$$X = \{ x^{\{1\}}, x^{\{2\}}, \dots, x^{\{d\}} \}$$

$$X_1 , X_2 \rightsquigarrow d(X_1, X_2) = \sqrt{\sum_{i=1}^{d} (x_1^{(i)} - x_2^{(i)})} \qquad X = \begin{bmatrix} \\ \\ \end{bmatrix}_{n \times d}$$

$n = 100$ $\qquad \underline{C_1 \quad C_2 \quad C_3}$

$$XD = \begin{array}{c} \\ X_1 \\ X_2 \\ \vdots \\ X_{100} \end{array} \begin{array}{c} C_1 \quad C_2 \quad C_3 \\ \begin{bmatrix} \\ \\ \\ \end{bmatrix}_{n \times 3} \end{array}$$

$$\mathrm{Min}(XD) = \begin{array}{c} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_{100} \end{array} \begin{bmatrix} C_1 \\ C_2 \\ C_1 \\ \\ C_3 \end{bmatrix}$$

$\rightsquigarrow C_1 = \{X_1, X_3, \ldots\} \Rightarrow \text{Mean}(C_1) \rightsquigarrow \overline{C_1}$

$C_2 = \{X_2, \ldots\} \Rightarrow \overline{C_2}$

$C_3 = \{\ldots, X_{100}\} \Rightarrow \overline{C_3}$

Lecture 07
Hierarchical Clustering

Mahdi Roozbahani
Georgia Tech

These slides are based on slides from Aarti Singh, Eric Xing, Andrew Moore, and Jiawei Han.

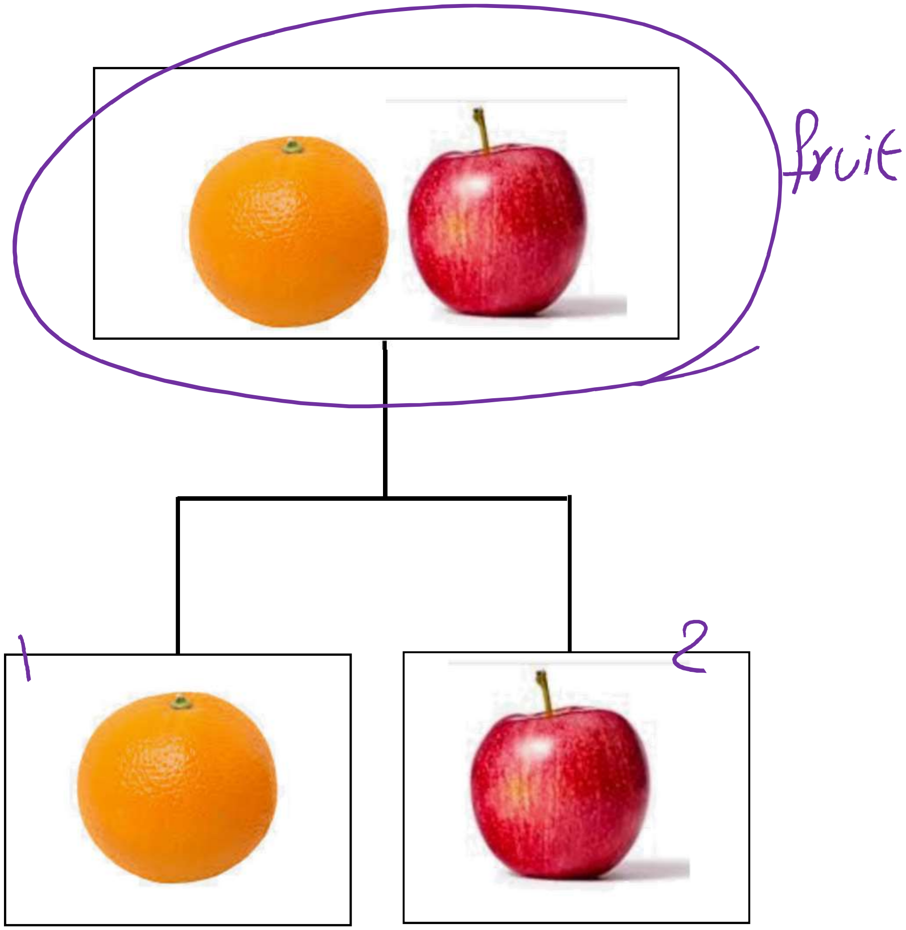
Outline

- Overview
- Bottom-Up vs Top-Down Clustering
- Measuring Distance between Clusters

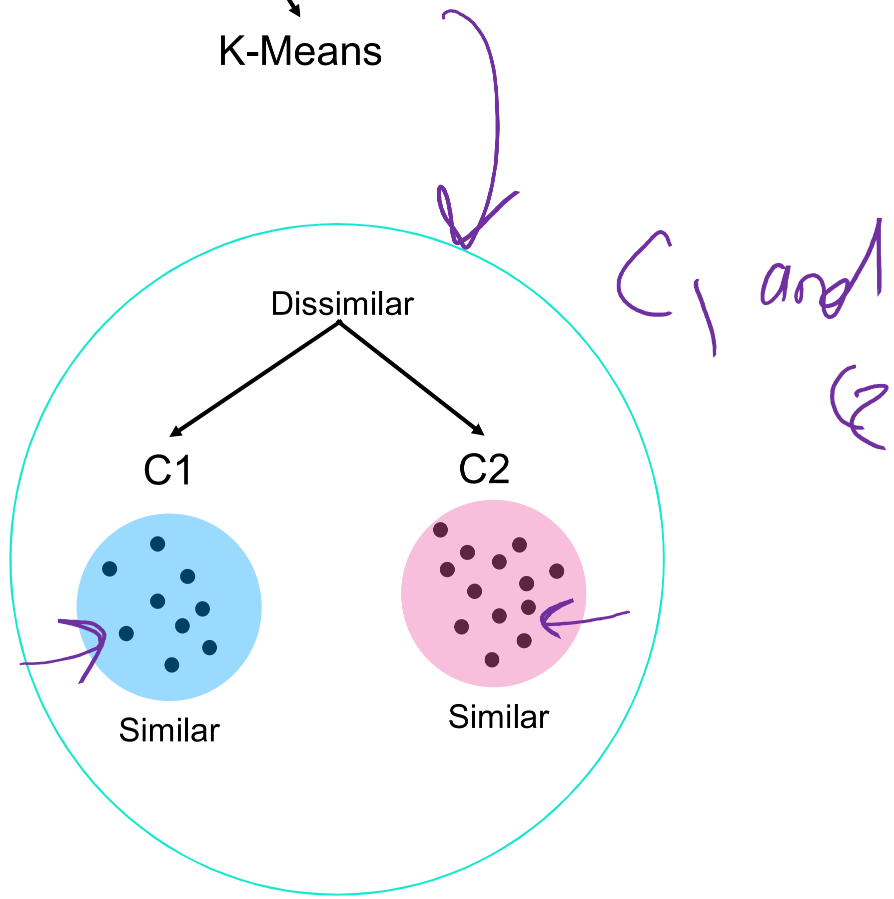
Hierarchical Clustering vs Partitional Clustering

Hierarchical Clustering → Agglomerative, Divisive

Partitional Clustering → K-Means

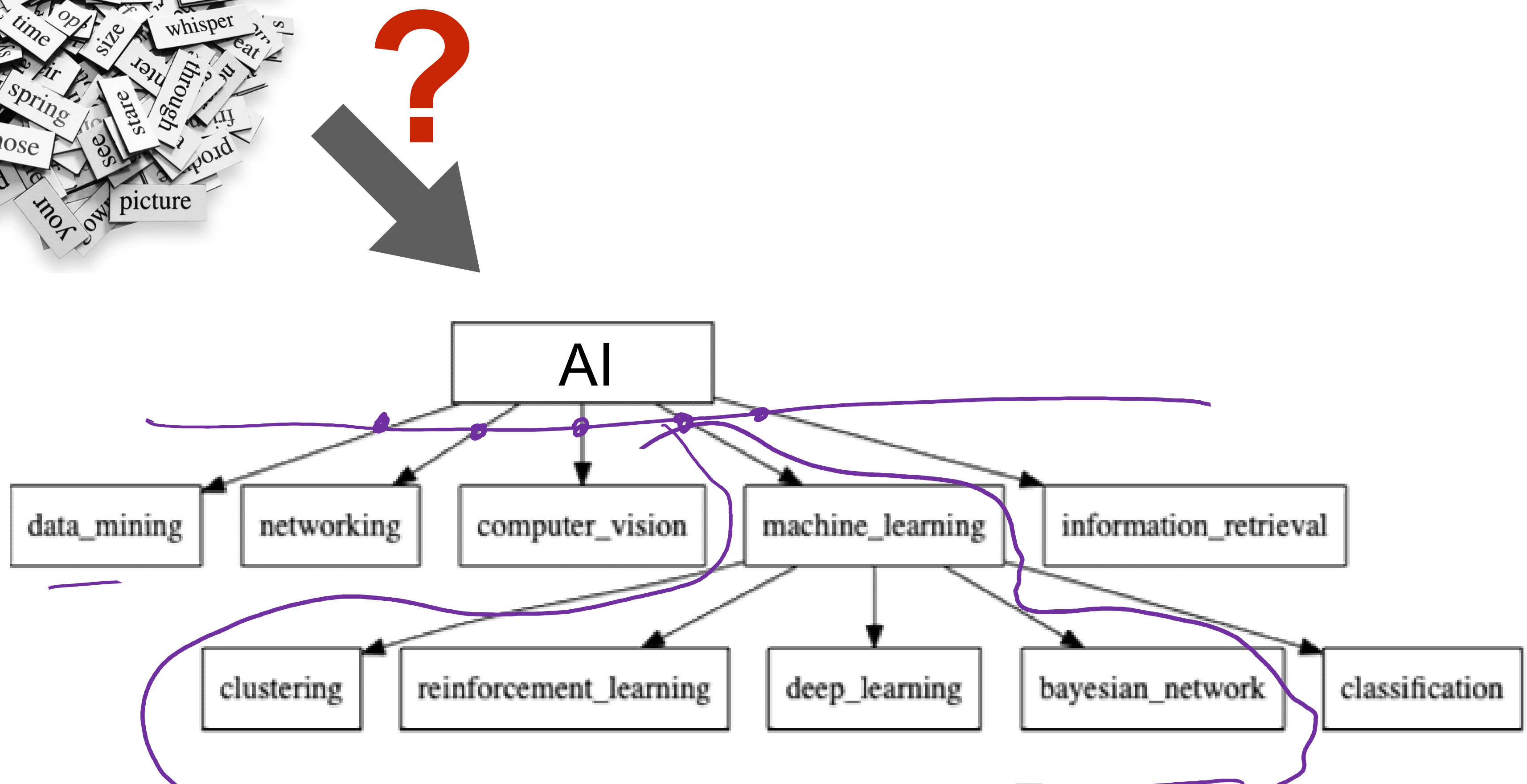


Tree structure (parent-child relationship)



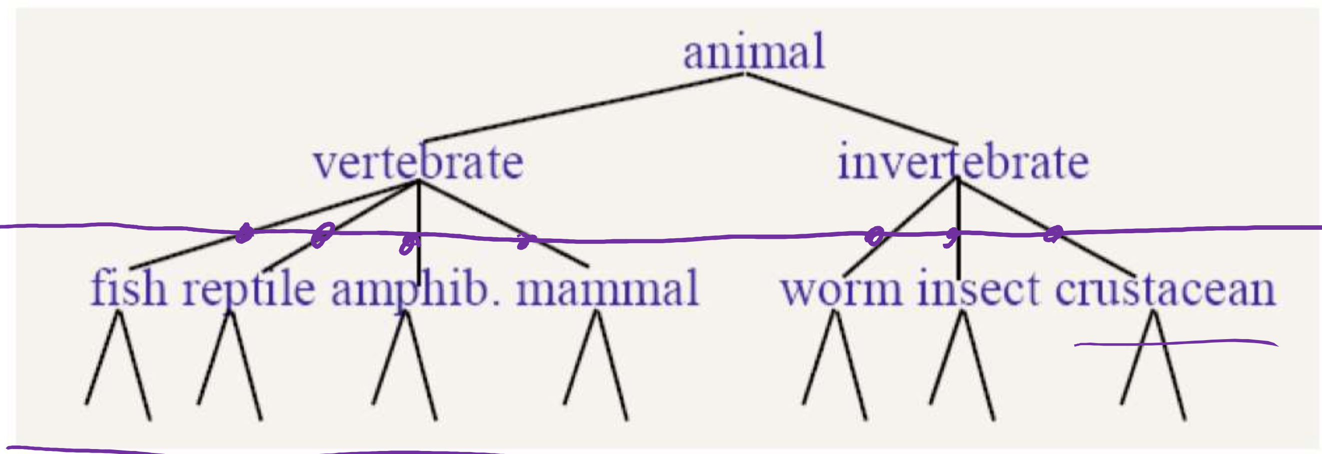
Hierarchical Clustering

- How to organize a set of CS papers into a hierarchy?



