Agglomerative

Hierarchical Clustering

March

Week	M	T	W	T	F	S	S
9					2	3	4
10	5	6	7	8	9	10	11
11	12	13	14	15	16	17	18
12	19	20	21	22	23	24	25
13	26	27	28	29	30	31	

April

Week	M	T	W	T	F	S	S
13/18	30						1
14	2	3	4	5	6	7	8
15	9	10	11	12	13	14	15
16	16	17	18	19	20	21	22
17	23	24	25	26	27	28	29

HAC

Good Friday

Friday 30

(089-276) Wk 13

- Starts with one cluster, individual item in its own cluster and iteratively merge clusters until all the items belong to one cluster.

- Bottom up approach is followed to merge the clusters together.

- Dendrograms are pictorially used to represent HAC.

Methods

- Single - nearest distance (or) Single linkage
- Complete - farthest distance (or) Complete linkage
- Average - average distance (or) Average linkage.

Single linkage - This is the distance b/w the closest members of the two cluster.

Saturday 31

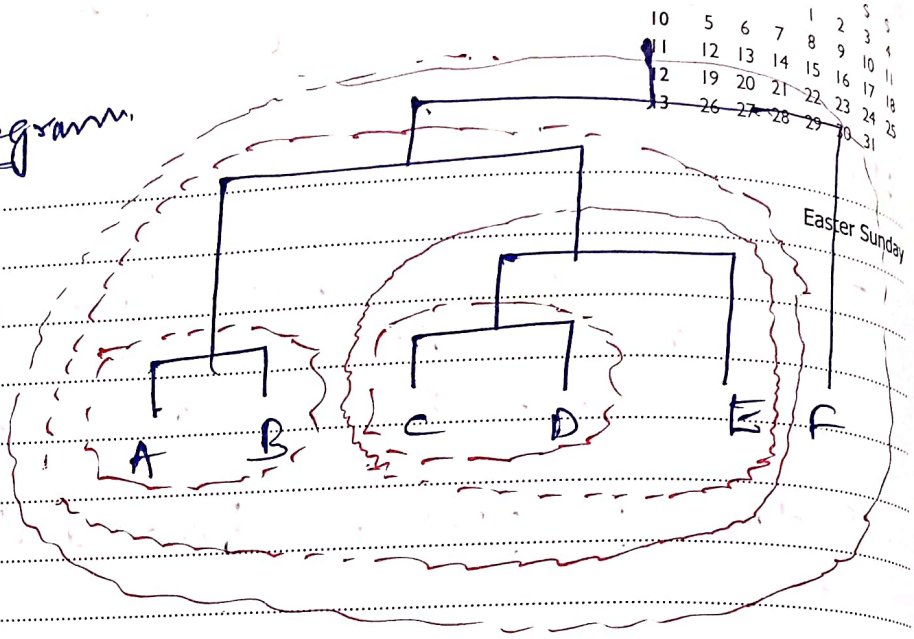
☺ (090-275) Wk 13

Complete linkage - This is the distance b/w the members that are farthest apart.

Average linkage - This method involves looking at the distance b/w all pairs and averages all of these distances.
This is also called Unweighted Pair Group Method Averaging.

April 2018

Sample Dendrogram



1 Sunday

(091-274) Wk 13

- Clusters using a Single link Technique
- Dataset

Sample no	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

2 Monday

Family Day / Easter Monday (Kenya / South Africa)

(092-273) Wk 14

Problem definition:

For the given dataset find the clusters using a single link technique.

Use Euclidean distance and draw the dendrogram.

Solution

Step 1: Compute the distance matrix

- So we have to find the Euclidean distance between each and every points.
- Let $A(x_1, y_1)$ and $B(x_2, y_2)$ are two points
- Then Euclidean distance between

$$d(A,B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

April

Week	M	T	W	T	F	S	S
13/18	30						1
14	2	3	4	5	6	7	8
15	9	10	11	12	13	14	15
16	16	17	18	19	20	21	22
17	23	24	25	26	27	28	29

May

Week	M	T	W	T	F	S	S
18		1	2	3	4	5	6
19	7	8	9	10	11	12	13
20	14	15	16	17	18	19	20
21	21	22	23	24	25	26	27
22	28	29	30	31			

