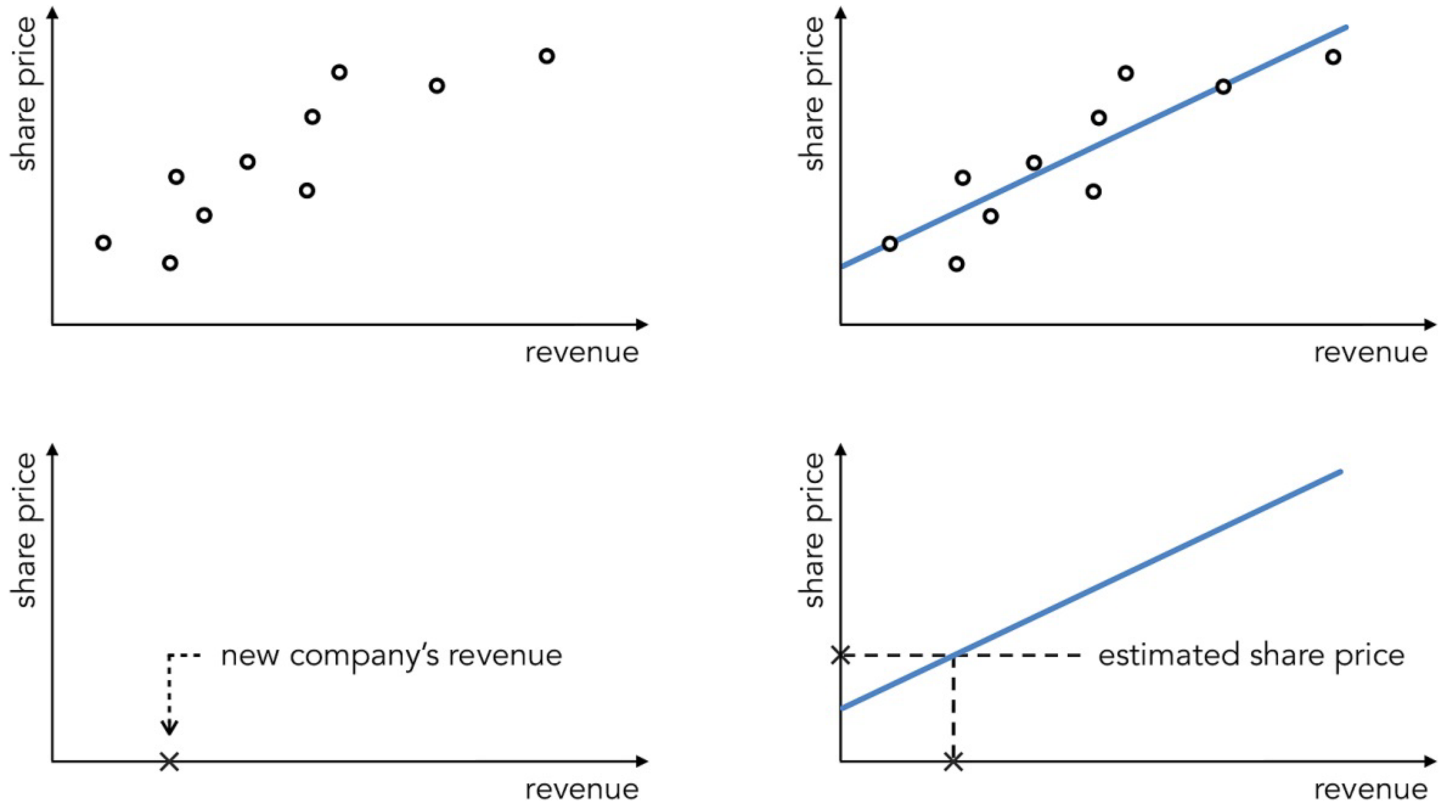
Dimensionality reduction: find a lower-dimensional space preserving some properties of the data.