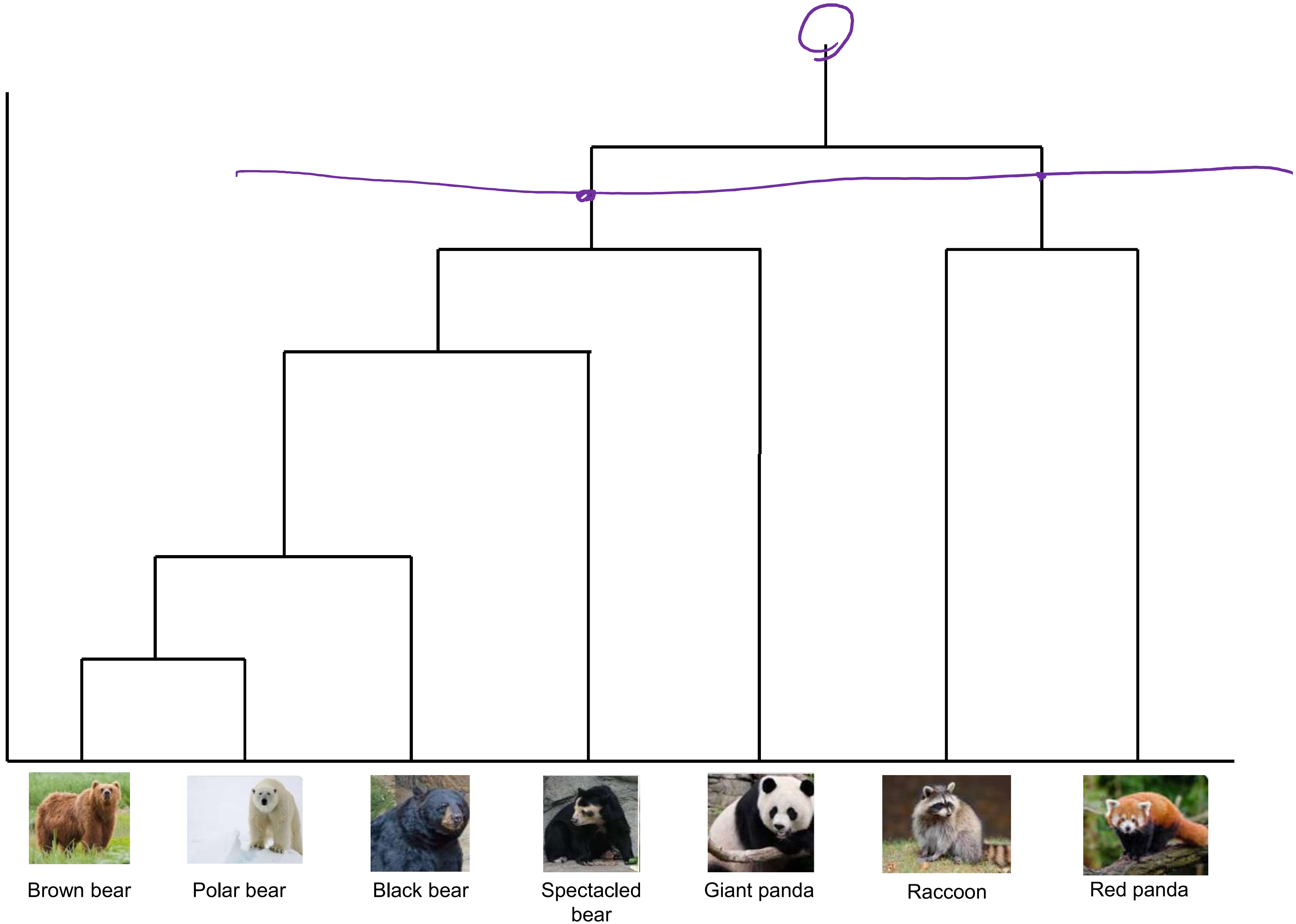
Hierarchical Clustering

- Organize objects into a tree-based hierarchical taxonomy (dendrogram)

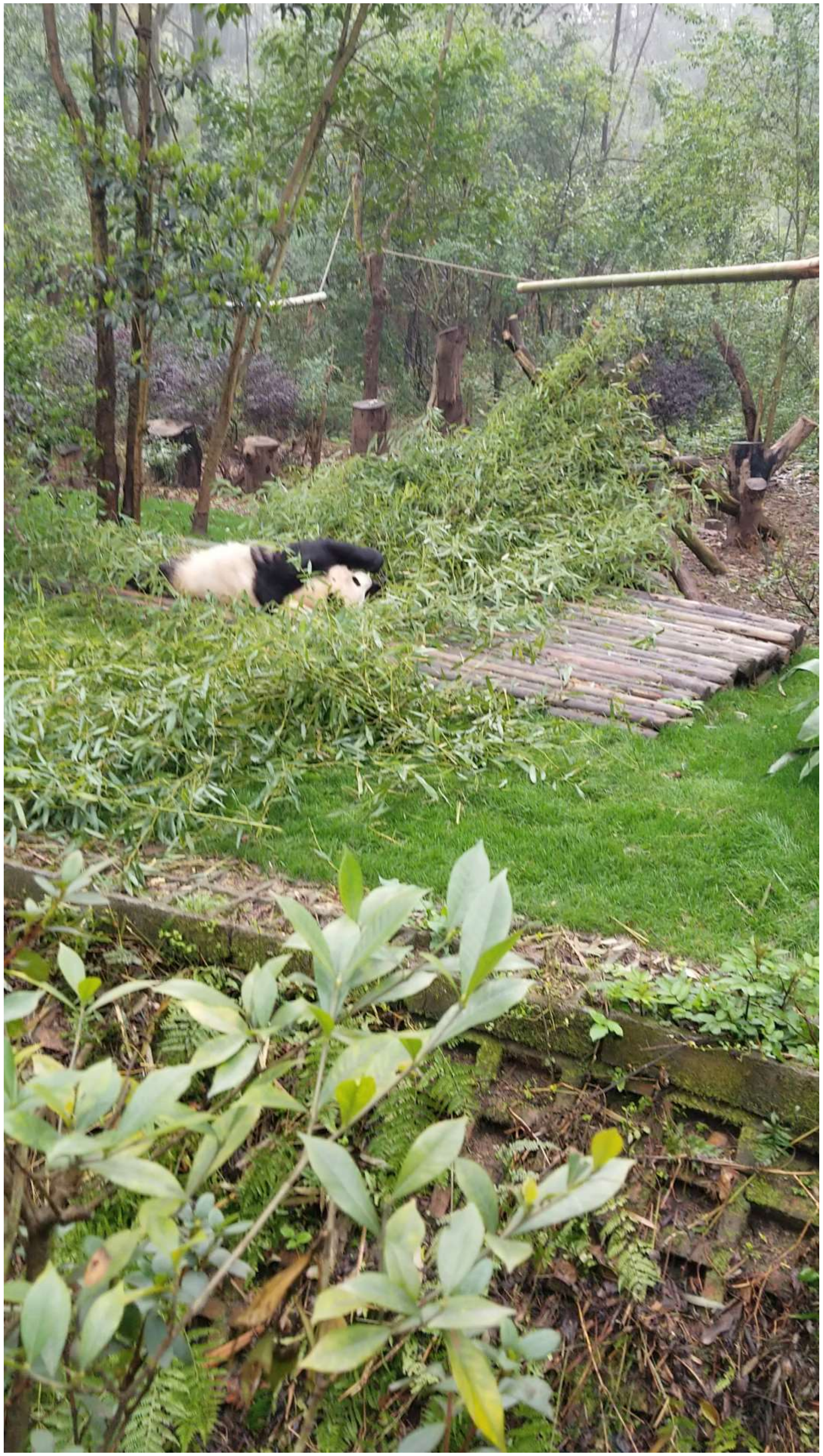


- Many applications in the real world
  - Web pages
  - News articles
  - Scientific papers

DNA sequencing and hierarchical clustering to find the phylogenetic tree of animal evolution

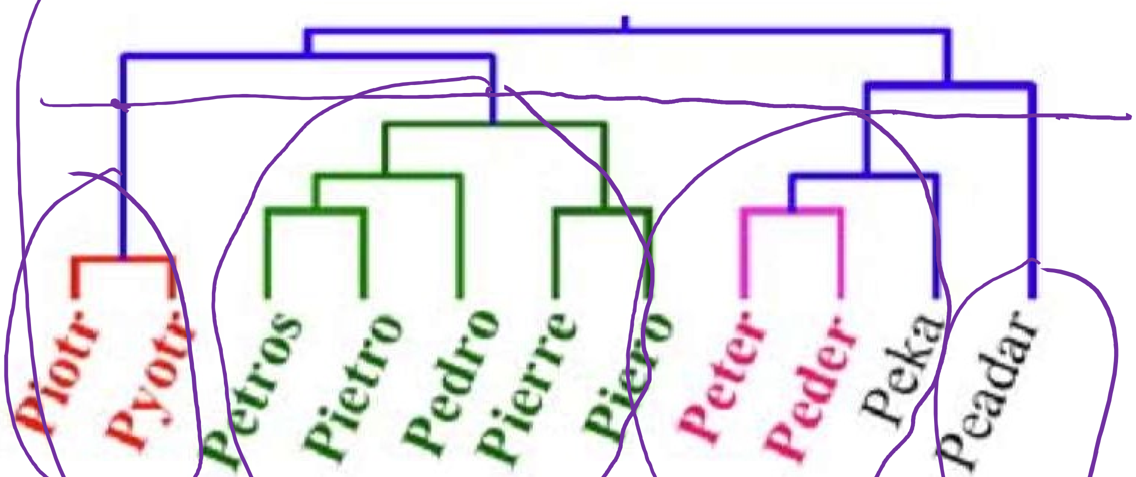


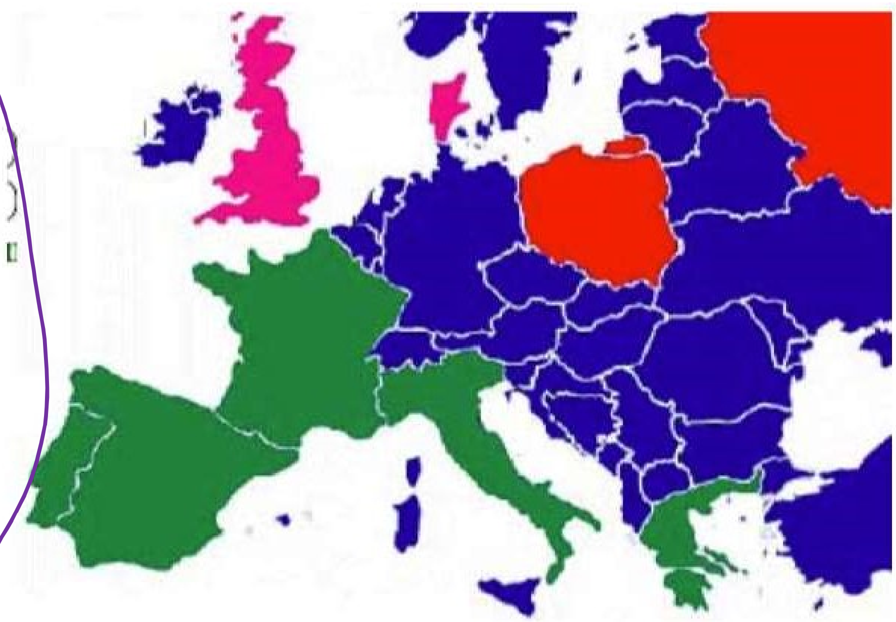
Using Hierarchical clustering, the researchers were able to place the giant pandas closer to bears



Hierarchical Clustering

- Organizing data at multiple granularities
- Cutting the dendrogram at a desired level leads to a sub-cluster: each connected component forms a cluster





Outline

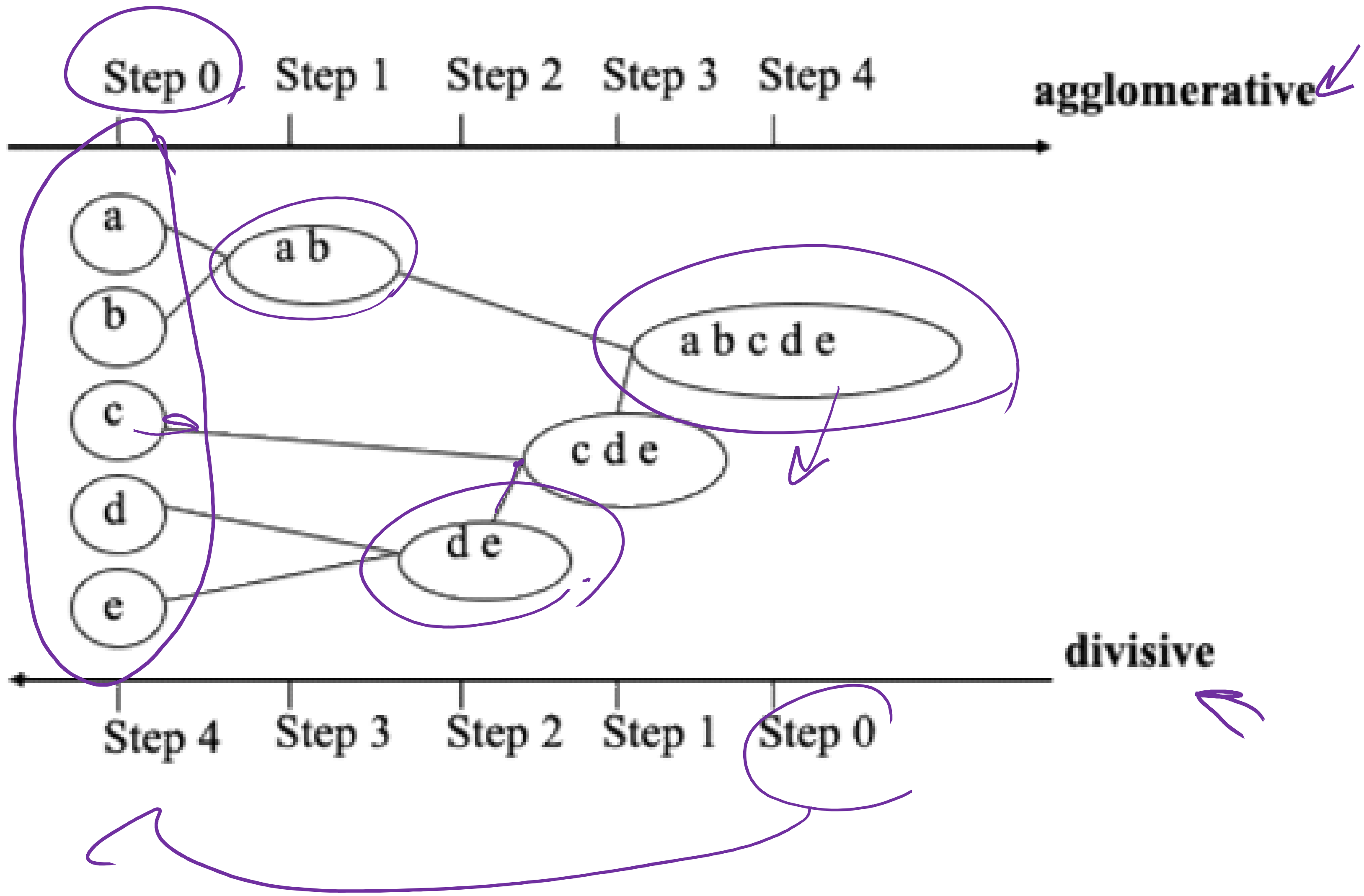
- Overview
- ~~Bottom-Up vs Top-Down Clustering~~
- Measuring Distance between Clusters

