

AGGLOMERATIVE CLUSTERING

PROBLEM

End Sem (X)

Given the dataset {a, b, c, d, e} and the following distance matrix.

Construct a dendrogram by single linkage, complete linkage, average linkage hierarchical clustering using agglomerative clustering model.

	a	b	c	d	e
a	0	9	3	6	11
b	9	0	7	5	10
c	3	7	0	9	2
d	6	5	9	0	8
e	11	10	2	8	0

Single Linkage :-

$$d(A, B) = \min (d(x, y) : x \in A, y \in B)$$

Complete Linkage :-

$$d(A, B) = \max (d(x, y) : x \in A, y \in B)$$

Average Linkage :-

$$\begin{aligned} d(A, B) &= \text{avg} (d(x, y) : x \in A, y \in B) \\ &= \frac{\sum d(x, y) : x \in A, y \in B}{|A| |B|} \end{aligned}$$

Solution :-

Single Linkage :-

STEP - 1 :- Construct a dendrogram by using the single linkage.

Iteration - 1 :-

	a	b	c	d	e
a	0	9	3	6	11
b	9	0	7	5	10
c	3	7	0	9	2
d	6	5	9	0	8
e	11	10	2	8	0

STEP - 1 :-

Find the minimum distance in the above table.

$$d(c, e) = 2$$

STEP - 2 :-

Cluster the data objects c and e.

	a	b	(c, e)	d
a	0	9	?	6
b	9	0	?	5
(c, e)	?	?	0	?
d	6	5	?	0

STEP - 3 :-

Find the distance for cluster {c, e} using single linkage.

$$\begin{aligned} d(\{c, e\}, \{a\}) &= \min(d(c, a), d(e, a)) \\ &= \min(3, 11) = 3 \end{aligned}$$

$$\begin{aligned} d(\{c, e\}, \{b\}) &= \min(d(c, b), d(e, b)) \\ &= \min(7, 10) = 7. \end{aligned}$$

$$d(\{c, e\}, \{d\}) = \min(d(c, d), d(e, d)) \\ = \min(9, 8) = 8.$$

	a	b	c, e	d
a	0	9	*3	6
b	9	0	*7	5
c, e	* <u>3</u>	*7	0	*8
d	6	5	*8	0

Iteration - 2 :-

STEP - 1 :-

Find the ⁿⁱminimum distance in the above table.

$$d(\{c, e\}, a) = 3$$

STEP - 2 :-

Cluster the data objects c, e and a.

	a, c, e	b	d
a, c, e	0	?	?
b	?	0	5
d	?	5	0

STEP - 3 :-

Find the distance for cluster $\{a, c, e\}$ using single linkage.

$$d(\{a, c, e\}, b) = \min(d(a, b), d(\{c, e\}, b)) \\ = \min(9, 7) = 7.$$

$$d(\{a, c, e\}, d) = \min(d(a, d), d(\{c, e\}, d)) \\ = \min(6, 8) \\ = 6.$$

	ace	b	d
ace	0	*7	*6
b	*7	0	5
d	*6	5	0

Iteration - 3 :-

STEP-1 :-

Find the minimum distance in the above table.

$$d(b, d) = 5$$

STEP-2 :-

We create cluster 3 combining b and d data objects.

	ace	b, d
ace	0	?
b, d	?	0

STEP-3 :-

Find the distance for cluster {b, d} using single linkage.

$$d(\{b, d\}, \{a, c, e\}) = \min(d(b, \{a, c, e\}), d(d, \{a, c, e\}))$$

$$= \min(7, 6) = 6.$$

	ace	bd
ace	0	*6
bd	*6	0

Iteration - 4 :-

STEP - 1 :-

Find the minimum distance in the above table.