Task examples: classification

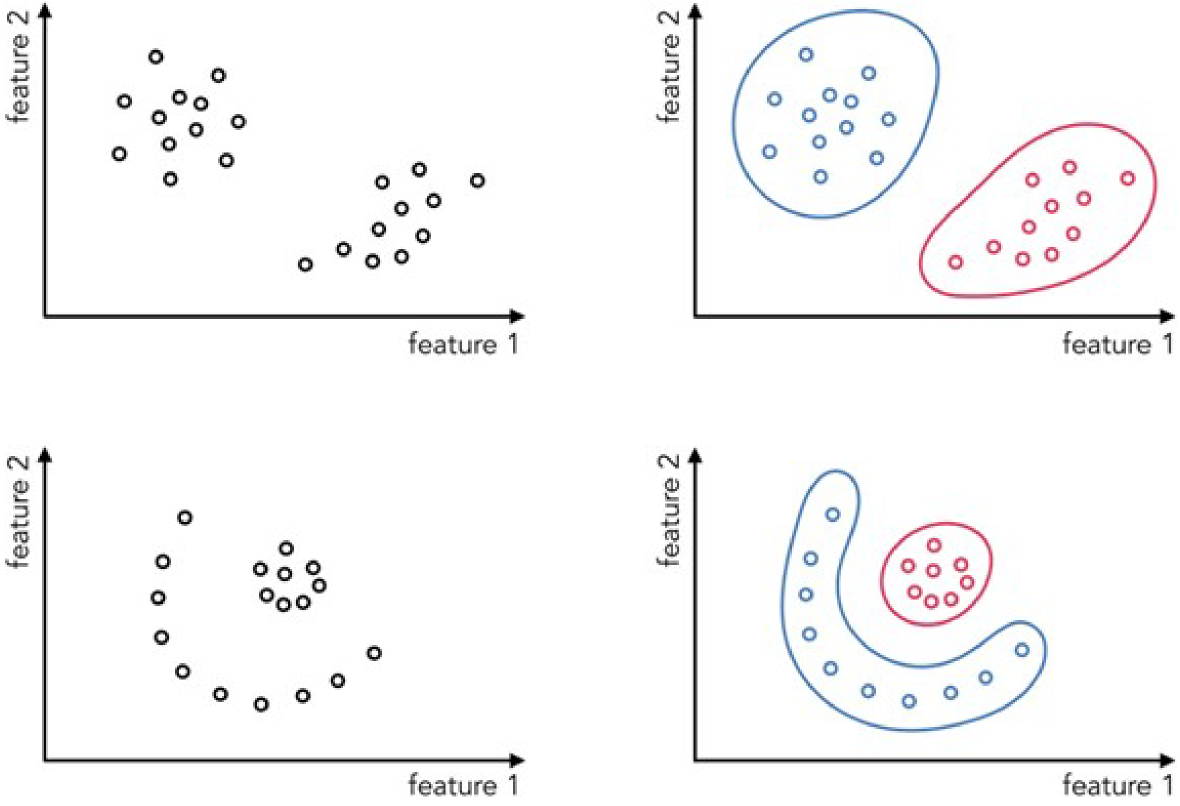


J. Watt, R. Borhani, A.K. Katsaggelos. *Machine Learning Refined*. © Cambridge University Press 2020

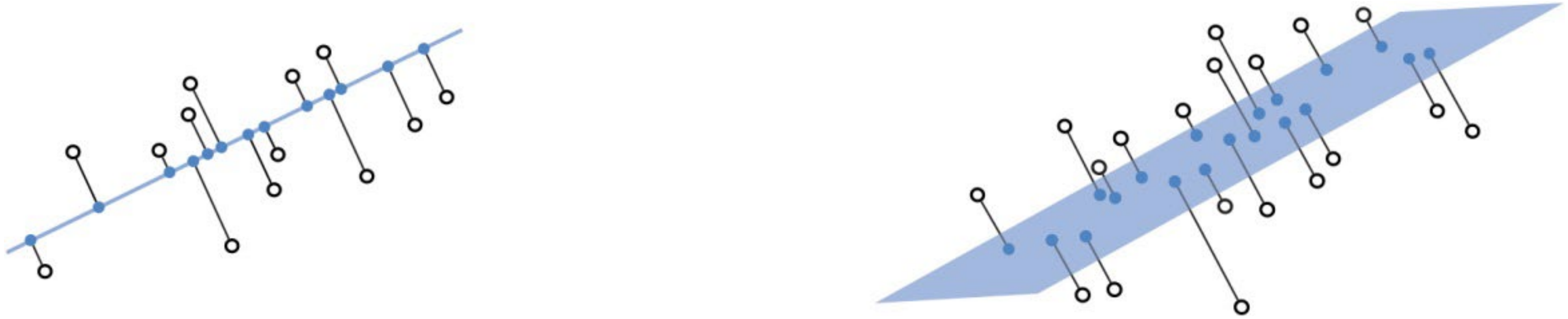
Task examples: regression



Task examples: clustering



Task examples: dimensionality reduction



Dimensionality reduction: PCA

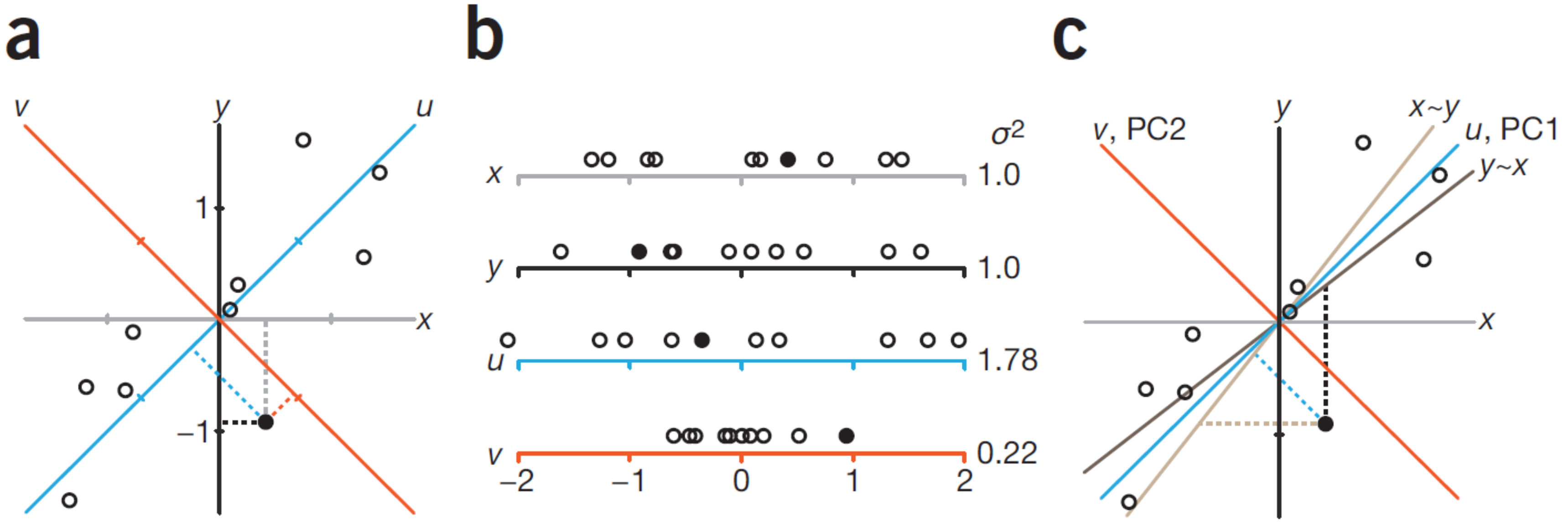


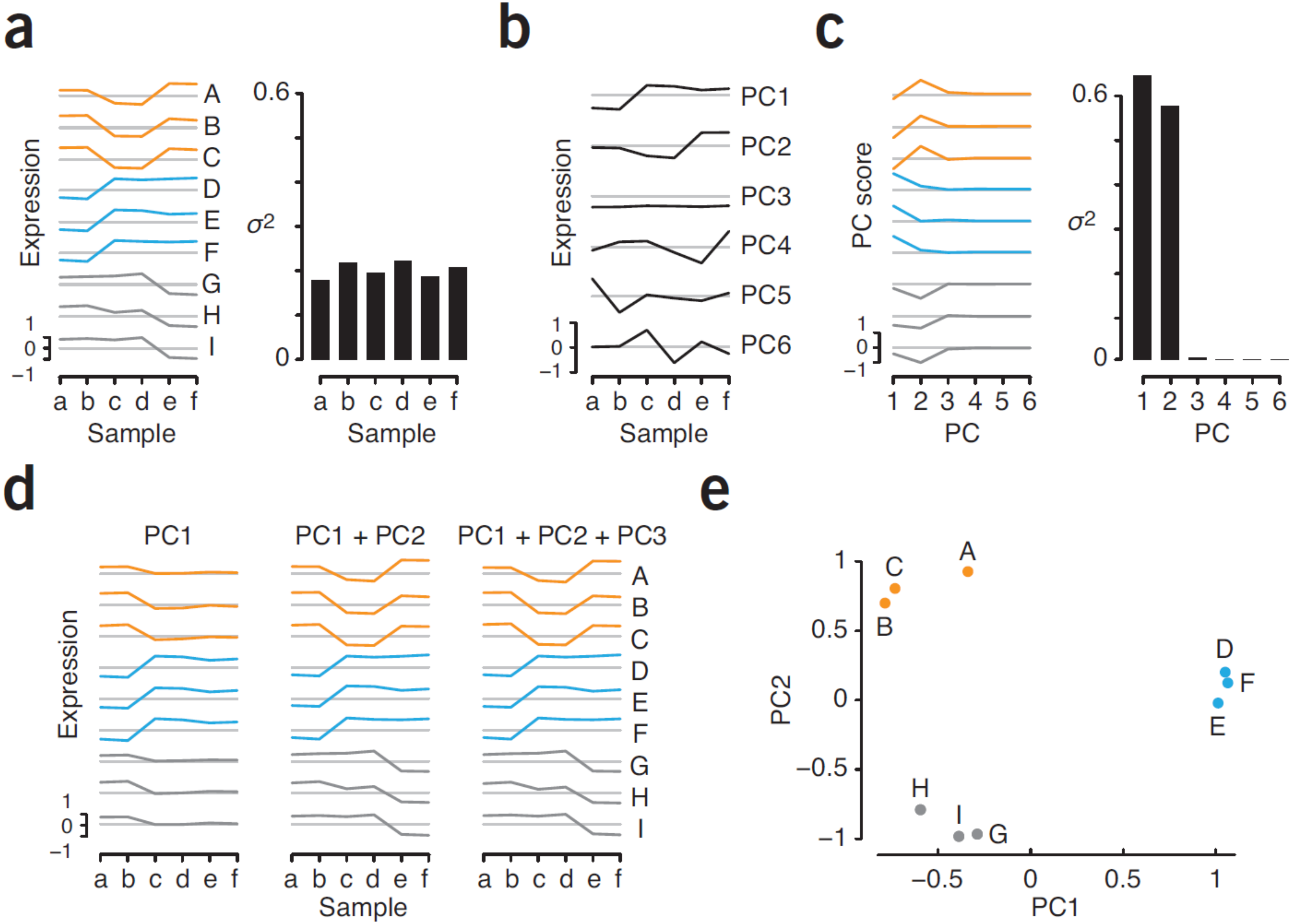
Figure 1 | PCA geometrically projects data onto a lower-dimensional space. (**a**) Projection is illustrated with 2D points projected onto 1D lines along a path perpendicular to the line (illustrated for the solid circle). (**b**) The projections of points in **a** onto each line. σ^2 for projected points can vary (e.g., high for *u* and low for *v*). (**c**) PC1 maximizes the σ^2 of the projection and is the line *u* from **a**. The second (*v*, PC2) is perpendicular to PC1. Note