Two Paradigms for Hierarchical Clustering

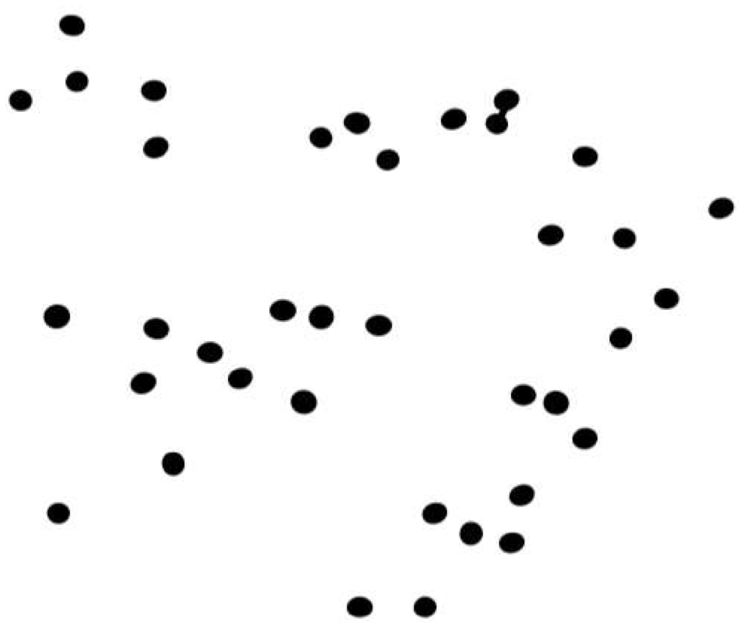
- Bottom-up Agglomerative Clustering
  - Start by considering each object as a separate cluster
  - Repeatedly join the closest pair of clusters
  - Stop when there is only one cluster left

- Top-Down Divisive Clustering
  - Start by considering all objects as one large cluster
  - Recursively divide each cluster into two sub-clusters
  - Stop when each cluster contains only one object

Bottom-Up v.s. Top-Down

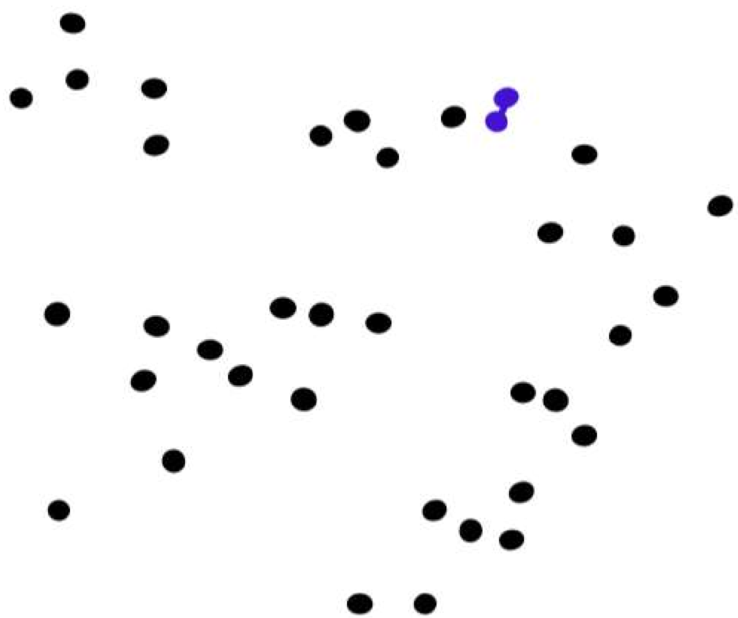


Bottom-Up Agglomerative Clustering

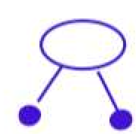


1. Say "Every point is it's own cluster"

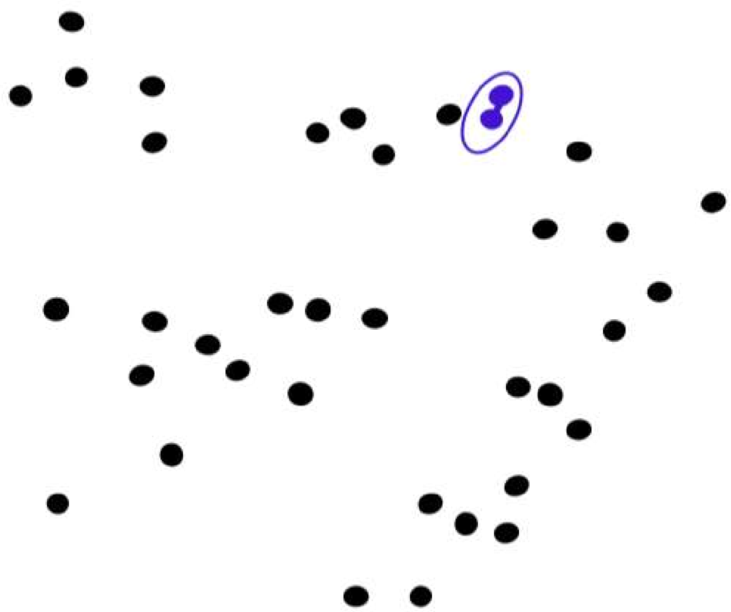
Bottom-Up Agglomerative Clustering



1. Say "Every point is it's own cluster"
2. Find "most similar" pair of clusters

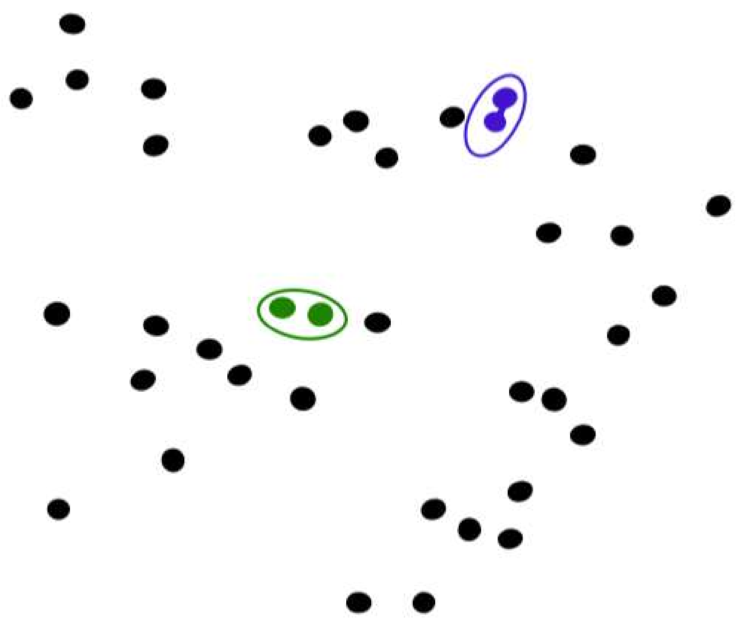


Bottom-Up Agglomerative Clustering

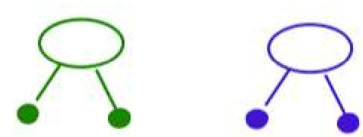


1. Say "Every point is it's own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster

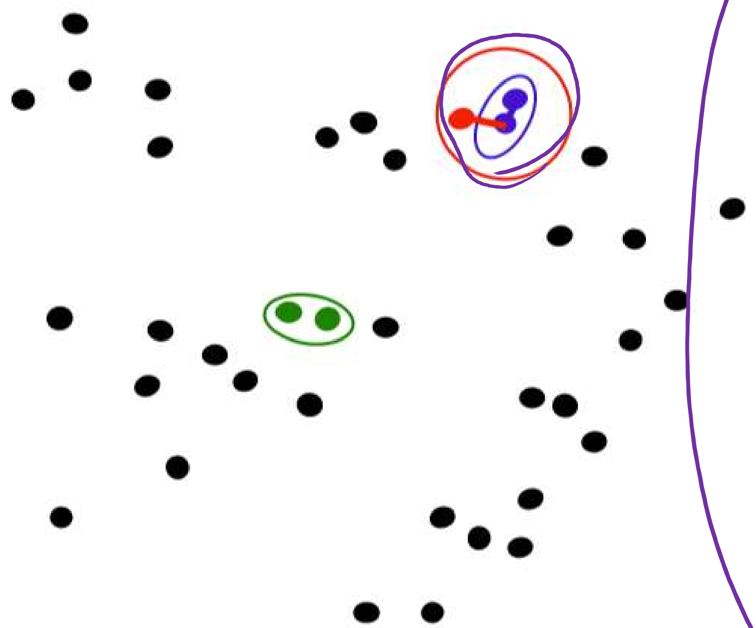
Bottom-Up Agglomerative Clustering



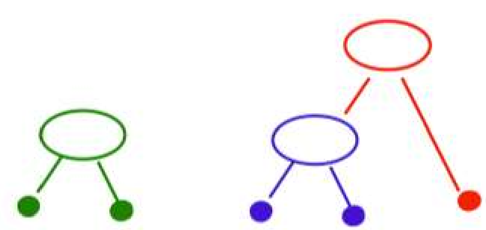
1. Say "Every point is it's own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster
4. Repeat



Bottom-Up Agglomerative Clustering



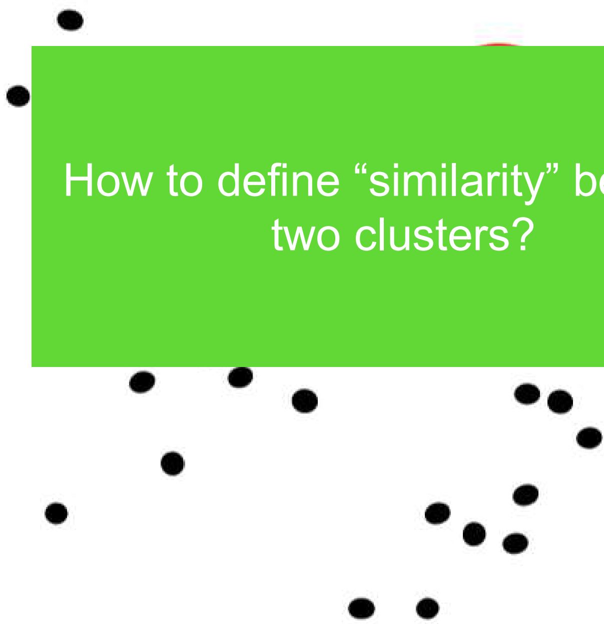
1. Say "Every point is it's own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster
4. Repeat



Outline

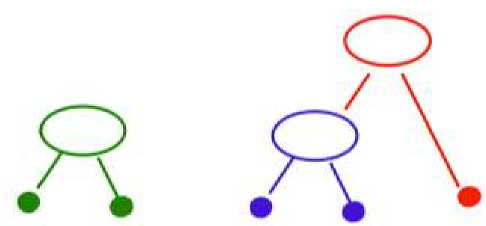
- Overview
- Bottom-Up vs Top-Down Clustering
- Measuring Distance between Clusters

Key Question: Similarity Function

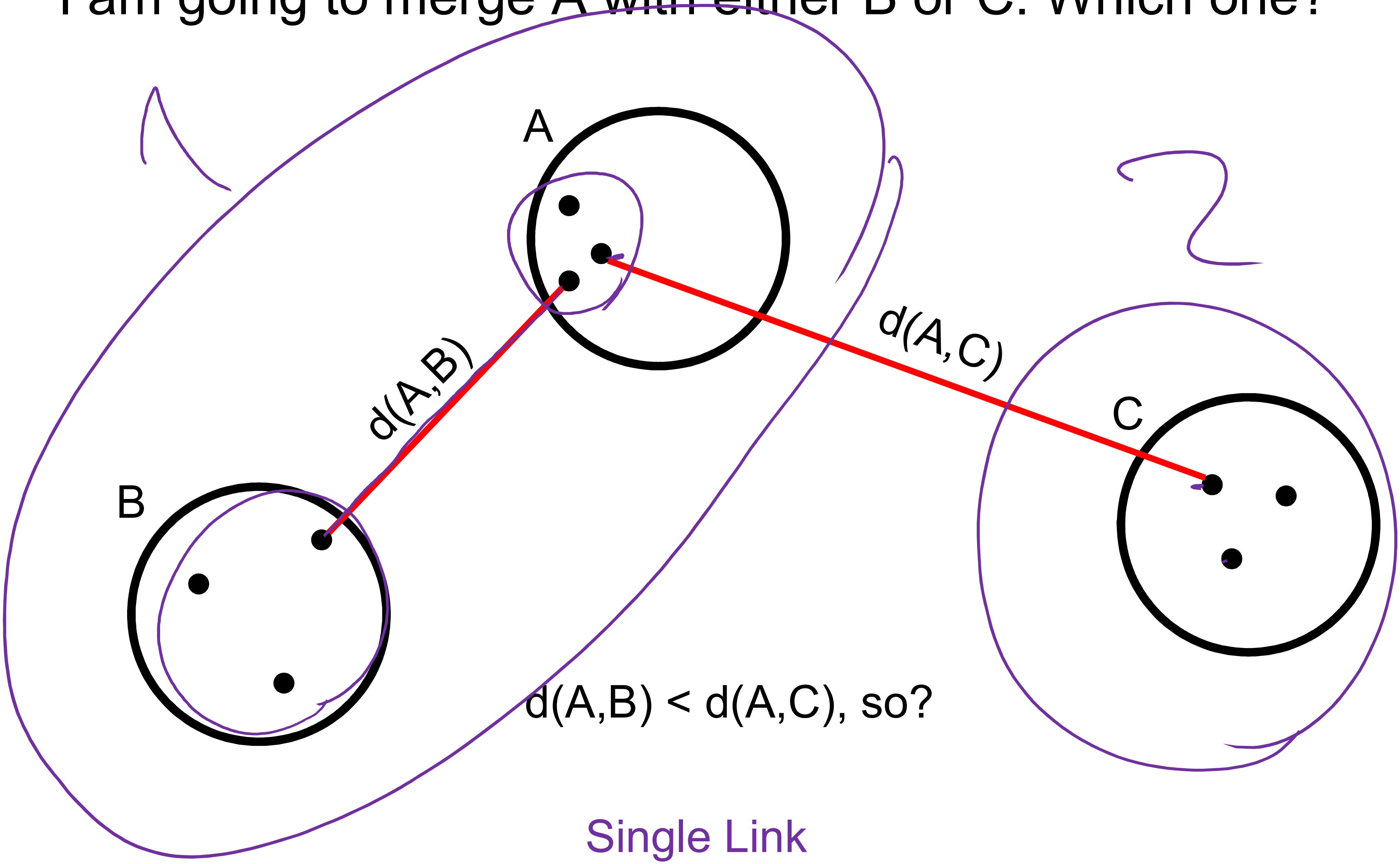


How to define "similarity" between two clusters?

