SAMPOERNA UNIVERSITY

GENERAL COURSE

COMP3315: Artificial Intelligence

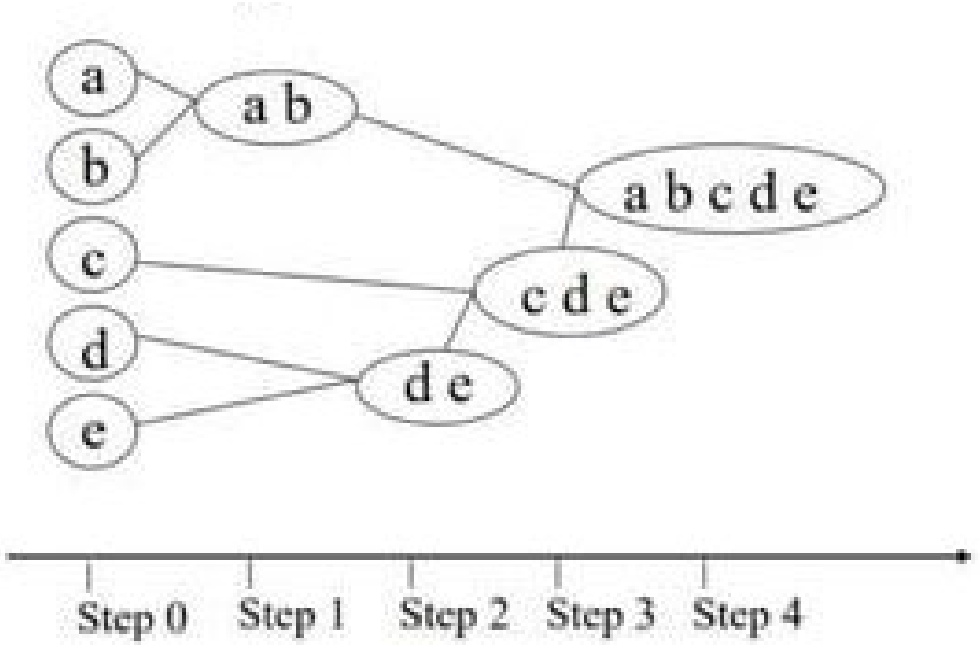
Unsupervised Learning – Agglomerative Hierarchical Clustering