Lever, J., Krzywinski, M., Altman, N., 2017. Principal component analysis. Nature Methods 14, 641–642.

Dimensionality reduction: PCA

- What do the components mean?
- How many components should I look at?
- …your questions?

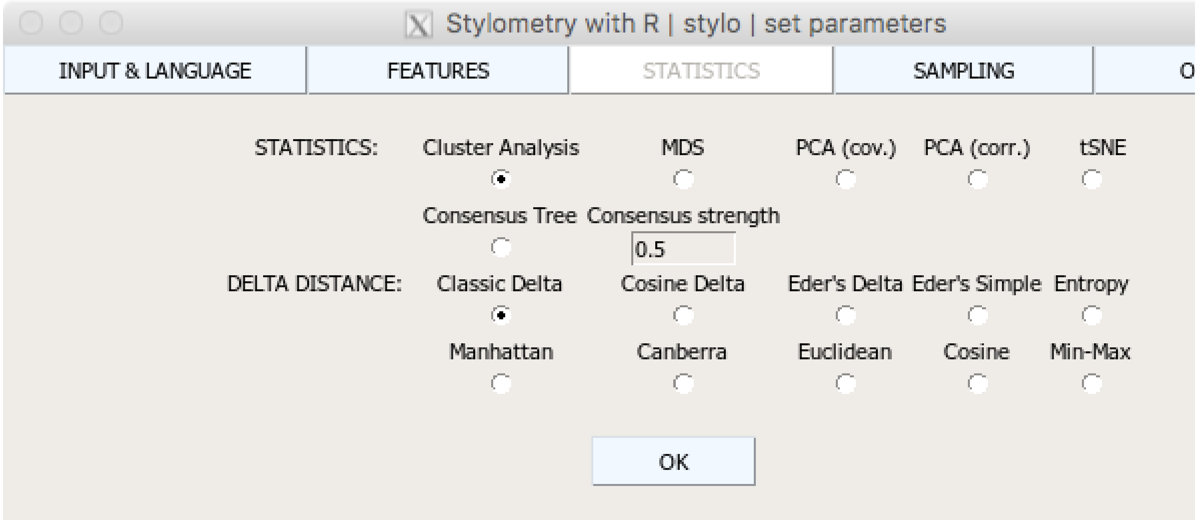
Lever, J., Krzywinski, M., Altman, N., 2017. Principal component analysis. Nature Methods 14, 641–642.



Lever, J., Krzywinski, M., Altman, N., 2017. Principal component analysis. Nature Methods 14, 641–642.

PCA: watch out for...