1. Say "Every point is it's own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster
4. Repeat



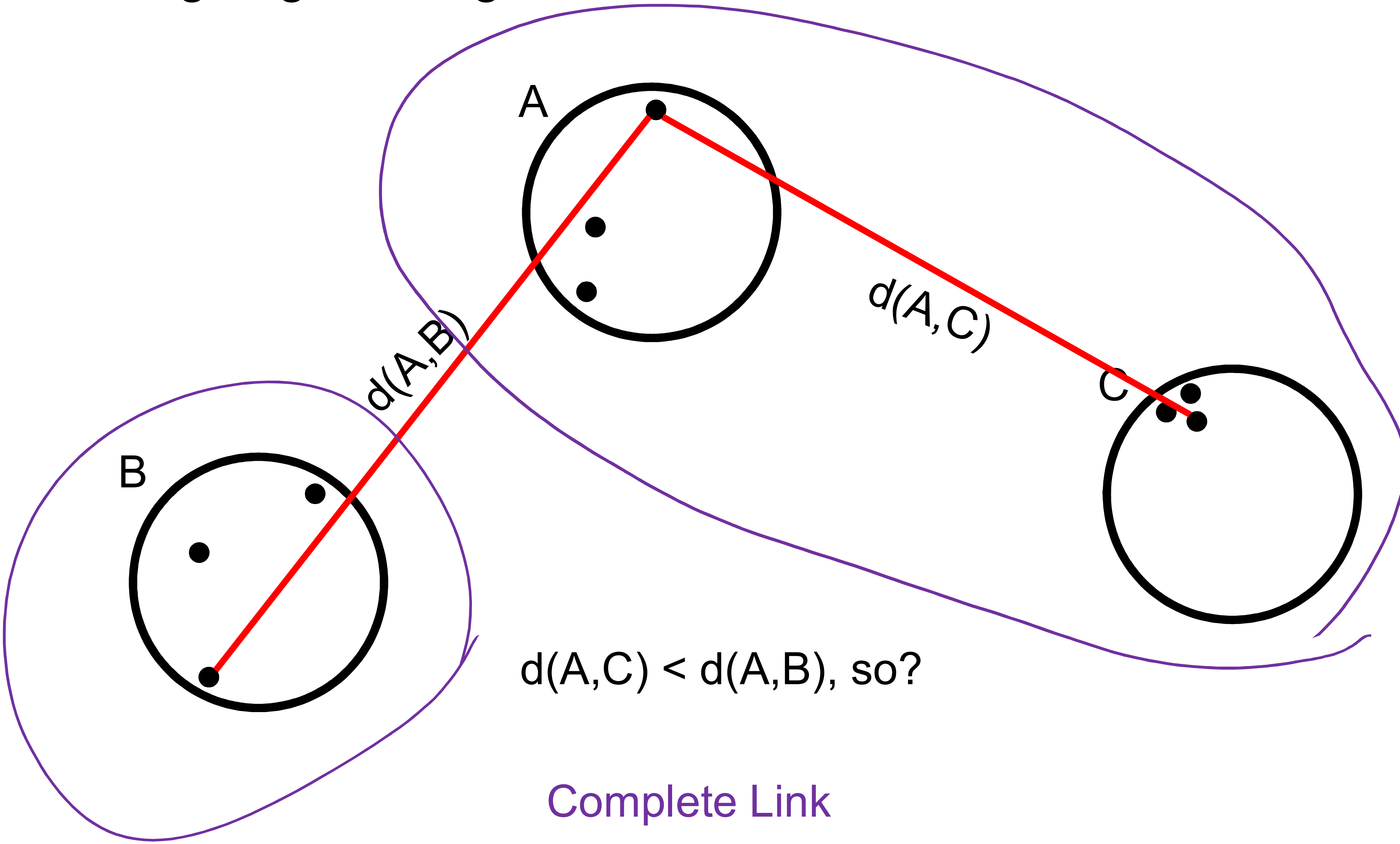
I am going to merge A with either B or C. Which one?



d(A,B) < d(A,C), so?

Single Link

I am going to merge A with either B or C. Which one?

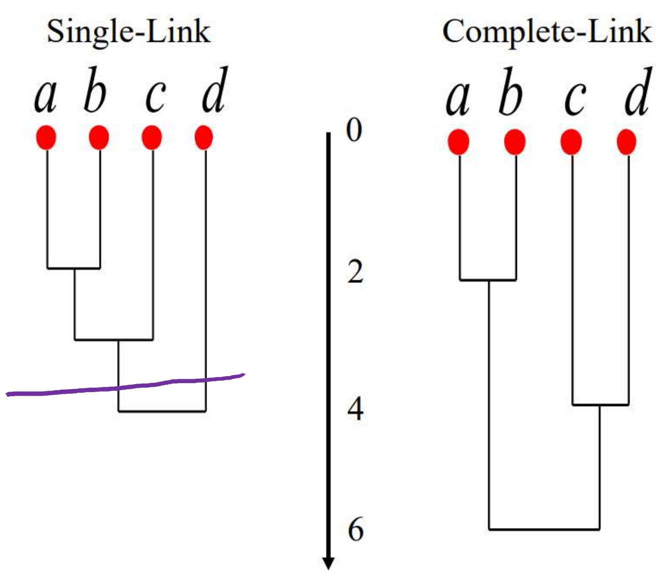


d(A,C) < d(A,B), so?

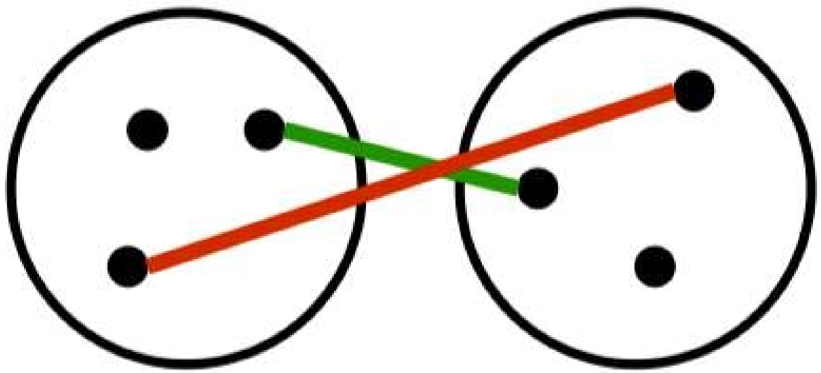
Complete Link

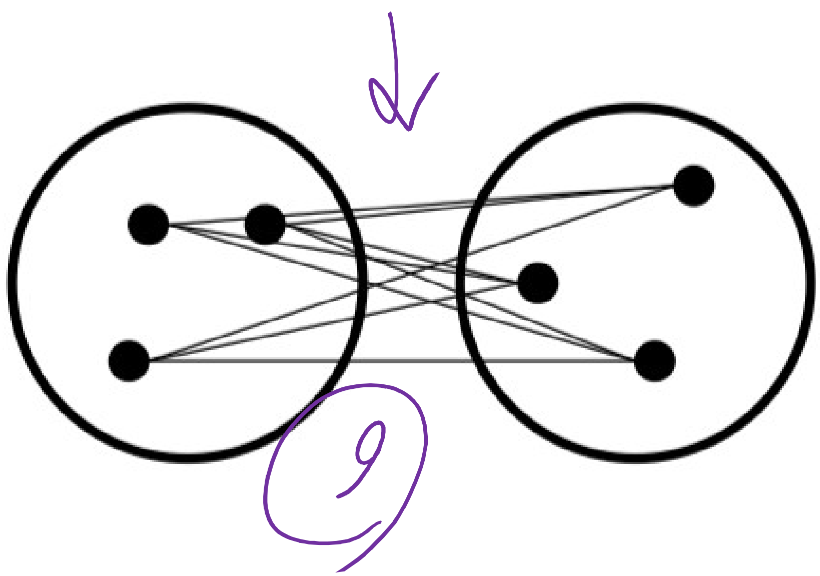
- Single link: A chain of points can be extended for long distances without regard to the overall shape of the emerging cluster. This effect is called *chaining*. It is also sensitive to outliers. It is faster in general.
- Complete link: Clusters are split into two groups of roughly equal size when we cut the dendrogram at the last merge. In general, this is a more useful organization of the data than a clustering with chains. It avoids chaining and more robust to outliers. Generally slower.
- Average link: When you don't know which one may be better for you, start it with the average link method.

Dendrograms



How to Define Distance Between Two Clusters?

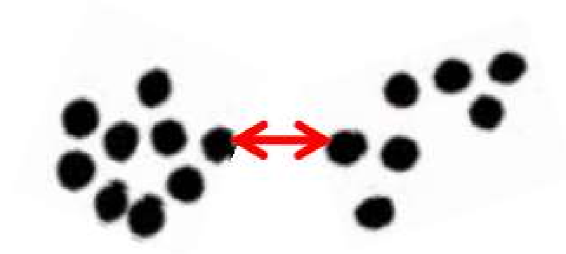




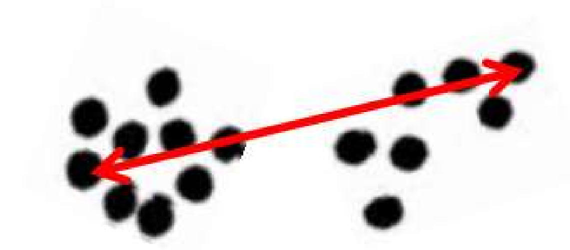
Bottom-up Agglomerative clustering

Different algorithms differ in how the similarities are defined (and hence updated) between two clusters

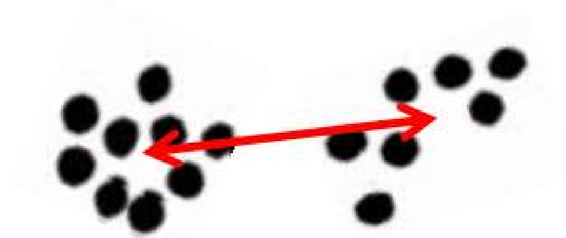
- Single-Link
  - Nearest Neighbor: similarity between their closest members.



- Complete-Link
  - Furthest Neighbor: similarity between their furthest members.



- Centroid
  - Similarity between the centers of gravity

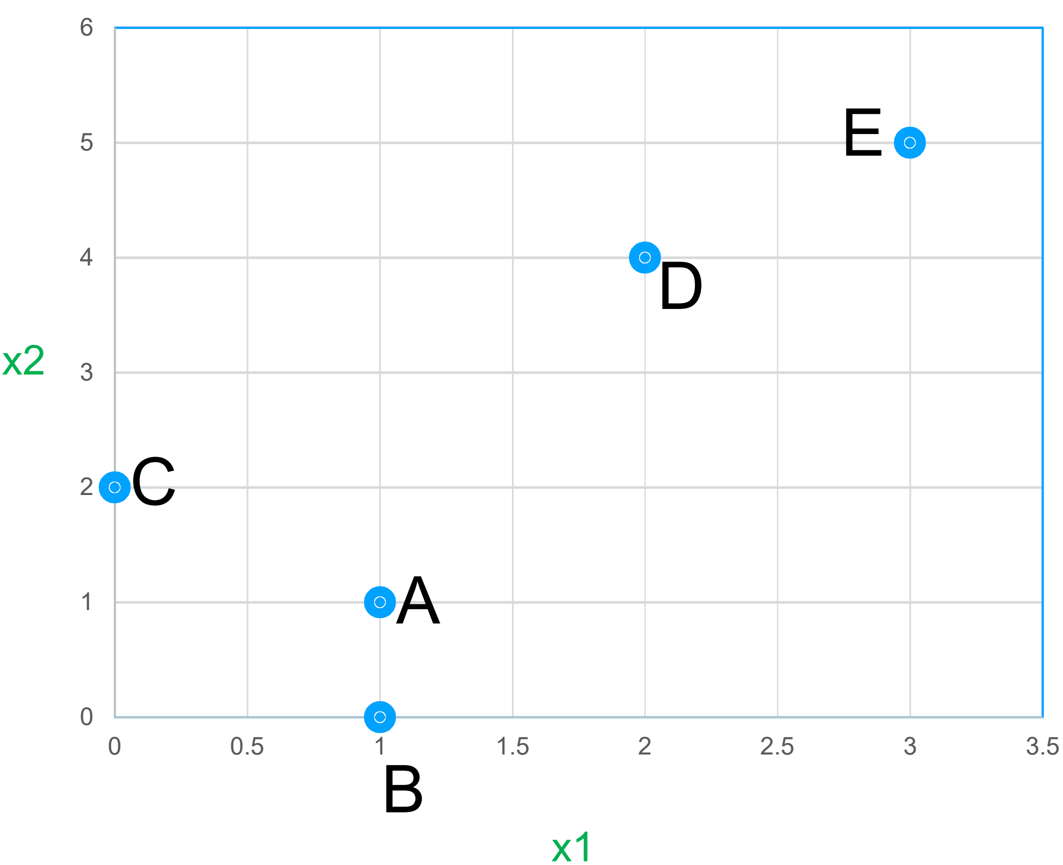


- Average-Link
  - Average similarity of all cross-cluster pairs.

