

# M.L Assignment - 6

Student ID: 700740230

15/11/2022

Name: Hanke SAGAR

USER ID: SXL02300

Question - 1: (Provide only mathematical Solutions for this Question) Six points with the following attributes are given, Calculate and find out Clustering representation and dendrogram Using Single, Complete and average link proximity function in hierarchical clustering technique.

Ans (Solution) : Mathematical Solution

Given

Point X-Coordinate Y-Coordinate

$P_1$  0.04005 0.5306

$P_2$  0.2148 0.3854

$P_3$  0.3457 0.3156

$P_4$  0.2652 0.1875

$P_5$  0.0789 0.4139

$P_6$  0.4548 0.3022

Single link proximity: Distance between two Clusters is the distance between the closest points. Also called "neighbour joining".  
Distance matrix:

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	P <sub>6</sub>
P <sub>1</sub>	0	0.2357	0.2218	0.3688	0.3421	0.2347
P <sub>2</sub>	0.2357	0	0.1483	0.2042	0.1388	0.254
P <sub>3</sub>	0.2218	0.1483	0	0.1513	0.2843	0.11
P <sub>4</sub>	0.3688	0.2042	0.1513	0	0.2932	0.2216
P <sub>5</sub>	0.3421	0.1388	0.2843	0.2932	0	0.3921
P <sub>6</sub>	0.2347	0.254	0.11	0.2216	0.3921	0

table - 1

The Smallest distance from above table-1 is 0.11

\* So P<sub>3</sub> and P<sub>6</sub> form the first Cluster.

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub> P <sub>6</sub>	P <sub>4</sub>	P <sub>5</sub>
P <sub>1</sub>	0	0.2357	0.2218	0.3688	0.3421
P <sub>2</sub>	0.2357	0	0.1483	0.2042	0.1388
P <sub>3</sub> P <sub>6</sub>	0.2218	0.1483	0	0.1513	0.2843
P <sub>4</sub>	0.3688	0.2042	0.1513	0	0.2932
P <sub>5</sub>	0.3421	0.1388	0.2843	0.2932	0

table - 2

The Smallest distance from table-2 is 0.1388

\*  $P_2$  and  $P_5$  forms 2<sup>nd</sup> Cluster

	$P_1$	$P_1P_5$	$P_3P_6$	$P_4$
$P_1$	0	0.2357	0.2218	0.3688
$P_2P_5$	0.2357	0	0.1483	0.2042
$P_3P_6$	0.2218	0.1483	0	0.1513
$P_4$	0.3688	0.2042	0.1513	0