- **PCA (cov.)/PCA (corr.)? – scale!**

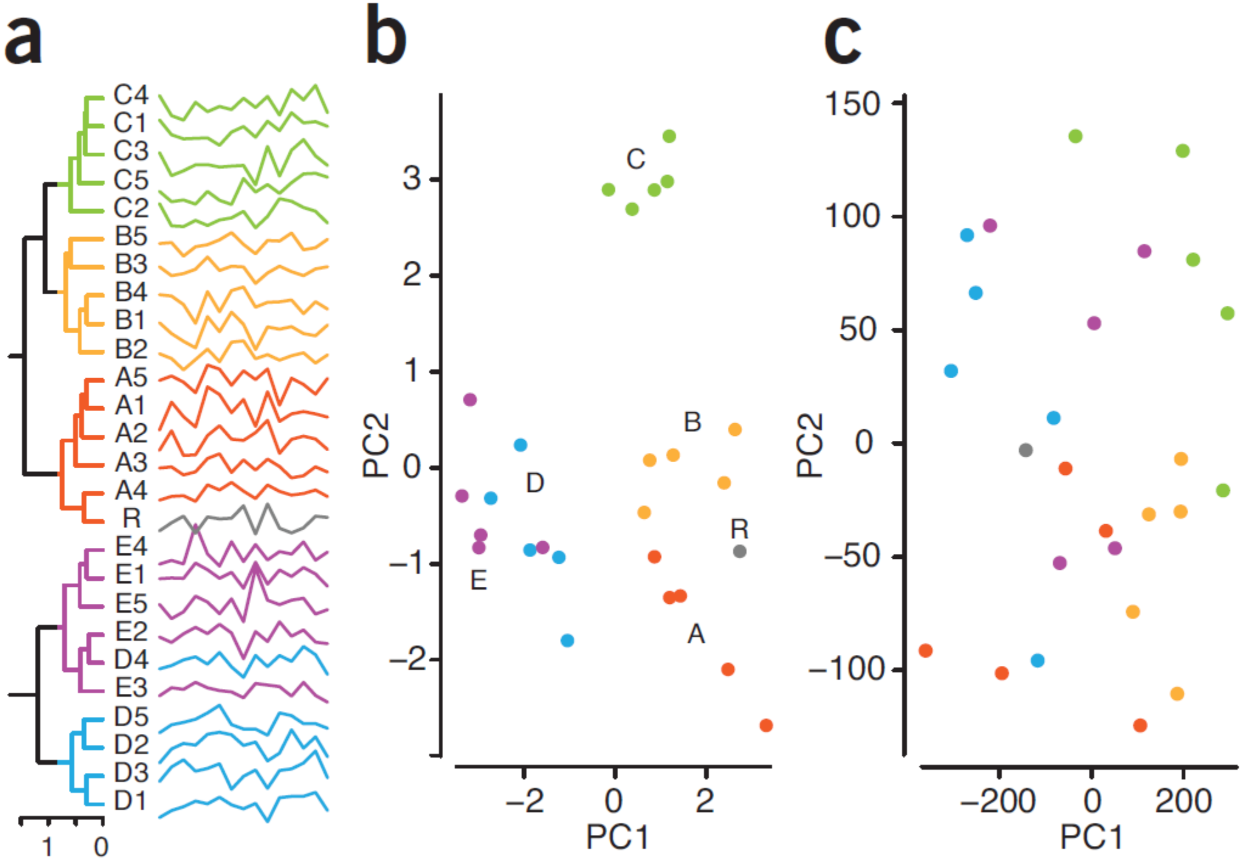


STATISTICS: ● Cluster Analysis ○ MDS ○ PCA (cov.) ○ PCA (corr.) ○ tSNE

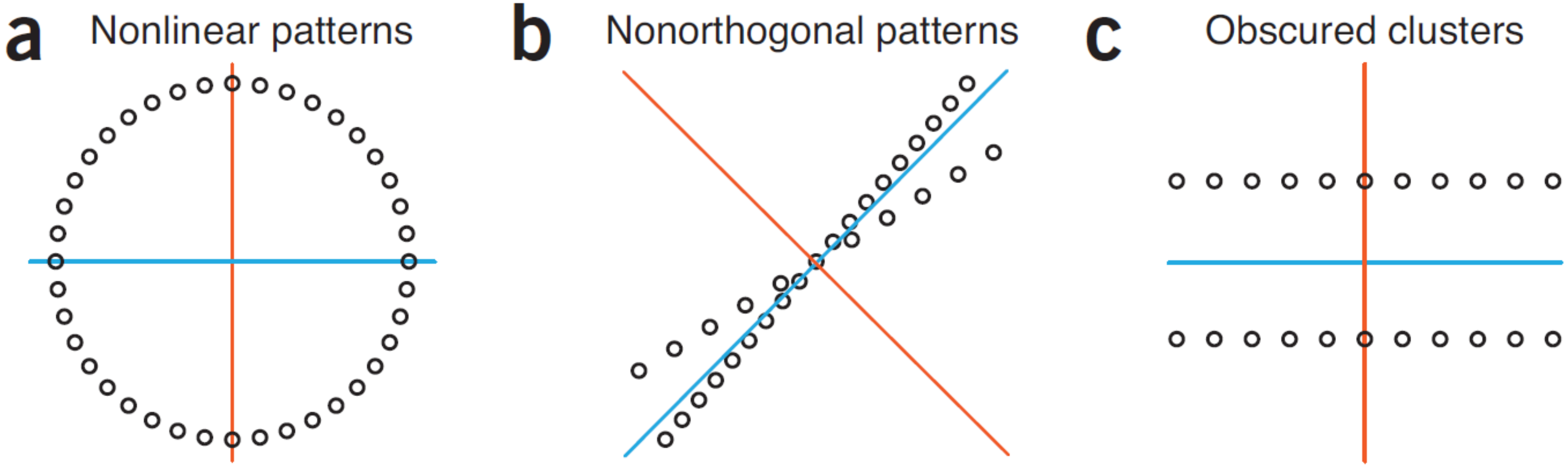
○ Consensus Tree Consensus strength 0.5

DELTA DISTANCE: ● Classic Delta ○ Cosine Delta ○ Eder's Delta ○ Eder's Simple ○ Entropy