Distance Between Clusters

Different distance functions can lead to different results!

i	X1	X2
A	1	1
B	1	0
C	0	2
D	2	4
E	3	5



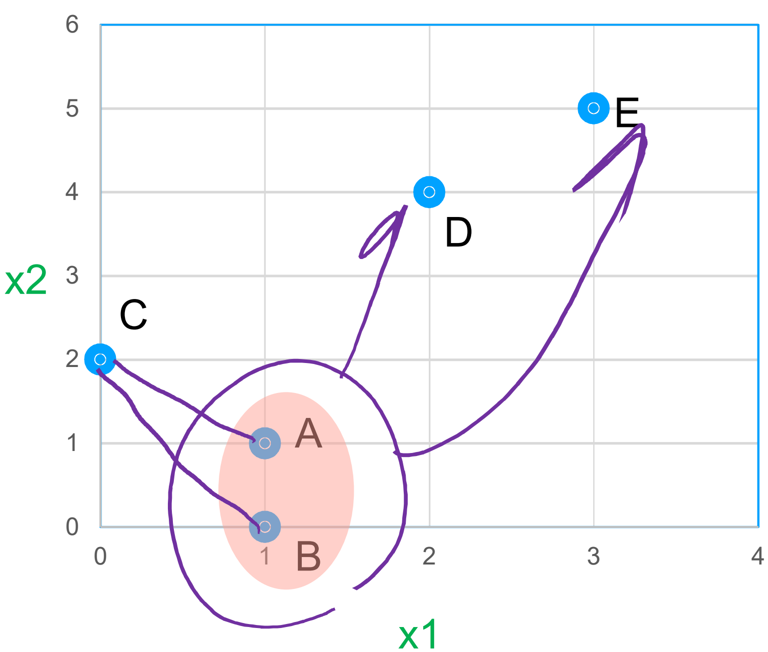
EUCLIDEAN DISTANCE

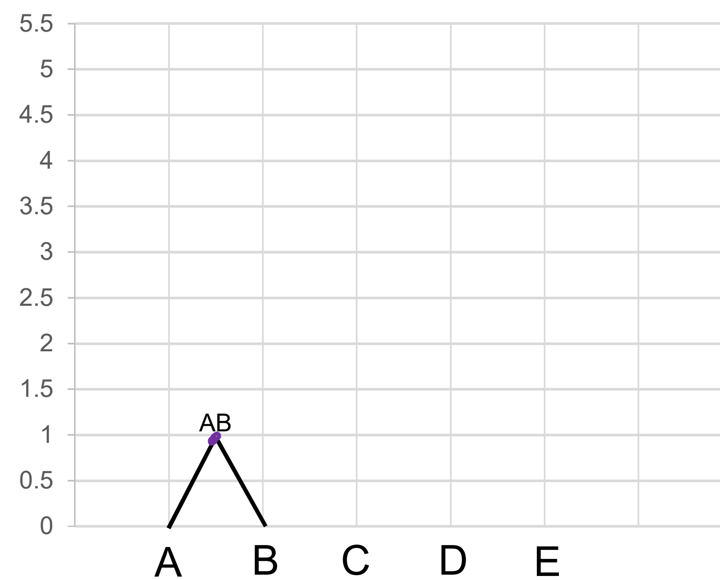
	A	B	C	D	E
A	0	1	1.4	3.2	4.5
B	1	0	2.2	4.1	5.4
C	1.4	2.2	0	2.8	4.2
D	3.2	4.1	2.8	0	1.4
E	4.5	5.4	4.2	1.4	0

Distance based on Average point (Bottom-Up Clustering)

	A	B	C	D	E
A	0	1	1.4	3.2	4.5
B	1	0	2.2	4.1	5.4
C	1.4	2.2	0	2.8	4.2
D	3.2	4.1	2.8	0	1.4
E	4.5	5.4	4.2	1.4	0

	(A,B)	C	D	E
(A,B)	0	1.8	3.6	4.9
C	1.8	0	2.8	4.2
D	3.6	2.8	0	1.4
E	4.9	4.2	1.4	0



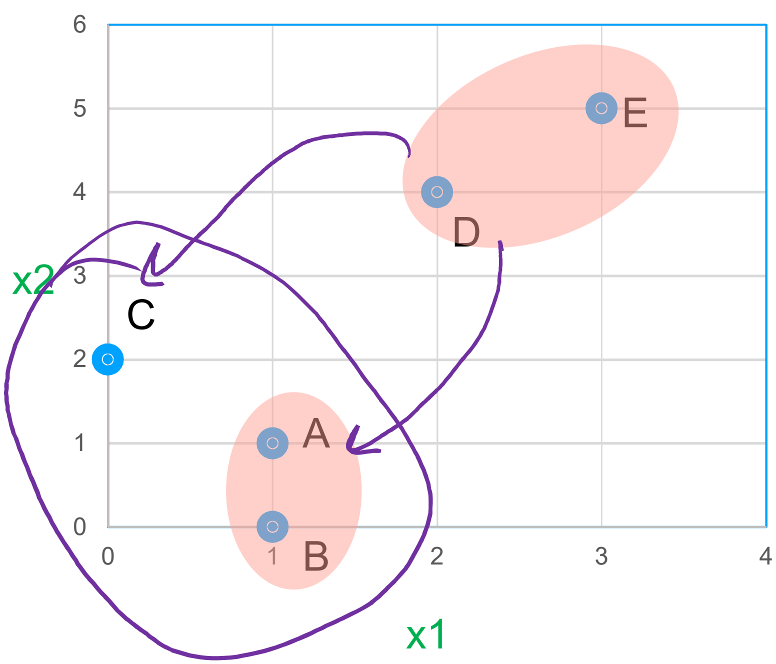


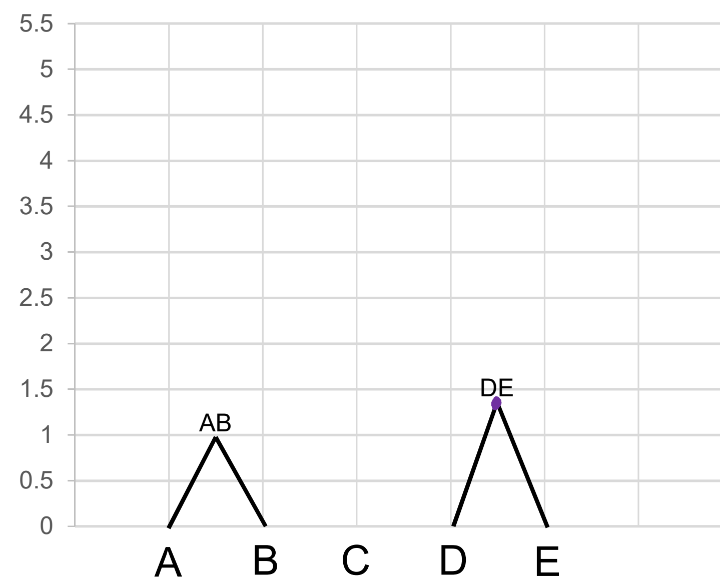
Dendrogram

Distance based on average point (Bottom-Up Clustering)

	(A,B)	C	D	E
(A,B)	0	1.8	3.6	4.9
C	1.8	0	2.8	4.2
D	3.6	2.8	0	1.4
E	4.9	4.2	1.4	0

	(A,B)	C	(D,E)
(A,B)	0	1.8	4.25
C	1.8	0	3.5
(D,E)	4.25	3.5	0



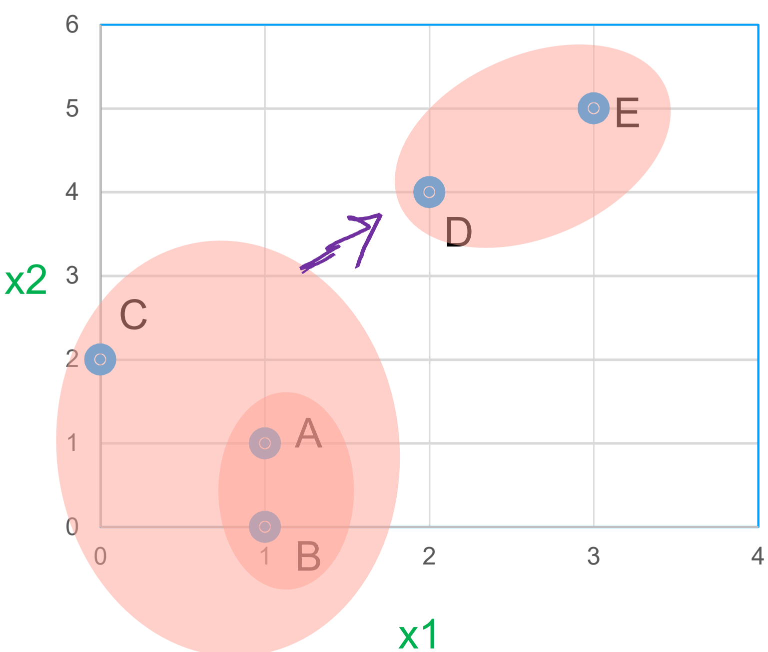


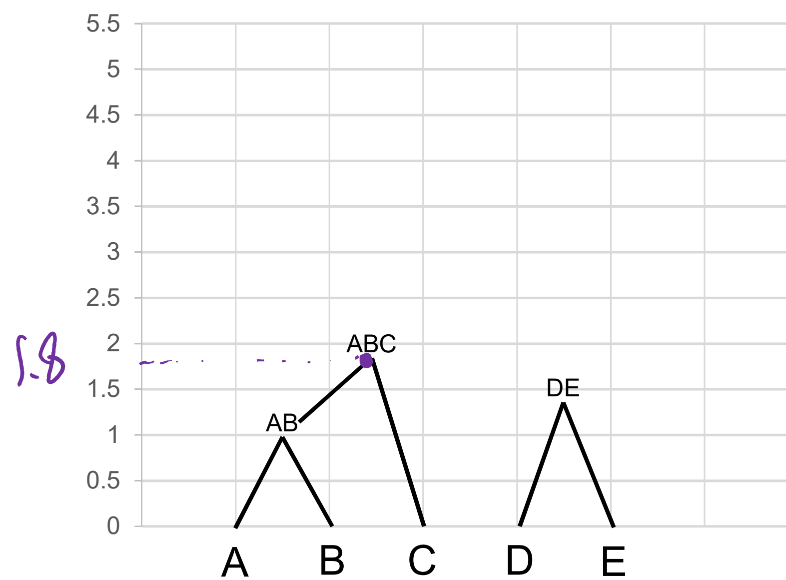
Dendrogram

Distance based on average point (Bottom-Up Clustering)

	(A,B)	C	(D,E)
(A,B)	0	1.8	4.25
C	1.8	0	3.5
(D,E)	4.25	3.5	0

	((A,B),C)	(D,E)
((A,B),C)	0	3.875
(D,E)	3.875	0



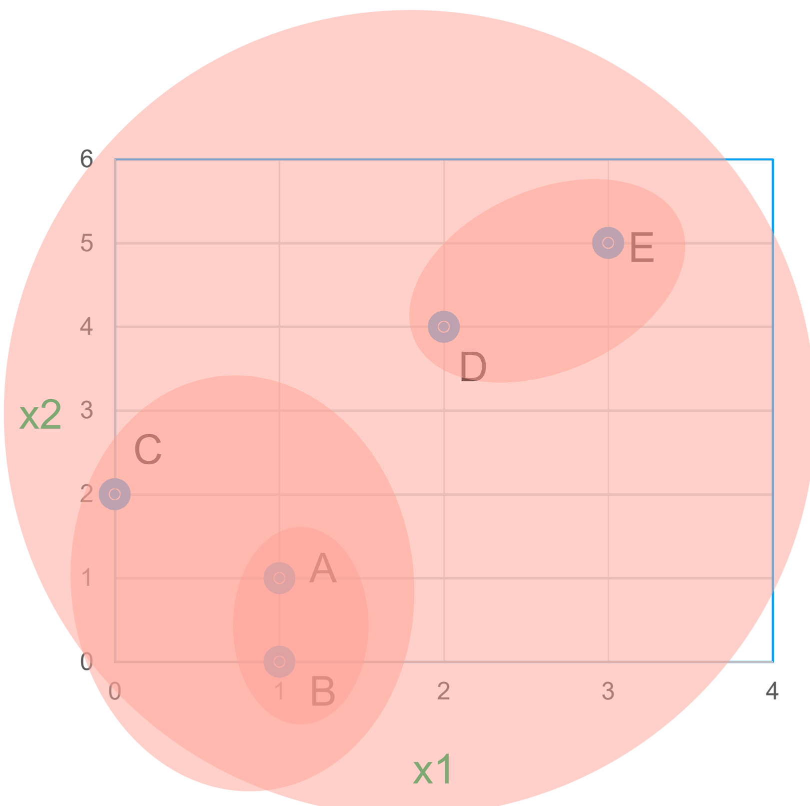


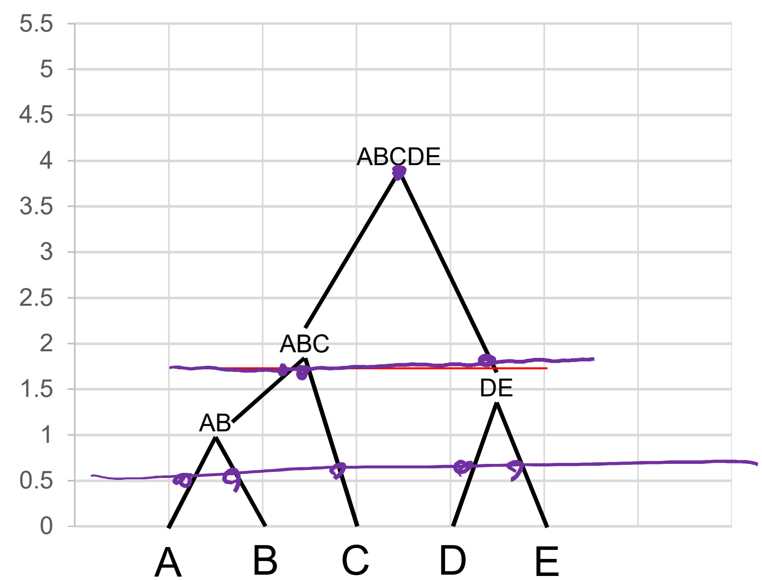
Dendrogram

Distance based on average point (Bottom-Up Clustering)