table-3

The Smallest distance from table-3 is 0.1483

\*  $P_2P_5$  forms the 3<sup>rd</sup> Cluster

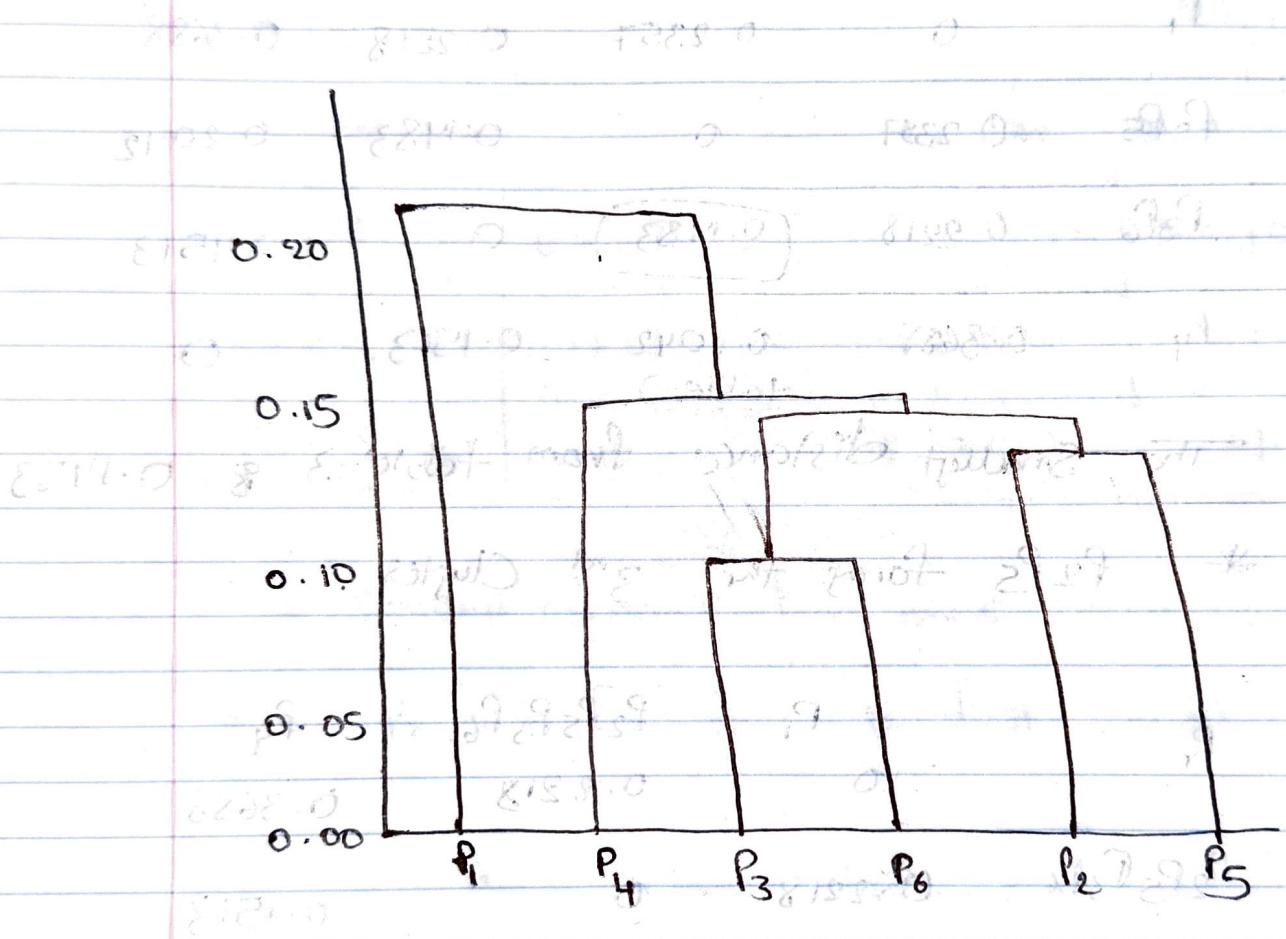
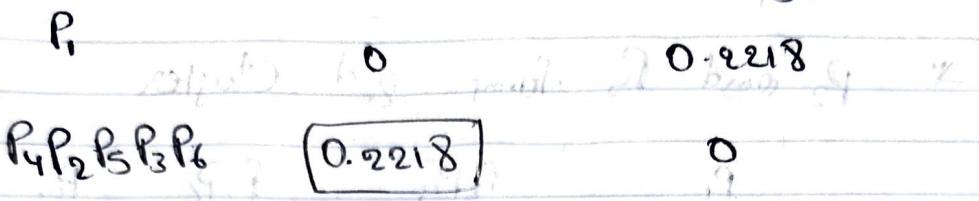
	$P_1$	$P_2P_5P_3P_6$	$P_4$
$P_1$	0	0.2218	0.3688
$P_2P_5P_3P_6$	0.2218	0	0.1513
$P_4$	0.3688	0.1513	0