By: Gilang Raka Rayuda Dewa

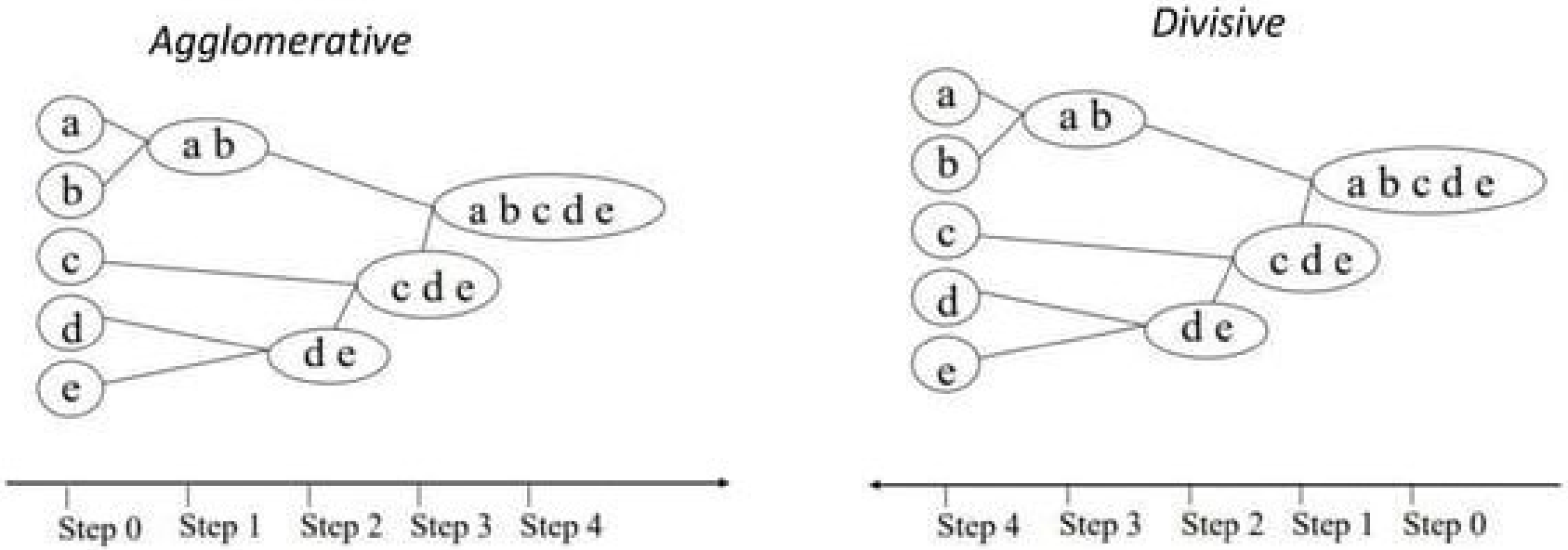
Hierarchical Clustering

- Hierarchical clustering is an unsupervised machine learning algorithm that groups data into a tree of nested clusters.

.



Agglomerative vs Divisive



Initial Dataset

SAMPOERNA UNIVERSITY

$$X=\{(2,3),(5,4),(9,6),(4,7),(8,1)\}$$

We will use Agglomerative clustering

Step 1

Calculate ED (Euclidean distance) of each points

	(2,3)	**(5,4)**	**(9,6)**	**(4,7)**	**(8,1)**
(2,3)	0	3.16	7.62	4.47	6.32
(5,4)	3.16	0	4.47	3.16	4.24
(9,6)	7.62	4.47	0	5.1	5.1
(4,7)	4.47	3.16	5.1	0	7.21
(8,1)	6.32	4.24	5.1	7.21	0

Step 2

Find the smallest point (besides 0) between them. And group them as clusters.

	(2,3)	**(5,4)**	**(9,6)**	**(4,7)**	**(8,1)**
(2,3)	0	3.16	7.62	4.47	6.32
(5,4)	3.16	0	4.47	3.16	4.24
(9,6)	7.62	4.47	0	5.1	5.1
(4,7)	4.47	3.16	5.1	0	7.21
(8,1)	6.32	4.24	5.1	7.21	0

We have two lowest ED here.

Ccand1= (2,3) and (5,4) and Ccand1= (5,4) and (4,7)

Because we use single linkage (=if we have more than one lowest ED, we can choose one randomly), then we choose C1 = Ccand1= (2,3) and (5,4)

C1 = {(2,3), (5,4)}

Remaining Points = {(9,6), (4,7), (8,1)}

Step 3

Find the minimum distance between clusters.

For example:

D(C1,(9,6))=min{ED((2,3),(9,6)),ED((5,4),(9,6))}

- ED (2,3) to (9,6) = 7.62
- ED (5,4) to (9,6) = 4.47

D(C1,(9,6)) = 4.47

Step 4

Find the smallest point and form cluster again.

Here, the smallest value is

- **C1 and (4,7)**

Here, we use (4,7).

C1 = {(2,3), (5,4)}
C2 = {(2,3), (5,4), (4,7)}
Remaining Points = {(9,6), (8,1)}

	C1	(9,6)	(4,7)	(8,1)
C1	0	4.47	3.16	4.24
(9,6)	4.47	0	5.1	5.1
(4,7)	3.16	5.1	0	7.21
(8,1)	4.24	5.1	7.21	0

Step 5

Redo steps 1-4 for the remaining points. Then assign to one of them.

C1 = {(2,3), (5,4)}
C2 = {(2,3), (5,4), (4,7)}
C3 = {(2,3), (5,4), (4,7),(8,1)}
C4 = {(2,3), (5,4), (4,7),(8,1),(9,6)}

Example 1

You have four animals with the following weights (in kg):

Animal	Weight (kg)
Cat	3
Dog	10
Rabbit	5
Goat	20

Using Single Linkage Agglomerative Clustering, if the first merge is between Cat and Rabbit, which cluster will Dog belong to in the next step?

Example 2

Imagine you have four students in a class, and you want to group them based on their height (in cm).

Student	Height (cm)
Alice	150
Bob	160
Charlie	170
David	180

We will use Agglomerative Clustering with Single Linkage to group them step by step.