	((A,B),C)	(D,E)
((A,B),C)	0	3.875
(D,E)	3.875	0

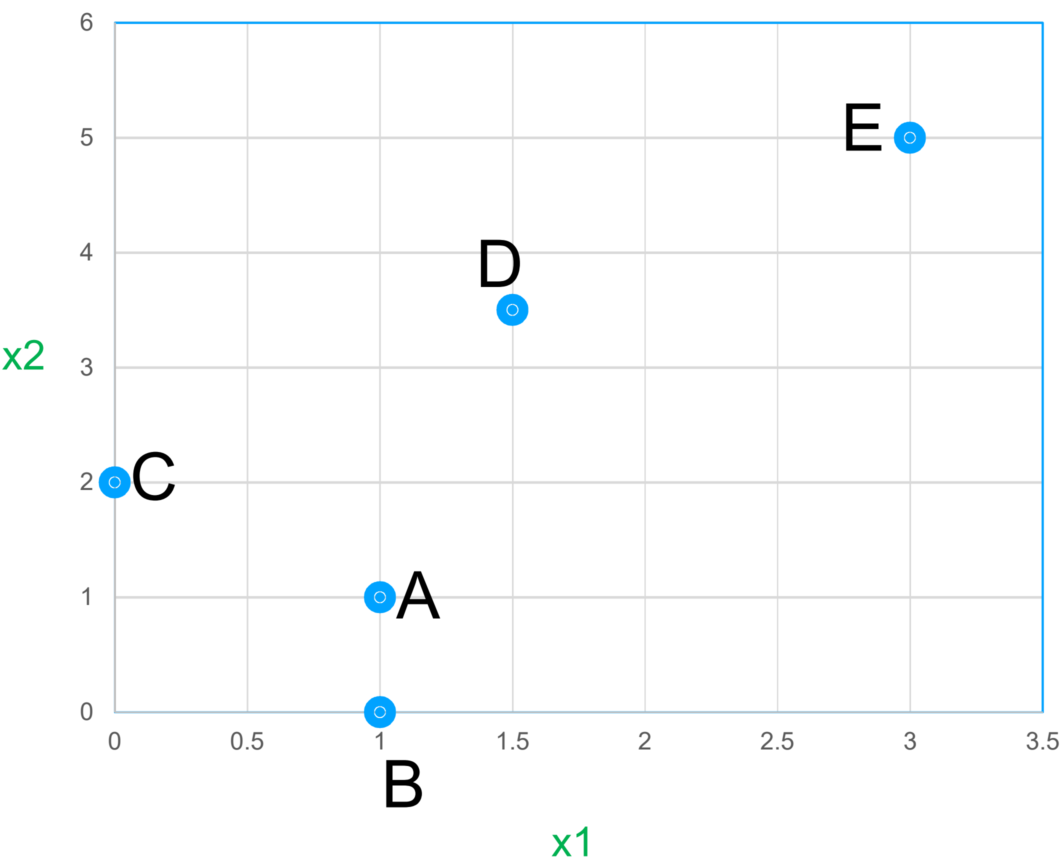
	(((A,B),C),(D,E))
(((A,B),C),(D,E))	0





Dendrogram

i	X1	X2
A	1	1
B	1	0
C	0	2
D	1.5	3.5
E	3	5



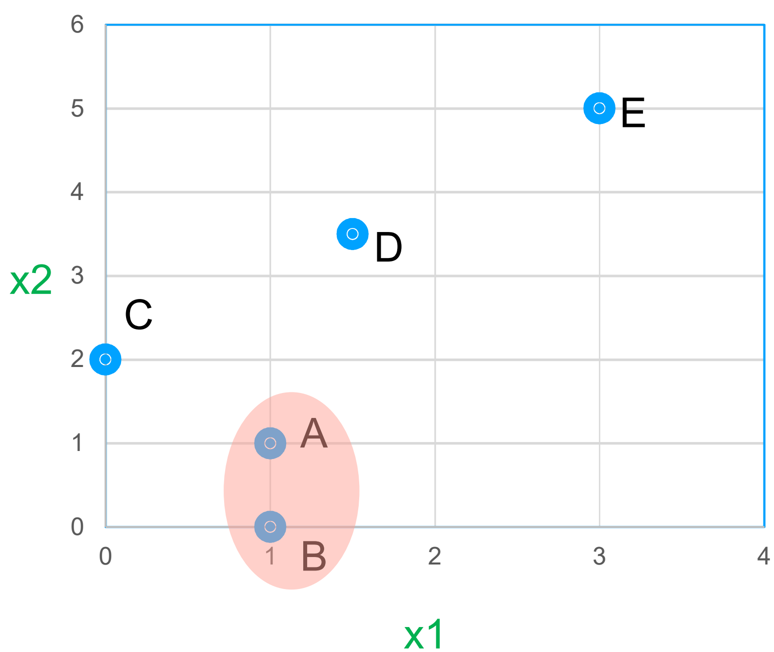
EUCLIDEAN DISTANCE

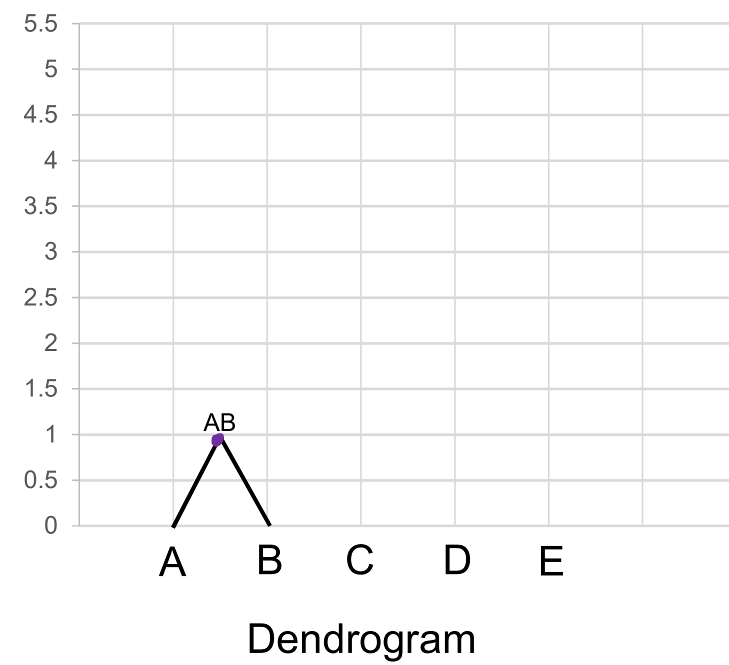
	A	B	C	D	E
A	0	1	1.4	2.55	4.5
B	1	0	2.2	3.53	5.4
C	1.4	2.2	0	2.12	4.2
D	2.55	3.53	2.12	0	2.12
E	4.5	5.4	4.2	2.12	0

Distance based on Single Link (Bottom-Up Clustering)

	A	B	C	D	E
A	0	1	1.4	2.55	4.5
B	1	0	2.2	3.53	5.4
C	1.4	2.2	0	2.12	4.2
D	2.55	3.53	2.12	0	2.12
E	4.5	5.4	4.2	2.12	0

	(A,B)	C	D	E
(A,B)	0	1.4	2.55	4.5
C	1.4	0	2.12	4.2
D	2.55	2.12	0	2.12
E	4.5	4.2	2.12	0

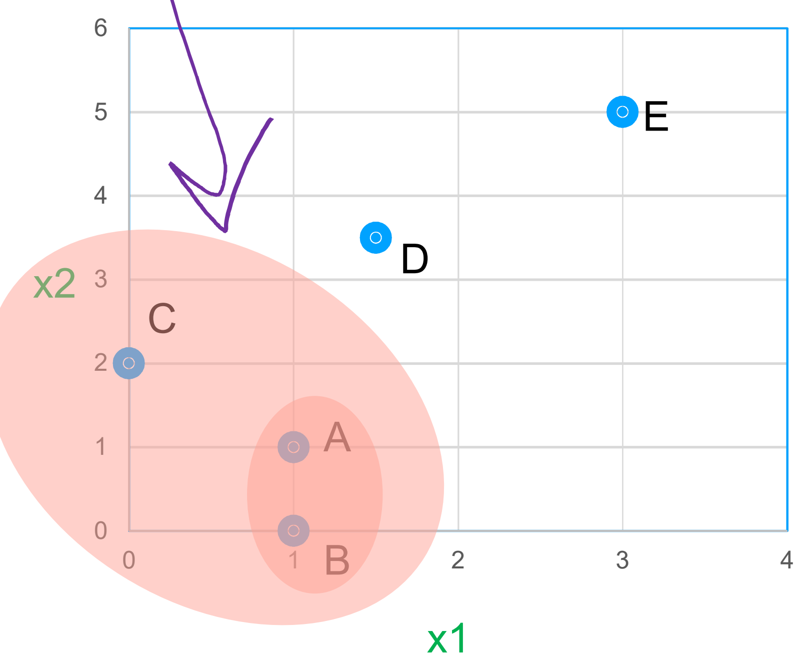


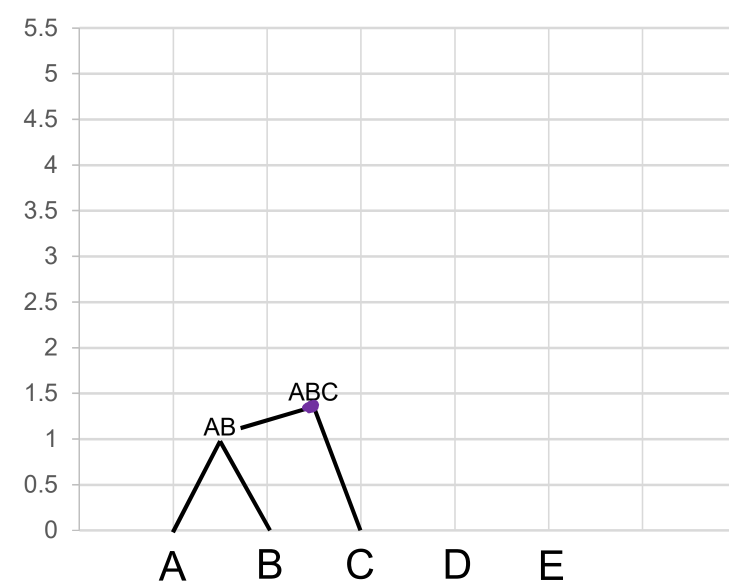


Distance based on Single Link (Bottom-Up Clustering)

	(A,B)	C	D	E
(A,B)	0	1.4	2.55	4.5
C	1.4	0	2.12	4.2
D	2.55	2.12	0	2.12
E	4.5	4.2	2.12	0

	(A,B),C	D	E
(A,B),C	0	2.12	4.2
D	2.12	0	2.12
E	4.2	2.12	0



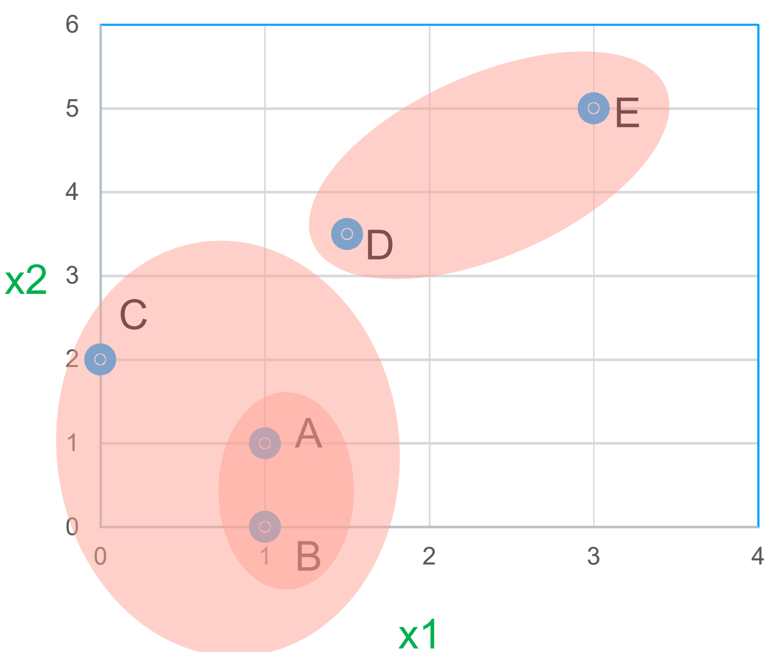


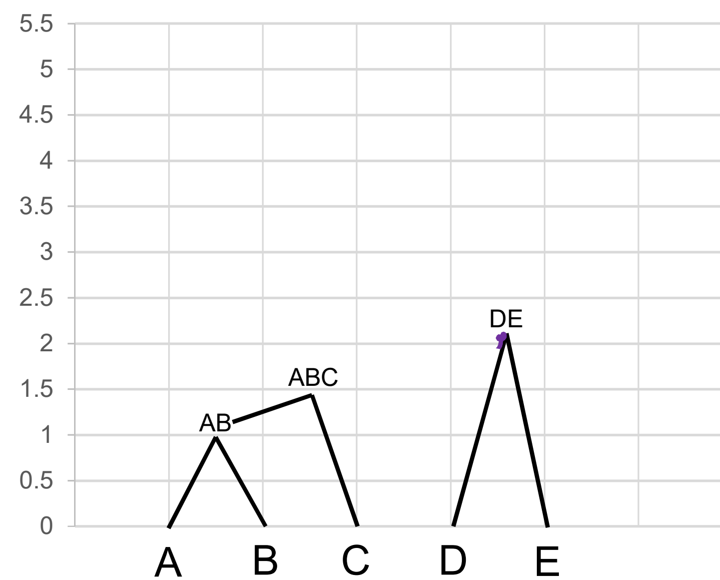
Dendrogram

Distance based on Single Link (Bottom-Up Clustering)