$$d(a, b, c, d, e) = 6.$$

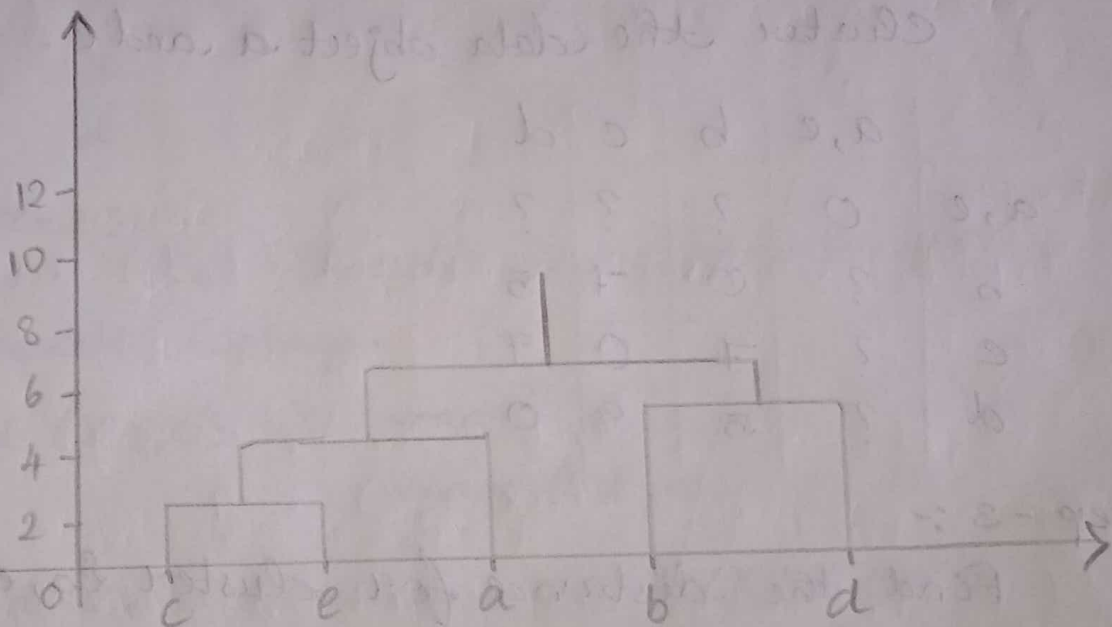
STEP - 2 :-

We create cluster & combining abcde data objects.

	abcde
abcde	6

STEP - 3 :-

Final Dendrogram -



COMPLETE LINKAGE :-

	a	b	c	d	e
a	0	9	3	6	11
b	9	0	7	5	10
c	3	7	0	9	12
d	6	5	9	0	8
e	11	10	2	8	0

Iteration - 1 :-

Step-1 :- Find the minimum distance in the above table

$$d(c, e) = 2$$

Step-2 :- Cluster the data object c and e.

	a	b	c, e	d
a	0	9	?	6
b	9	0	?	5
c, e	?	?	0	?
d	6	5	?	0

STEP-3 :-

Find the distance for cluster $\{c, e\}$ using complete linkage.

$$\begin{aligned} d(\{c, e\}, b) &= \max(d(c, b), d(e, b)) \\ &= \max(7, 10) = 10. \end{aligned}$$

$$\begin{aligned} d(\{c, e\}, d) &= \max(d(c, d), d(e, d)) \\ &= \max(9, 8) = 9 \end{aligned}$$

$$\begin{aligned} d(\{c, e\}, a) &= \max(d(c, a), d(e, a)) \\ &= \max(3, 11) = 11 \end{aligned}$$

	a	b	c, e	d
a	0	9	*11	6
b	9	0	*10	(5)
c, e	*11	*10	0	*9
d	6	5	*9	0

Iteration - 2 :-

STEP-1:- Find the minimum distance in the above table.

$$d(b, d) = 5$$

STEP-2:-

Cluster the data object b, d.

	a	b, d	c, e
a	0	?	11
b, d	?	0	?
c, e	11	?	0

STEP-3:-

Find the distance for cluster {b, d} using Complete linkage.

$$\begin{aligned} d(\{b, d\}, a) &= \max(d(b, a), d(d, a)) \\ &= \max(9, 6) = 9 \end{aligned}$$

$$\begin{aligned} d(\{b, d\}, \{c, e\}) &= \max(d(b, c, e), d(d, c, e)) \\ &= \max(10, 9) = 10. \end{aligned}$$

	a	b, d	c, e
a	0	*9	11
b, d	*9	0	*10
c, e	11	*10	0

Iteration - 3 :-

STEP - 1 :- Find the minimum distance in the above table.

$$d(a, b, d) = 9$$

STEP - 2 :-

Cluster the data object a, b, d.

	abd	ce
abd	0	?
ce	?	0

STEP - 3 :-

Find the distance for cluster {abd} using complete linkage.

$$d(\{a, b, d\}, \{c, e\}) = \max(d(a, c, e), d(b, d, c, e)) \\ = \max(11, 10) = 11$$

	abd	ce
abd	0	* 11
ce	* 11	0

Iteration - 4 :-

Step - 1 :- Find the minimum distance in the above table.