Tuesday 3

(093-272) Wk 14

Sample No	x	y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

Calculate the distance b/w (P1,P2) (P1 P3) (P1,P4) (P1,P5) (P1,P6) (P2,P3) (P2,P3) (P2,P4) (P2,P5) (P2,P6) (P3,P4) (P3,P5) (P3,P6) (P4,P5) & (P4,P6) (P5,P6)

Sample distance calculation

$$d(P1,P2) = \sqrt{(0.22-0.40)^2 + (0.38-0.53)^2}$$

$$= 0.23$$

$$d(P1,P3) = \sqrt{(0.35-0.40)^2 + (0.22-0.53)^2}$$

$$= 0.22$$

Wednesday 4

(094-271) Wk 14

Distance matrix:

	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.23	0				
P3	0.22	0.14	0			
P4	0.37	0.19	0.13	0		
P5	0.34	0.14	0.28	0.23	0	
P6	0.24	0.24	0.10	0.22	0.39	0

P3 P6

Step2: Merging the two closest members

- Here the minimum value is 0.10 and hence we combine P3 and P6 (P3 column & P6 column)
- Now form clusters of elements corresponding to the minimum value & update the distance matrix

5 Thursday

(095-270) Wk 14

iter1

	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.23	0				
P3	0.22	0.14	0			
P4	0.37	0.19	0.13	0		
P5	0.34	0.14	0.28	0.23	0	
P6	0.24	0.24	0.10	0.22	0.39	0

updated distance matrix:

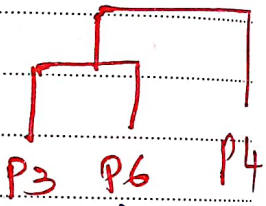
P3 & P6 merged & removed P6 column and row.

	P1	P2	P3 P6	P4	P5
P1	0				
P2	0.23	0			
P3,P6	0.22	0.14	0		
P4	0.37	0.19	0.13	0	
P5	0.34	0.14	0.28	0.23	0

{(P3,P6), P4}

iter2 : 0.13 is the smallest value.

∴ P4 is merged with (P3 & P6) & P4 is removed.



Updated distance matrix is below.

6 Friday

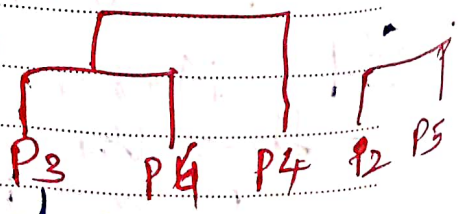
(096-269) Wk 14

	P1	P2	P3 P6 P4	P5
P1	0			
P2	0.23	0		
P3,P6,P4	0.22	0.14	0	
P5	0.24	0.14	0.28	0

{(P3,P6), P4 and (P2,P5)}

0.14 is the minimum value

∴ P5 is merged with P2 & P5 is removed.



Updated distance matrix is below

April

Week	M	T	W	T	F	S	S
13/18	30						1
14	2	3	4	5	6	7	8
15	9	10	11	12	13	14	15
16	16	17	18	19	20	21	22
17	23	24	25	26	27	28	29

May

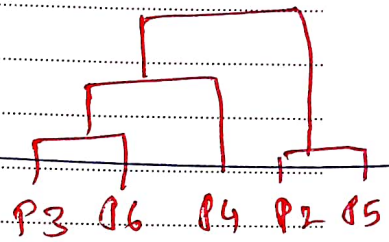
Week	M	T	W	T	F	S	S
18		1	2	3	4	5	6
19	7	8	9	10	11	12	13
20	14	15	16	17	18	19	20
21	21	22	23	24	25	26	27
22	28	29	30	31			

	P1	P2 P5	P3 P6 P4
P1	0		
P2,P5	0.23	0	
P3,P6,P4	0.22	0.14	0

Saturday 7

(097-268) Wk 14

iter3: {{(P3,P6),P4},(P2,P5)}



∴ (P3, P6), P4 is merged with P2,P5 &
(P3,P2),P4 is removed.

	P1	{(P3 P6), P4}	{P2 P5}
P1	0		
{(P3,P6),P4} {P2 P5}	0.22	0	

Sunday 8

(098-267) Wk 14

(P3,P6) P4 {P2,P5} is the minimum value

∴ (P3,P6) P4 {P2 P5} is merged with P1 &
(P3,P6) P4 {P2 P5} is removed

Updated distance matrix is

	P1	{P3 P6, P4}	(P2 P5)
{P1, (P2 P6) P4}	0		
{P2 P5}