	(A,B), C	D	E
(A,B), C	0	2.12	4.2
D	2.12	0	2.12
E	4.2	2.12	0

	((A,B),C)	(D,E)
((A,B),C)	0	2.12
(D,E)	2.12	0



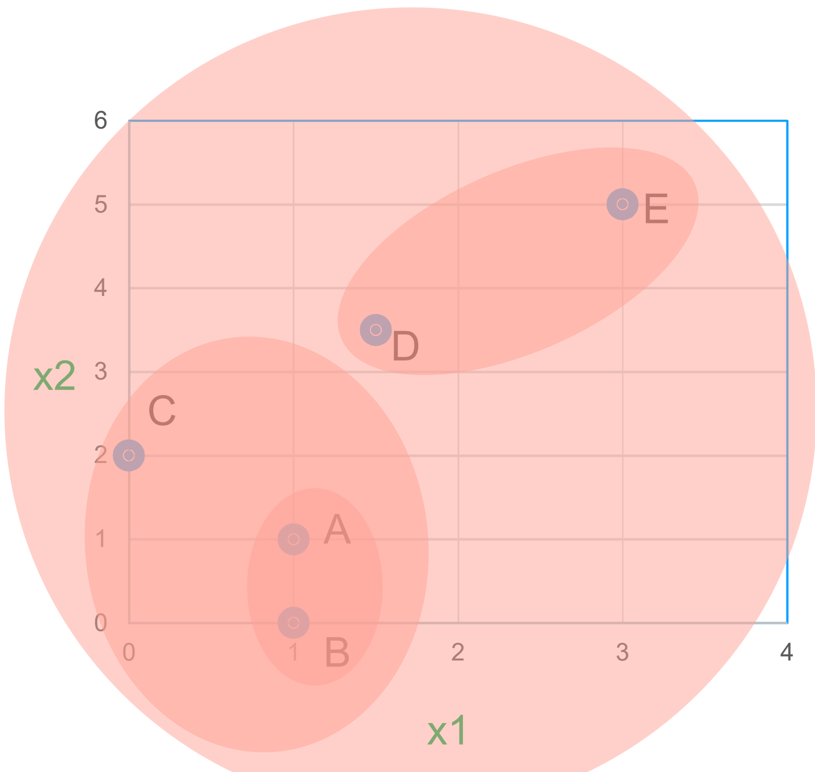


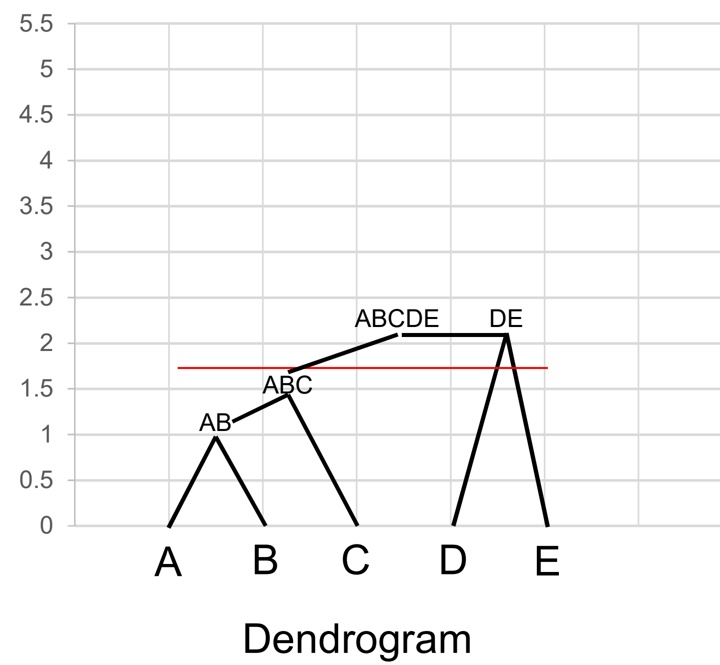
Dendrogram

Distance based on Single Link (Bottom-Up Clustering)

	((A,B),C)	(D,E)
((A,B),C)	0	2.12
(D,E)	2.12	0

	(((A,B),C),(D,E))
(((A,B),C),(D,E))	0

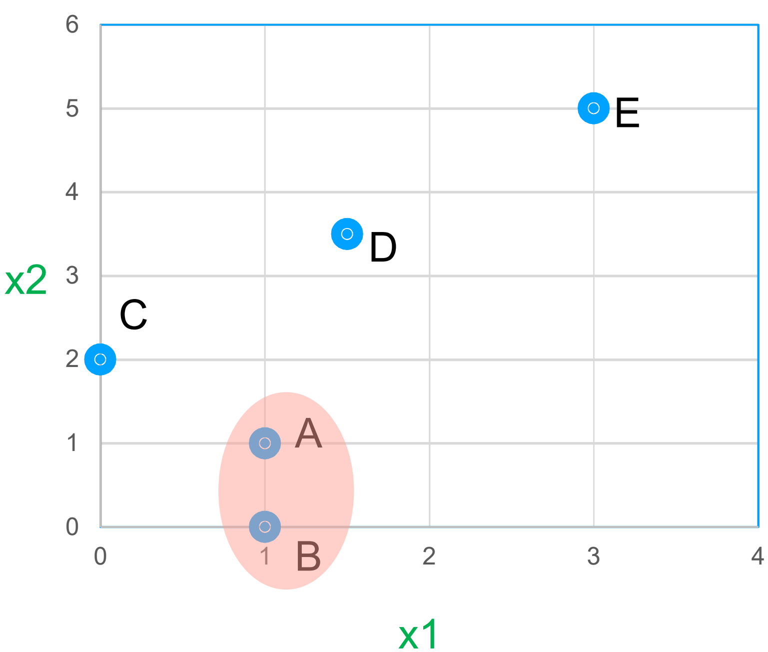


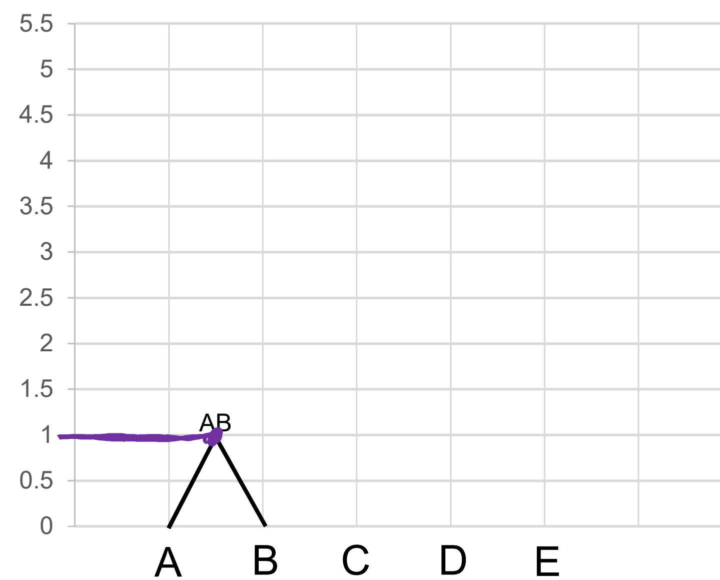


Distance based on Complete Link (Bottom-Up Clustering)

	A	B	C	D	E
A	0	1	1.4	2.55	4.5
B	1	0	2.2	3.53	5.4
C	1.4	2.2	0	2.12	4.2
D	2.55	3.53	2.12	0	2.12
E	4.5	5.4	4.2	2.12	0

	(A,B)	C	D	E
(A,B)	0	2.2	3.55	5.4
C	2.2	0	2.12	4.2
D	3.55	2.12	0	2.12
E	5.4	4.2	2.12	0



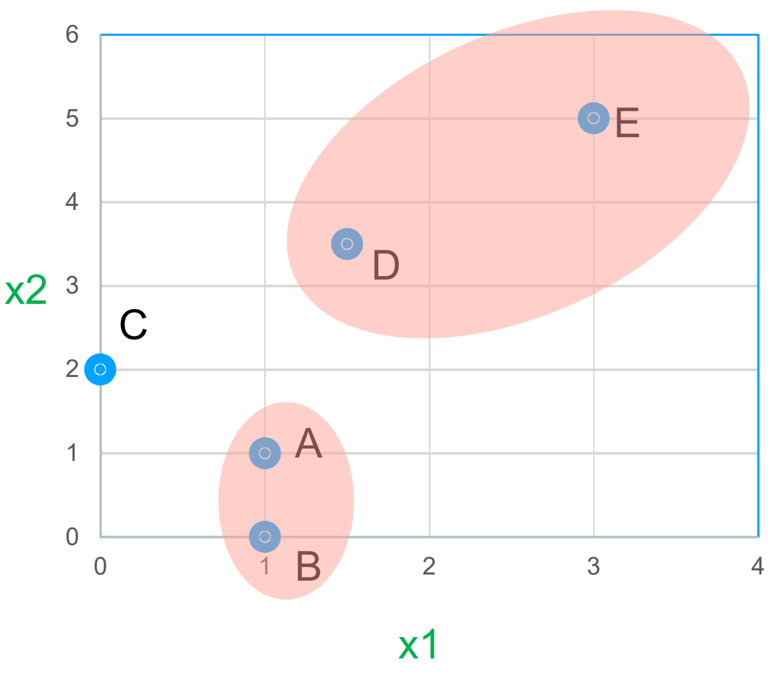


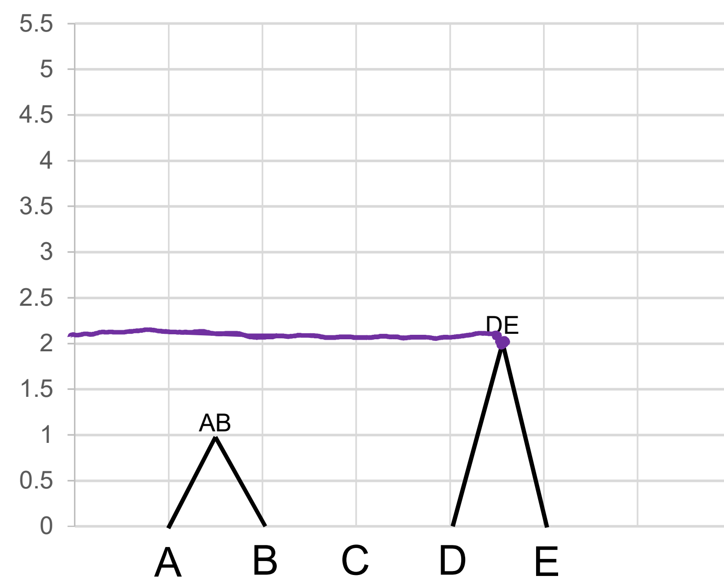
Dendrogram

Distance based on Complete Link (Bottom-Up Clustering)

	(A,B)	C	D	E
(A,B)	0	2.2	3.55	5.4
C	2.2	0	2.12	4.2
D	3.55	2.12	0	2.12
E	5.4	4.2	2.12	0

	(A,B)	C	(D,E)
(A,B)	0	2.2	5.4
C	2.2	0	4.2
(D,E)	5.4	4.2	0



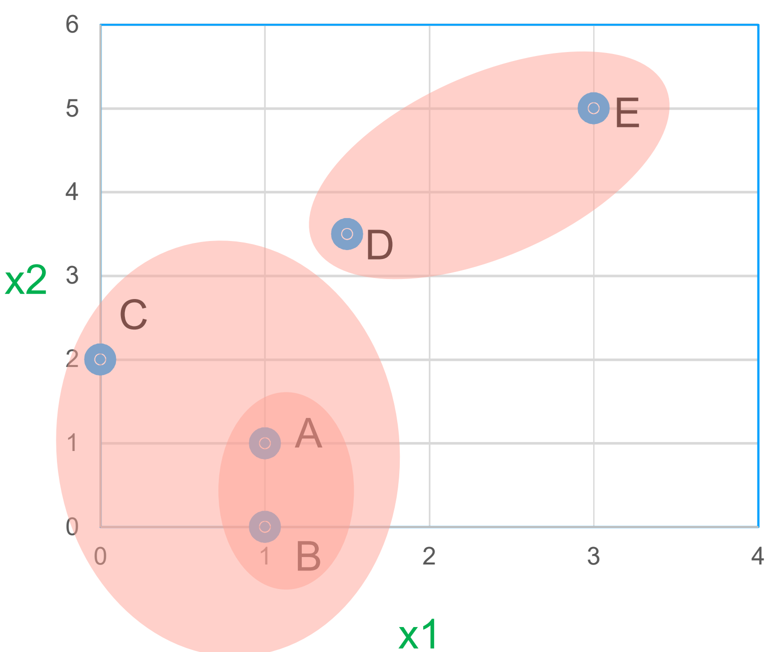


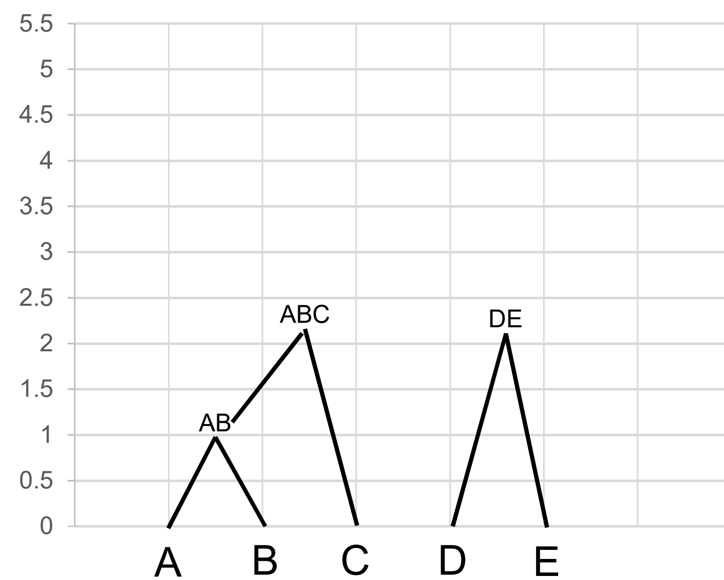
Dendrogram

Distance based on Single Link (Bottom-Up Clustering)