○ Manhattan ○ Canberra ○ Euclidean ○ Cosine ○ Min-Max



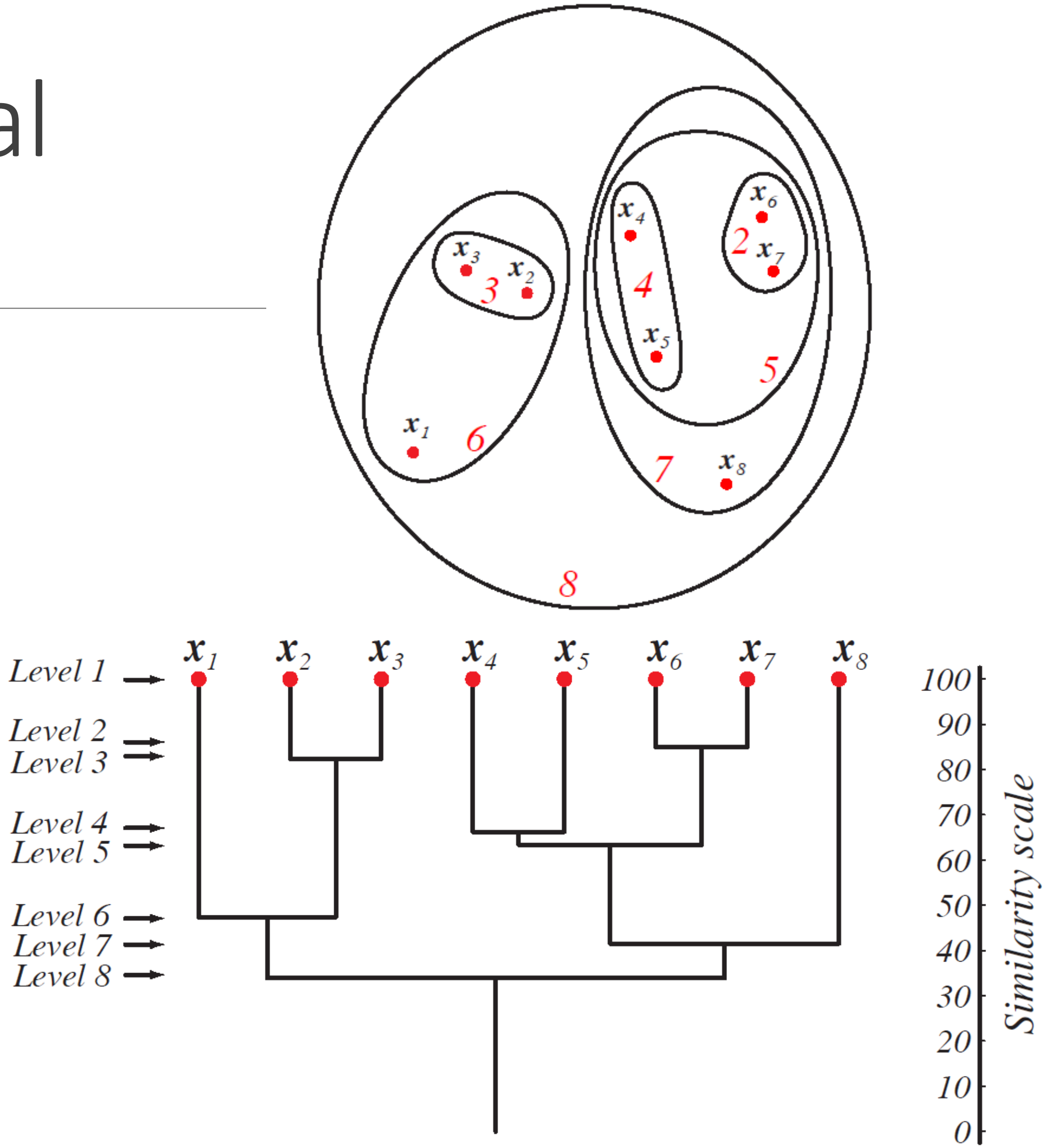
PCA: watch out for…



Clustering: hierarchical

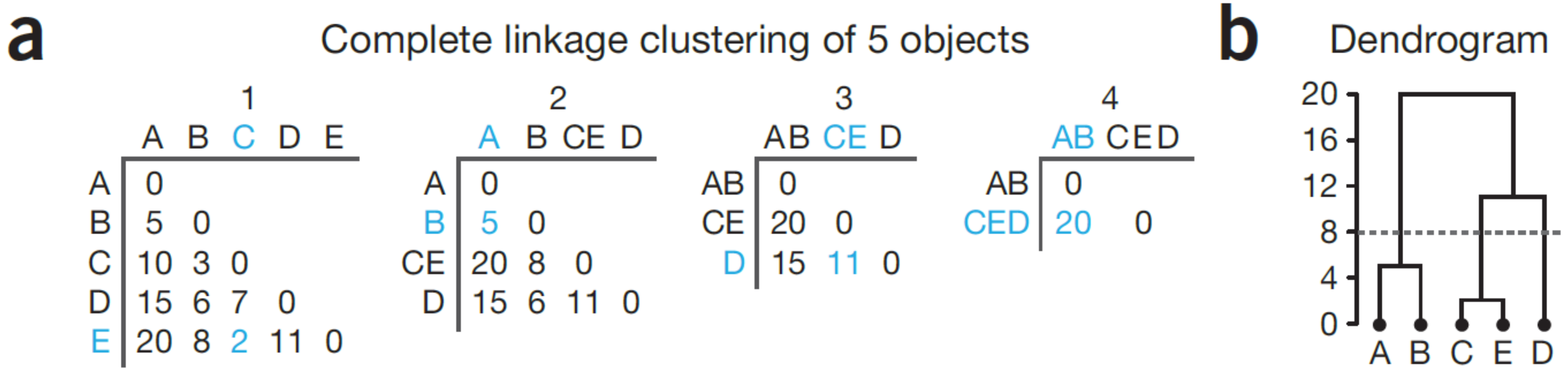
- **Type**:
  agglomerative
  divisive.

- order of branches is arbitrary (horizontal direction in the dendrogram here has no meaning)



Richard O. Duda, Peter E. Hart, David G. Stork, *Pattern classification*, Wiley (2001), Chapter 10.9.

Clustering: hierarchical



1

	A	B	C	D	E
A	0				
B	5	0			
C	10	3	0		
D	15	6	7	0	
E	20	8	2	11	0

2

	A	B	CE	D
A	0			
B	5	0		
CE	20	8	0	
D	15	6	11	0

3

	AB	CE	D
AB	0		
CE	20	0	
D	15	11	0

4

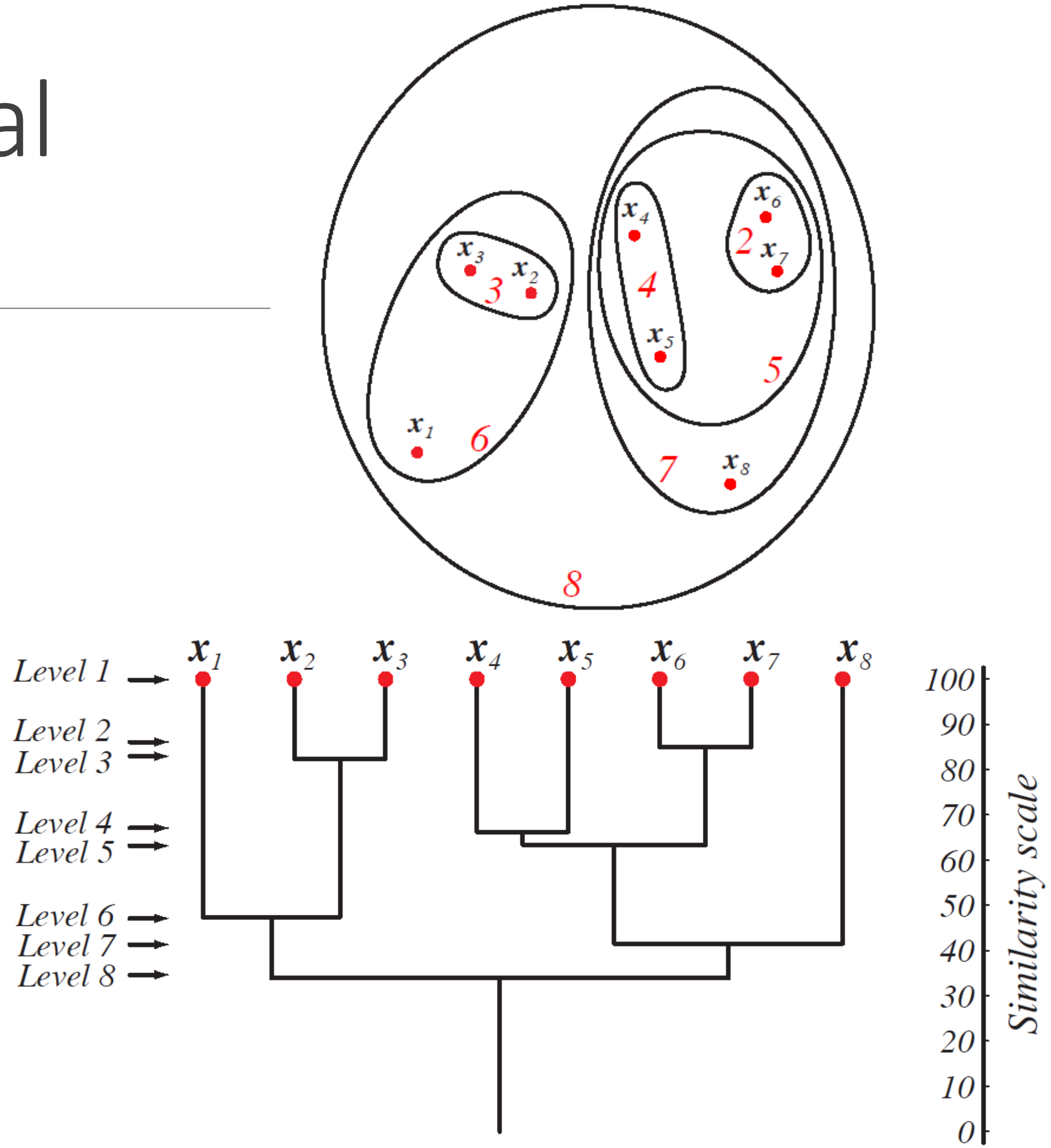
	AB	CED
AB	0	
CED	20	0

Figure 2 | Complete linkage clustering of five objects. (**a**) Pairwise distances (step 1) are used to merge objects (steps 2–4) where the maximum of all pairwise distances is used. At each merging step, the shortest distance is chosen (blue). (**b**) A dendrogram with a vertical axis showing the distance between merged nodes. To create clusters, one can cut the tree at a fixed height (dashed line).