table-4

The Smallest distance from table-4 is 0.1513

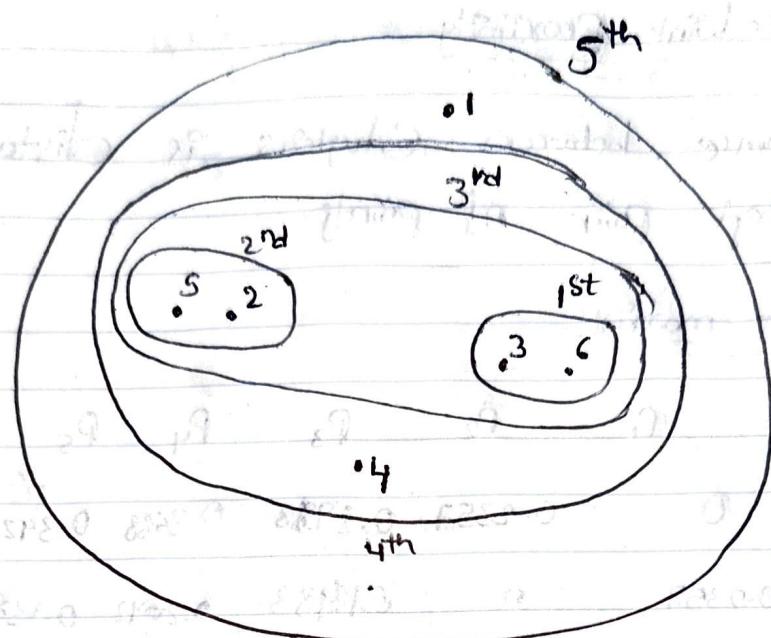
\*  $P_2P_5P_3P_6$  forms the 4<sup>th</sup> Cluster

602.0 7.5 sample  $P_1$  first sample  $P_4 P_2 P_5 P_3 P_6$



Single Linkage Dendrogram

Representation



### Single Link Clustering Representation

1100 0011 + 1000 0010 = 1100 0011 0000 0010

0010 1000 0011 + 0000 0010 = 0010 1000 0011 0000 0010

1100 0011 0000 0010 0000 0010 + 0000 0010 0000 0010 = 1100 0011 0000 0010 0000 0010

1100 0011 0000 0010 0000 0010 + 0000 0010 0000 0010 = 1100 0011 0000 0010 0000 0010

1100 0011 0000 0010 0000 0010 + 0000 0010 0000 0010 = 1100 0011 0000 0010 0000 0010

1100 0011 0000 0010 0000 0010 + 0000 0010 0000 0010 = 1100 0011 0000 0010 0000 0010

## Complete Link Proximity:

Distance between Clusters is distance between farthest pair of points.

### Distance matrix:

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	P <sub>6</sub>
P <sub>1</sub>	0	0.2357	0.2913	0.3688	0.3421	0.2347
P <sub>2</sub>	0.2357	0	0.1483	0.2042	0.1388	0.254
P <sub>3</sub>	0.2913	0.1483	0	0.1513	0.2843	0.11
P <sub>4</sub>	0.3688	0.2042	0.1513	0	0.2932	0.2216
P <sub>5</sub>	0.3421	0.1388	0.2843	0.2932	0	0.3921
P <sub>6</sub>	0.2347	0.254	0.11	0.2216	0.3921	0

table - 1

The smallest distance from table 1 is 0.11

∴ So P<sub>3</sub> and P<sub>6</sub> form 1<sup>st</sup> cluster

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub> , P <sub>6</sub>	P <sub>4</sub>	P <sub>5</sub>
P <sub>1</sub>	0	0.2357	0.2347	0.3688	0.3421
P <sub>2</sub>	0.2357	0	0.254	0.2042	0.1388
P <sub>3</sub> , P <sub>6</sub>	0.2347	0.254	0	0.2216	0.3921
P <sub>4</sub>	0.3688	0.2042	0.2216	0	0.2932
P <sub>5</sub>	0.3421	0.1388	0.3921	0.2932	0

table - 2

The Smallest distance from table-2 is 0.1388

\*  $P_2$  and  $P_5$  forming 2<sup>nd</sup> Cluster.