$$d(a, b, c, d, e) = 11$$

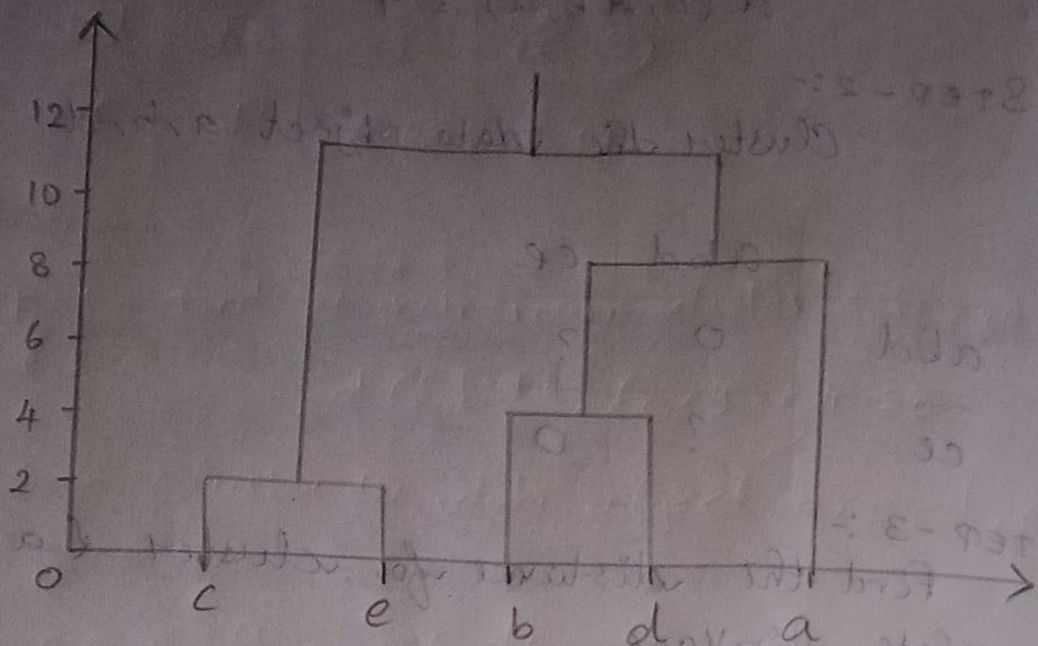
Step - 2 :-

Cluster the data object a, b, c, d, e.

	abcde
abcde	11

STEP - 3 :-

Final Dendrogram.



AVERAGE LINKAGE :-

	a	b	c	d	e
a	0	9	3	6	11
b	9	0	7	5	10
c	3	7	0	9	2
d	6	5	9	0	8
e	11	10	2	8	0

Iteration - 1 :-

Step - 1 :- Find the minimum distance in the above table.

$$d(c, e) = 2$$

Step - 2 :- Cluster the data object. c and e.

	a	b	c, e	d
a	0	9	?	6
b	9	0	?	5
c, e	?	?	0	?
d	6	5	?	0

Step-3:- Find the distance for cluster $\{c, e\}$ using average linkage.

$$d(\{c, e\}, a) = \frac{d(c, a) + d(e, a)}{2 \times 1} = \frac{3 + 11}{2} = \frac{14}{2} = 7.$$

$$d(\{c, e\}, b) = \frac{d(c, b) + d(e, b)}{2 \times 1} = \frac{7 + 10}{2} = \frac{17}{2} = 8.5$$

$$d(\{c, e\}, d) = \frac{d(c, d) + d(e, d)}{2 \times 1} = \frac{9 + 8}{2} = \frac{17}{2} = 8.5.$$

	a	b	c, e	d
a	0	9	*7	6
b	9	0	*8.5	5
c, e	*7	*8.5	0	*8.5
d	6	5	*8.5	0

Iteration - 2:-

Step-1:- Find the minimum distance in the above table.

$$d(b, d) = 5$$

Step-2:- Cluster the data object b and d.

	a	b, d	c, e
a	0	?	7
b, d	?	0	?
c, e	7	?	0

Step-3:- Find the distance for cluster $\{b, d\}$ using average linkage.

$$d(\{b, d\}, a) = \frac{d(b, a) + d(d, a)}{2 \times 1} = \frac{9 + 6}{2} = \frac{15}{2} = 7.5$$

$$d(\{b, d\}, \{c, e\}) = \frac{d(b, c) + d(b, e) + d(d, c) + d(d, e)}{2 \times 2}$$

$$= \frac{7 + 10 + 9 + 8}{4} = \frac{34}{4} = 8.5$$

	a	b, d	c, e
a	0	* 7.5	7
b, d	* 7.5	0	* 8.5
c, e	7	* 8.5	0

Iteration - 3 :-

Step - 1 :- Find the minimum distance from the above table.

$$d(\{c, e\}, a) = 7$$

Step - 2 :- Cluster the data object c, e, a.

	ace	bd
ace	0	?
bd	?	0

Step-3:- Find the distance for cluster {a, b, c, d, e} using average linkage.

$$d(\{a, b, c, d\}, \{e\}) = \frac{d(a, b) + d(a, d) + d(c, b) + d(c, d) + d(e, b) + d(e, d)}{3 \times 2}$$

$$= \frac{9 + 6 + 7 + 9 + 10 + 8}{6} = \frac{49}{6} = 8.16$$

	ace	bd
ace	0	*8.16
bd	*8.16	0

Iteration - 4 :-

Step-1:- Find the minimum distance from the above table.

$$d(\{a, b, c, d, e\}) = 8.16$$

Step-2:- Cluster the data object a, b, c, d, e.

	a b c d e
a b c d e	8.16

Step-3:- Final dendrogram.