Altman, N., Krzywinski, M., 2017. Clustering. Nature Methods 14, 545–546. https://doi.org/10/gcx55n

Clustering: hierarchical

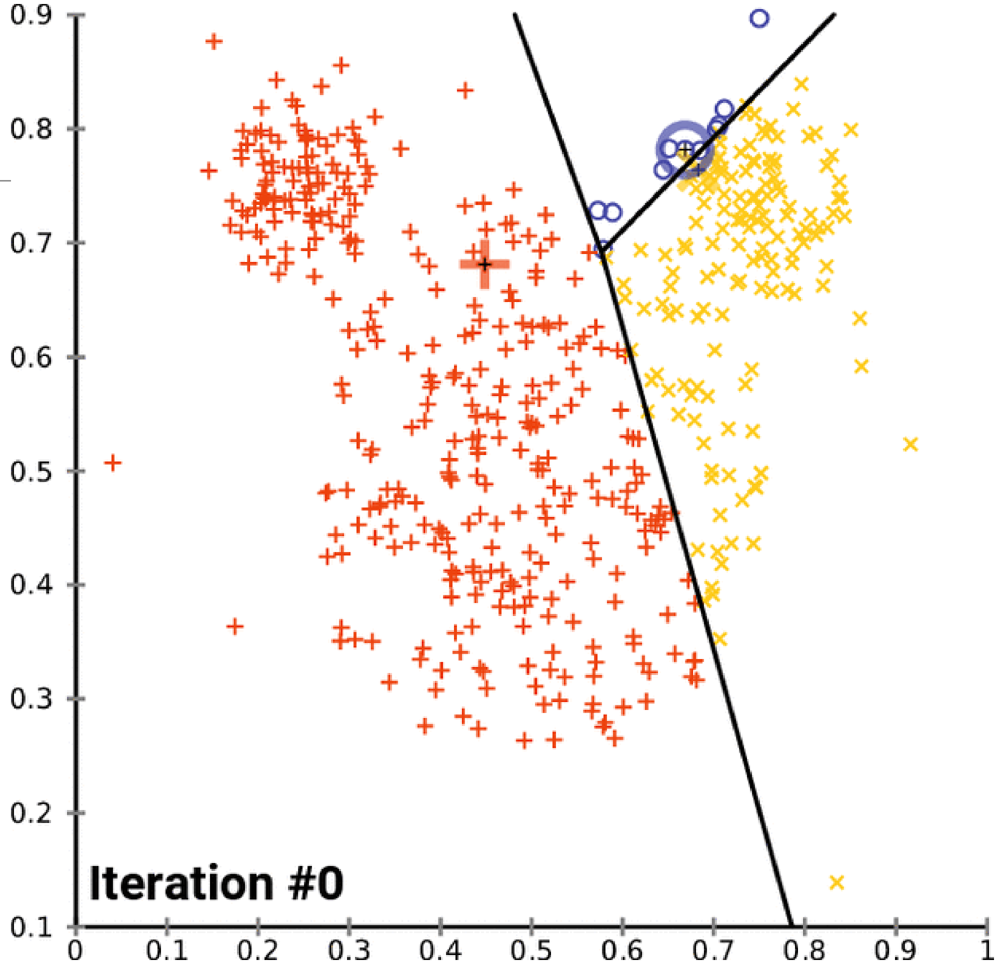
- **Type**:
  agglomerative
  divisive.
- „**Linkage**”:
  single (nearest neighbour),
  complete (farthest neighbour),
  average,
  etc.
- **Distance:**
  Euclidean



Richard O. Duda, Peter E. Hart, David G. Stork, *Pattern classification*, Wiley (2001), Chapter 10.9.

Clustering: k-means

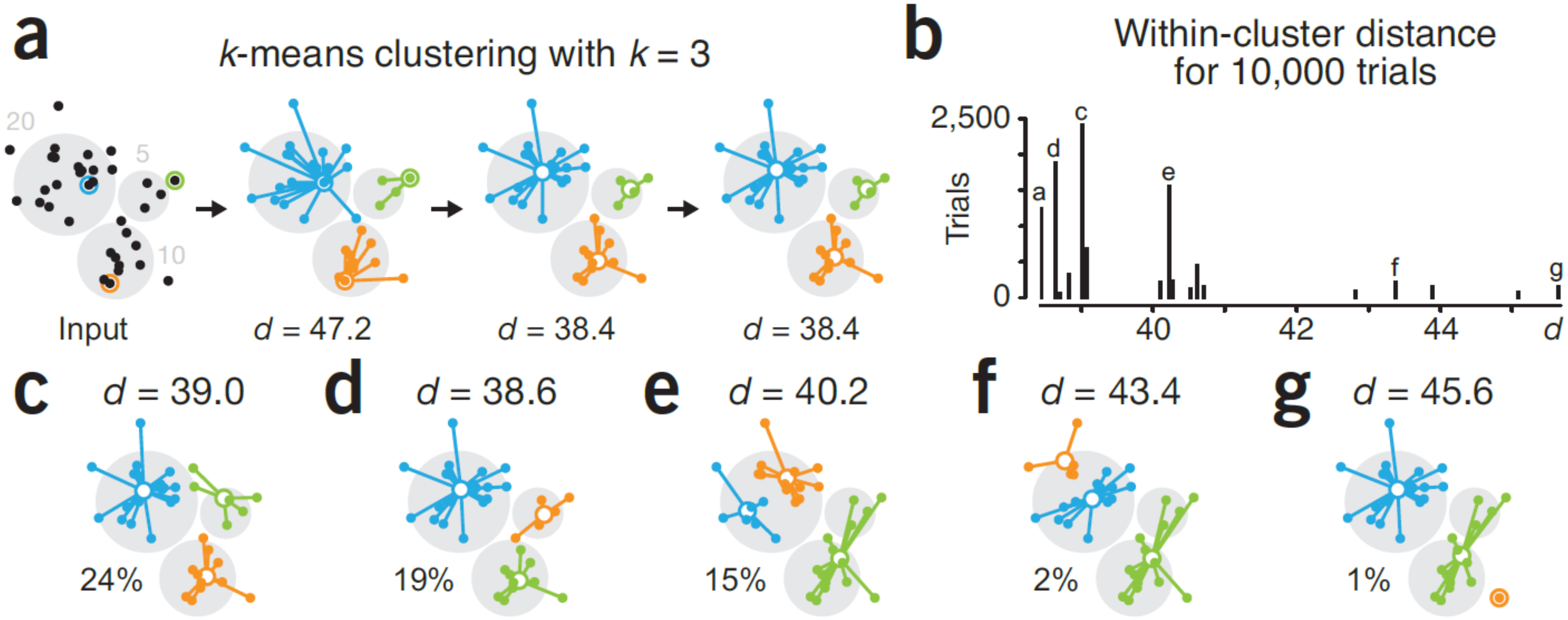
- **choose *k*** beforehand (e.g. „elbow method”).
- compare stability over random initialisations, etc.
- rate solutions: inter-/intra-cluster similarity
- clustering methods always find clusters, even if there are no natural clusters in the data.