	(A,B)	C	(D,E)
(A,B)	0	2.2	5.4
C	2.2	0	4.2
(D,E)	5.4	4.2	0

	((A,B),C)	(D,E)
((A,B),C)	0	5.4
(D,E)	5.4	0



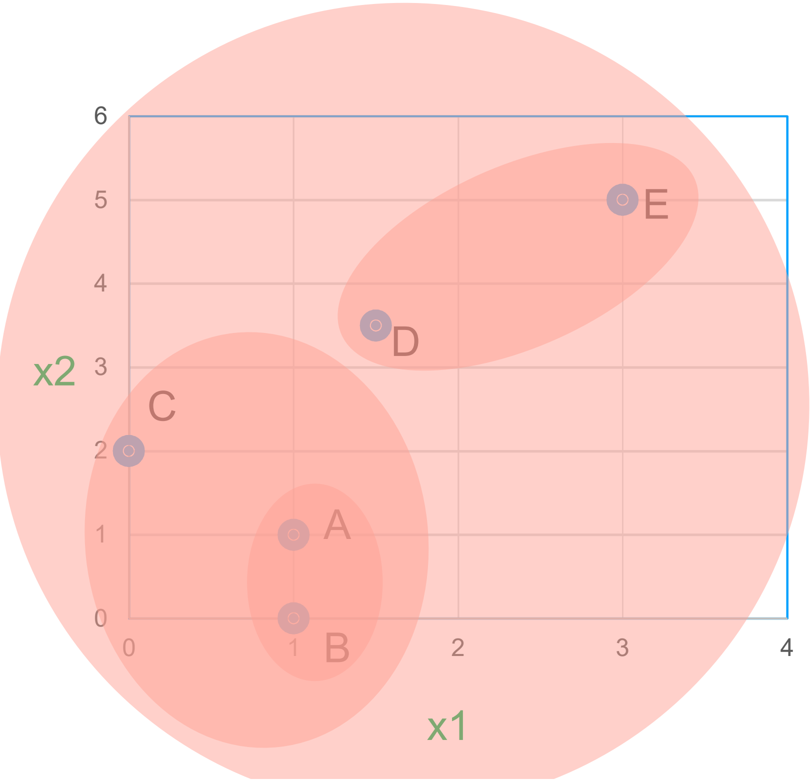


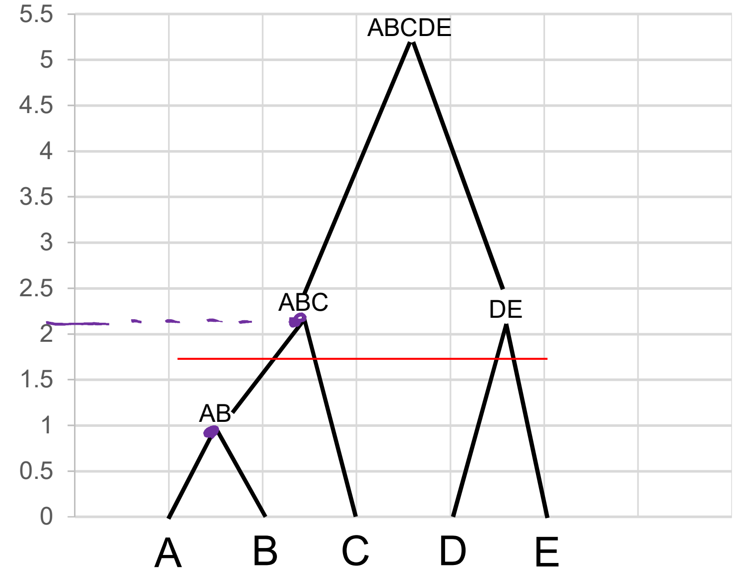
Dendrogram

Distance based on Complete Link (Bottom-Up Clustering)

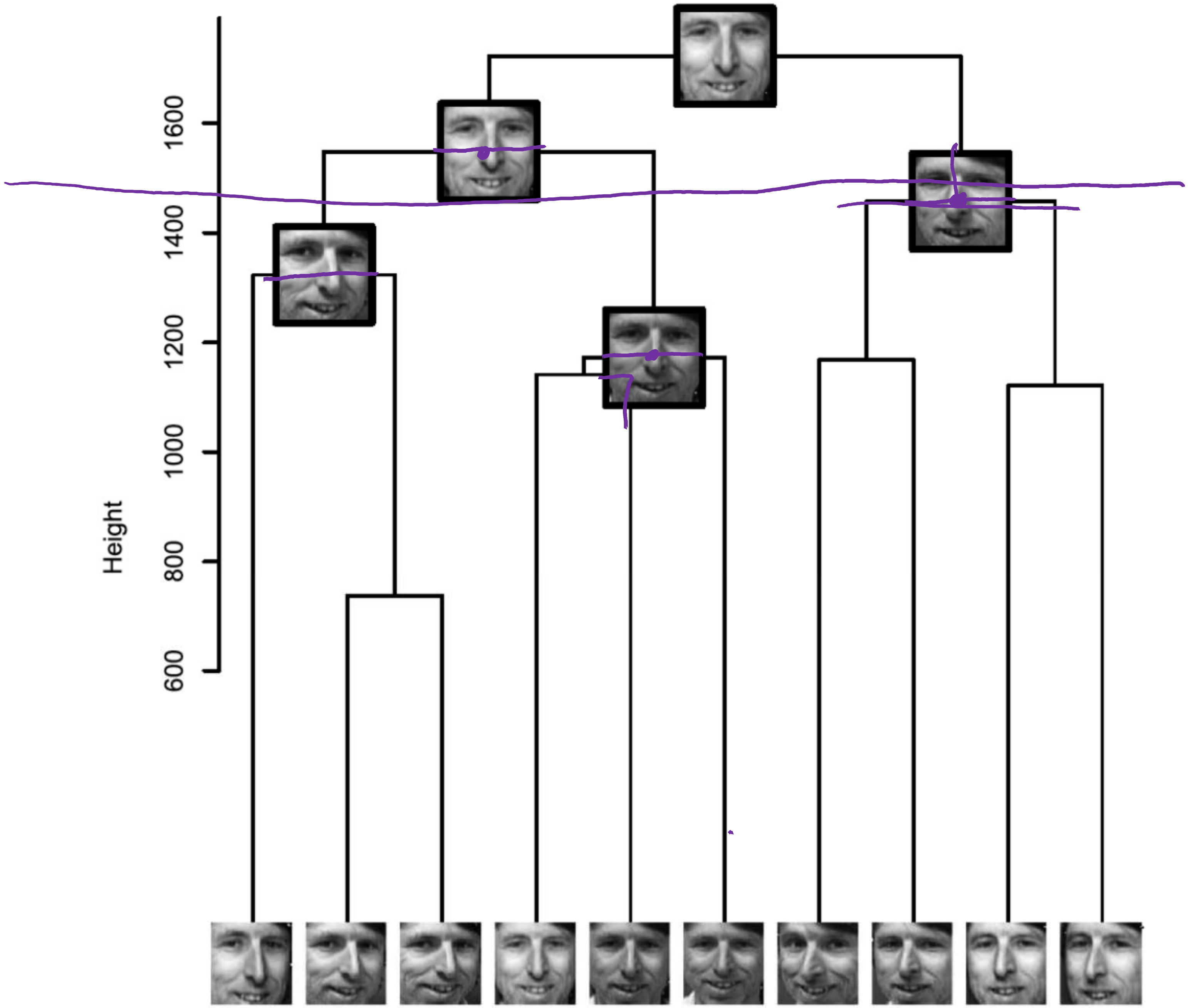
	((A,B),C)	(D,E)
((A,B),C)	0	5.4
(D,E)	5.4	0

	(((A,B),C),(D,E))
(((A,B),C),(D,E))	0



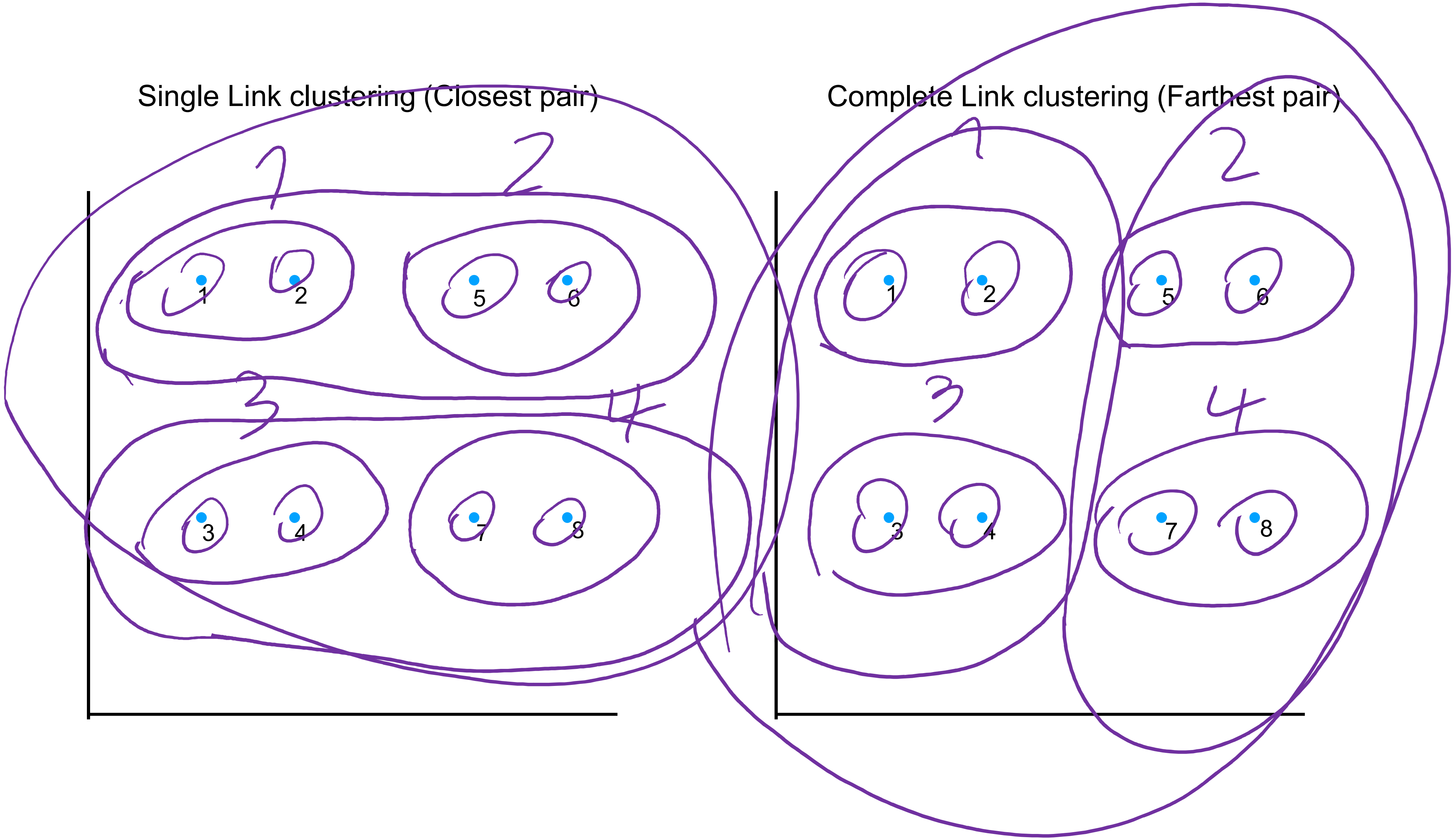


Dendrogram

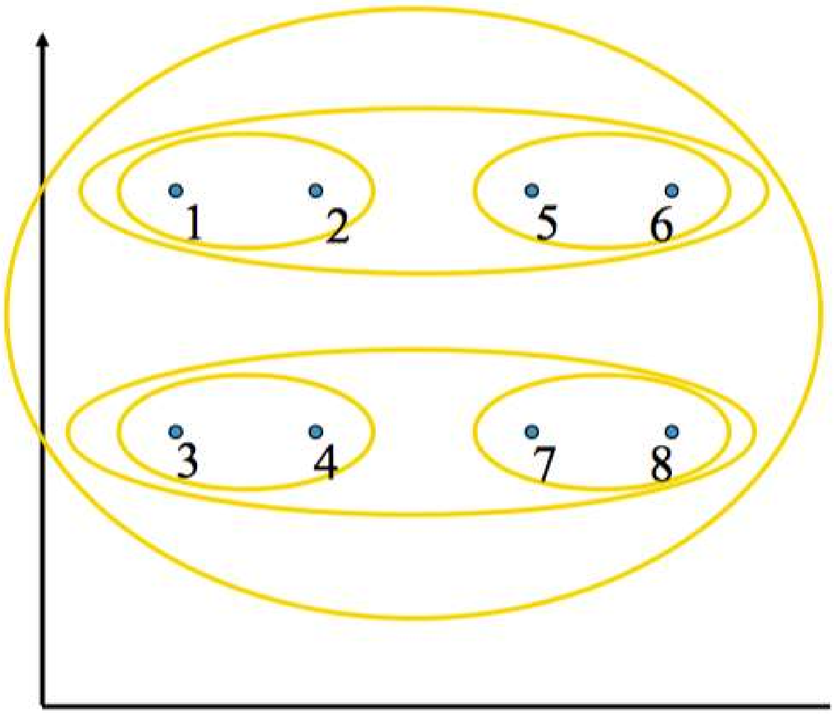


(From Bien et al. (2011))

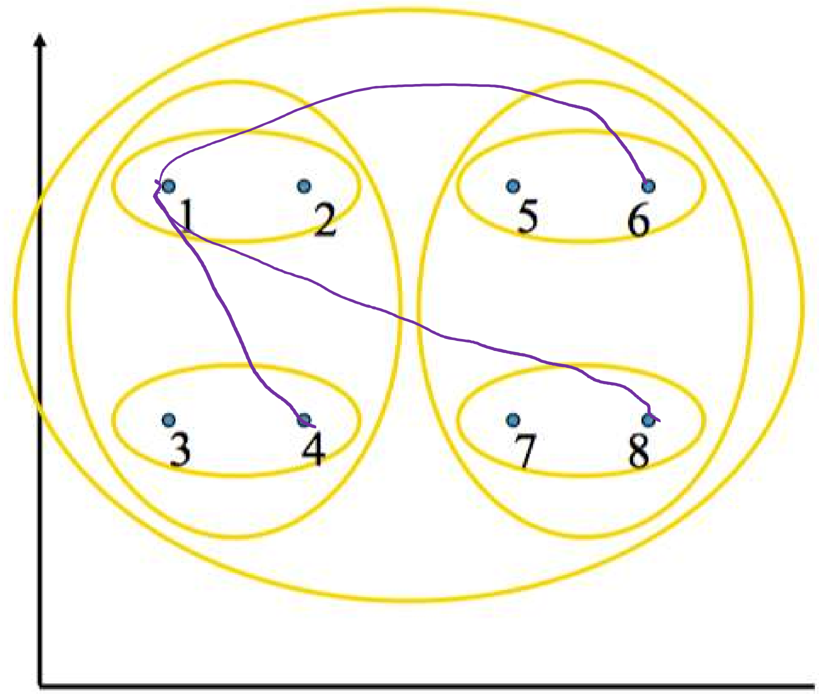
Another Example



Closest pair
(single-link clustering)



Farthest pair
(complete-link clustering)



Slide credit: David Sontag.