https://en.wikipedia.org/wiki/File:K-means_convergence.gif

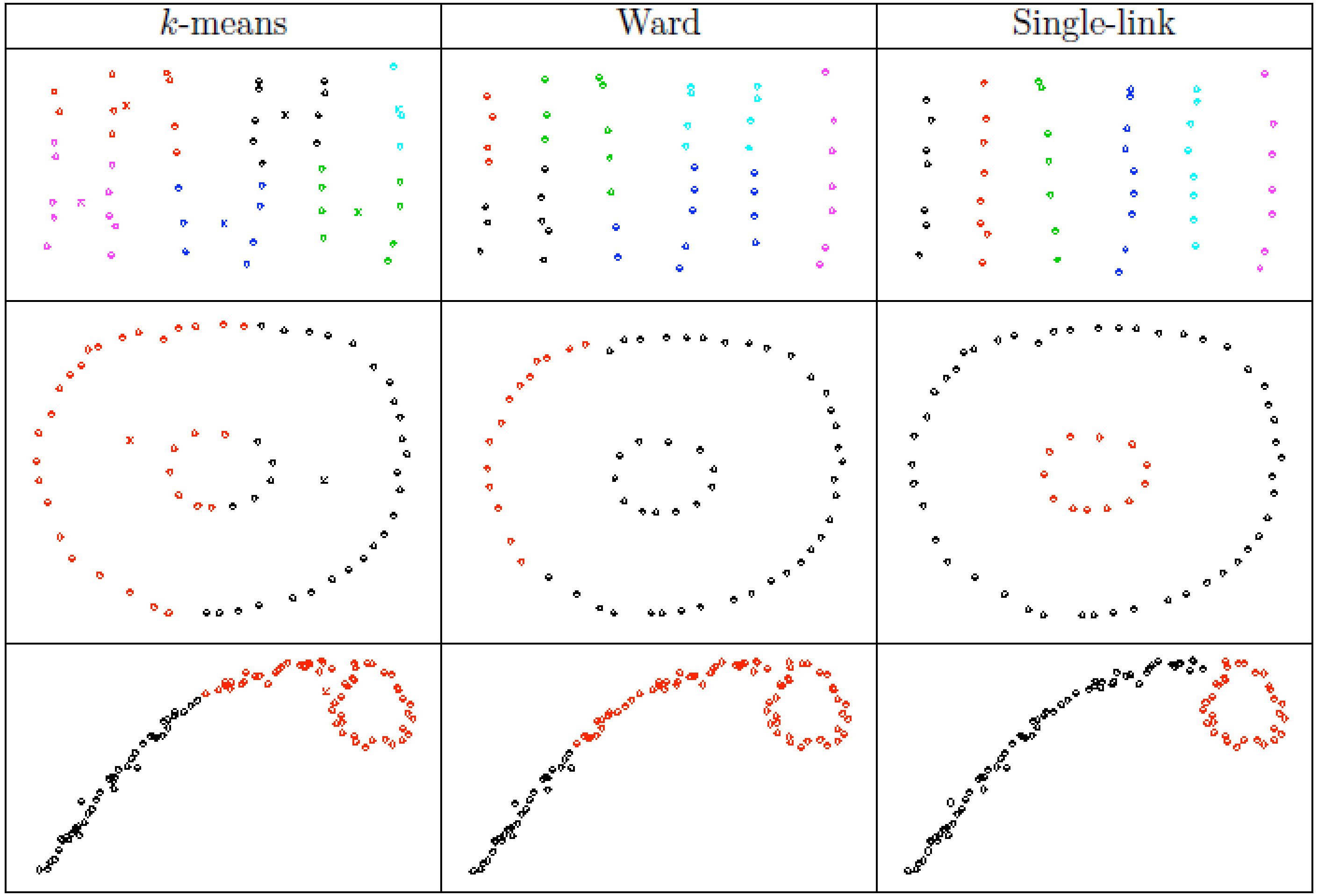
Clustering: k-means

- compare stability over random initialisations, etc.
  - » repeat, repeat, repeat…



Altman, N., Krzywinski, M., 2017. Clustering. Nature Methods 14, 545–546. https://doi.org/10/gcx55n

Clustering: examples



Cosma Shalizi, Carnegie Mellon : http://www.stat.cmu.edu/~cshalizi/350/lectures/08/lecture-08.pdf

Clustering: examples

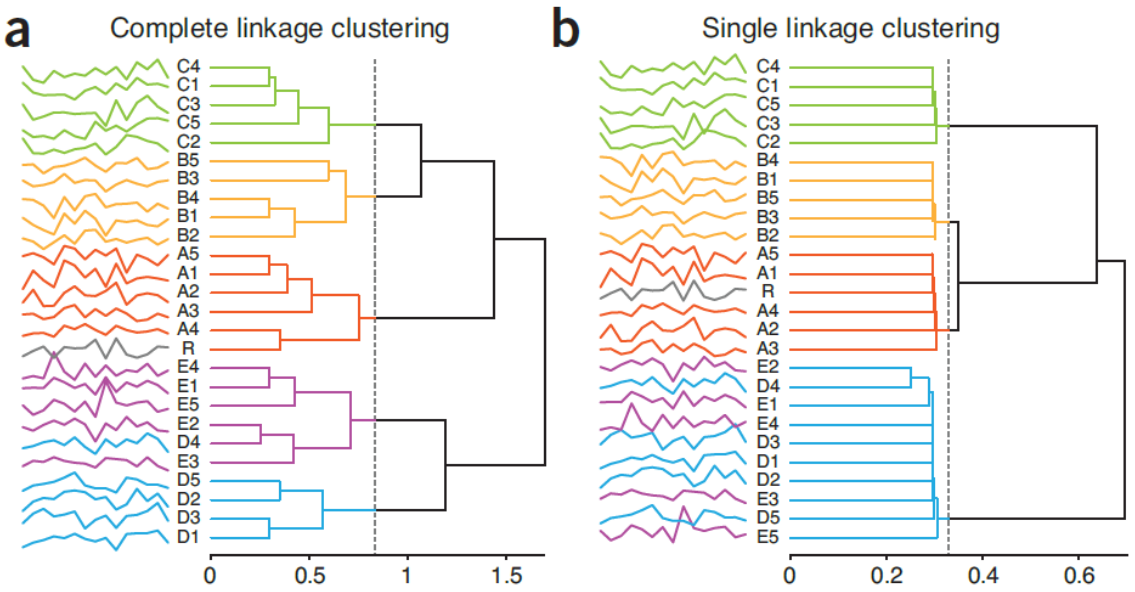
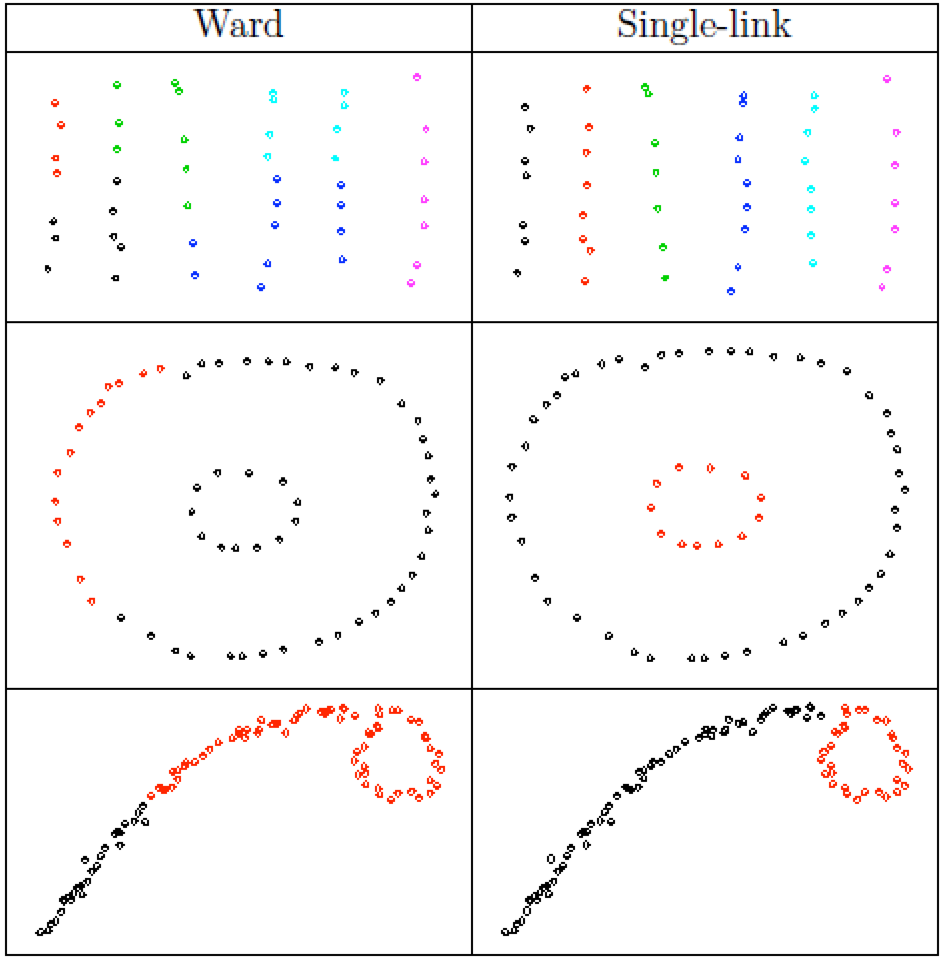
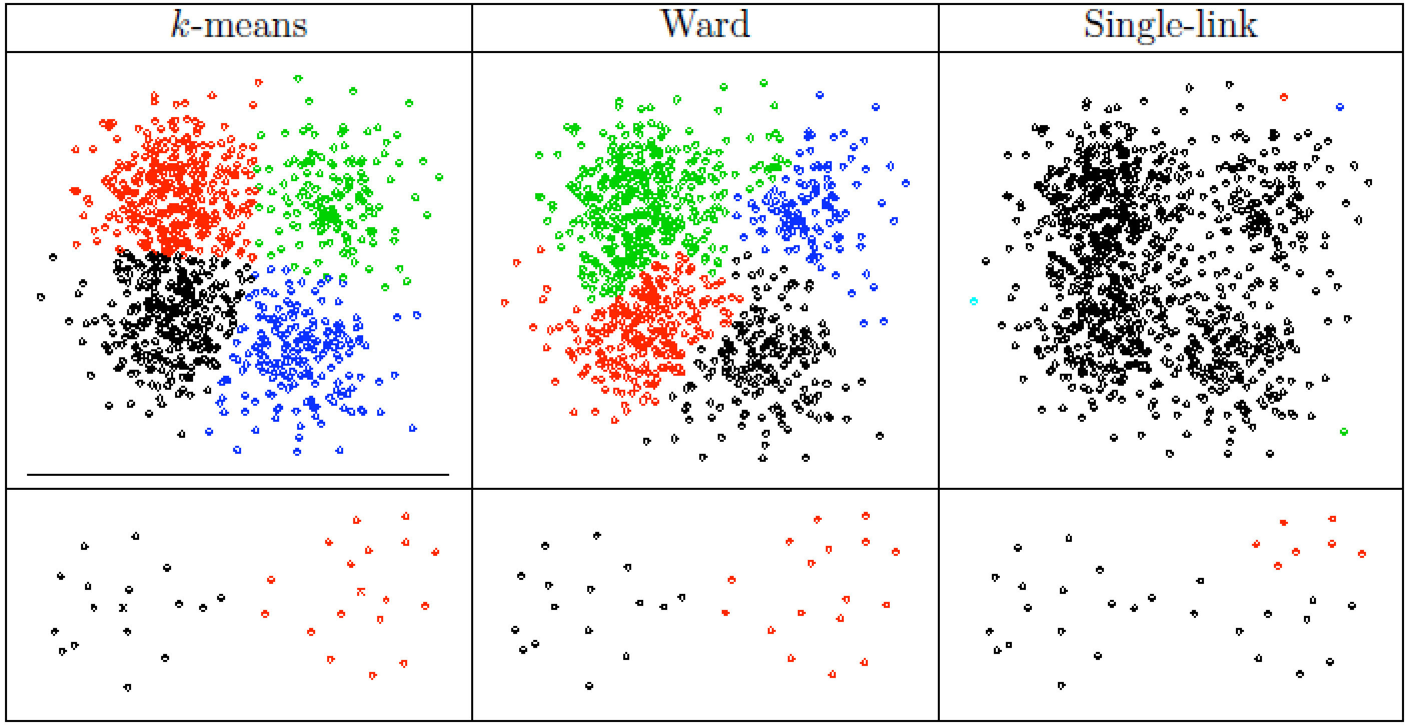


Figure 3 | Dendrograms of hierarchical clustering of gene expression profiles based on correlation distance. The data were generated by creating core profiles A1, B1, C1, D1, and E1 with correlation values of 0.7, 0.5, 0, -0.5, and -0.7 (respectively) with the reference profile R from **Figure 1.** For each core profile (e.g., A1), four additional highly correlated random profiles were generated (e.g., A2–A5). Profiles are colored by group and clusters formed by cutting at a fixed height (dashed line). (**a**) Complete linkage clustering tends to create balanced dendrograms by first clustering objects into small nodes and then clustering the nodes. (**b**) Single linkage clustering tends to create stringy dendrograms by first creating a few nodes and then adding objects to them one at a time.



Altman, N., Krzywinski, M., 2017. Clustering. Nature Methods 14, 545–546. https://doi.org/10/gcx55n

Clustering: examples



No free lunch theorem!!

Cosma Shalizi, Carnegie Mellon : http://www.stat.cmu.edu/~cshalizi/350/lectures/08/lecture-08.pdf