$P_i$	$P_1$	$P_2P_5$	$P_3P_6$	$P_4$
$P_2P_5$	0	0.3421	0.2347	0.3688
$P_3P_6$		0	0.3921	0.2932
$P_4$	0.3688	0.2932	0.2216	0

table-3

The Smallest distance from table-3 is 0.2216

\*  $P_2P_5$  and  $P_3P_6$  forming 3<sup>rd</sup> cluster.

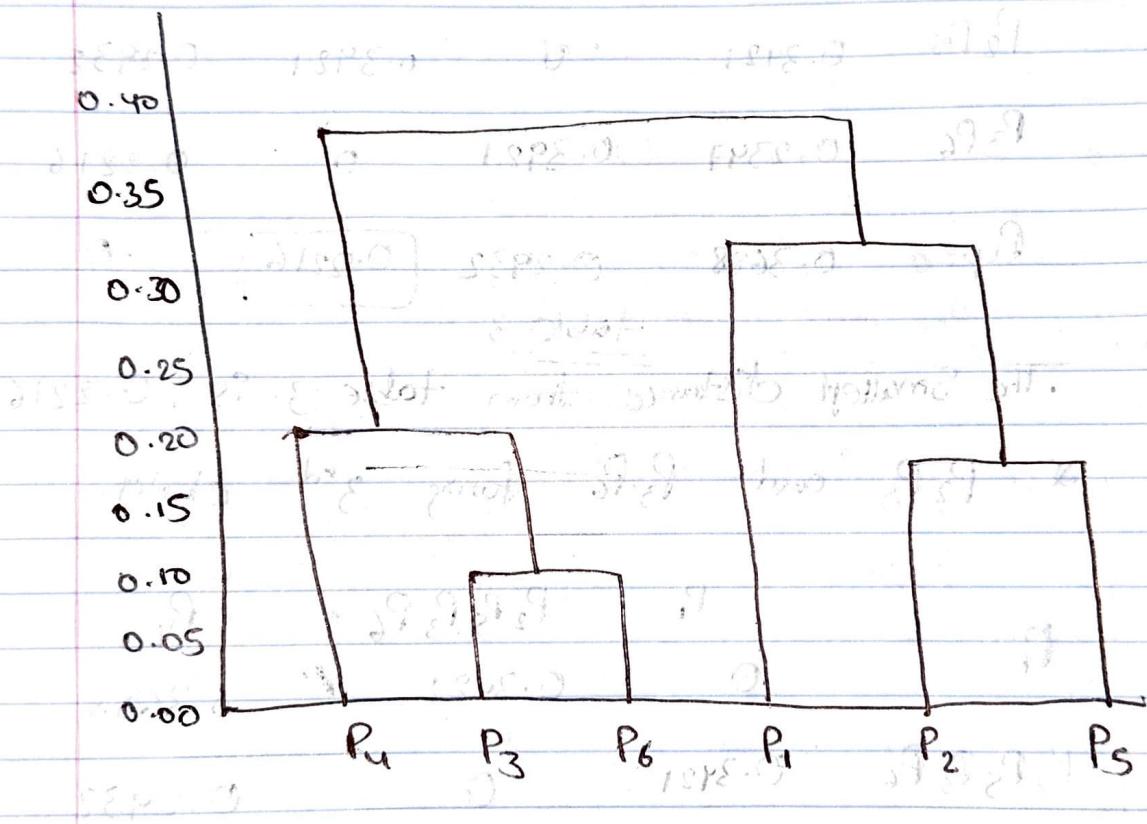
$P_i$	$P_1$	$P_2P_5P_3P_6$	$P_4$
$P_2P_5P_3P_6$	0	0.3421	0.3688
$P_4$		0	0.2932

table-4

The Smallest distance from table-4 is 0.2932

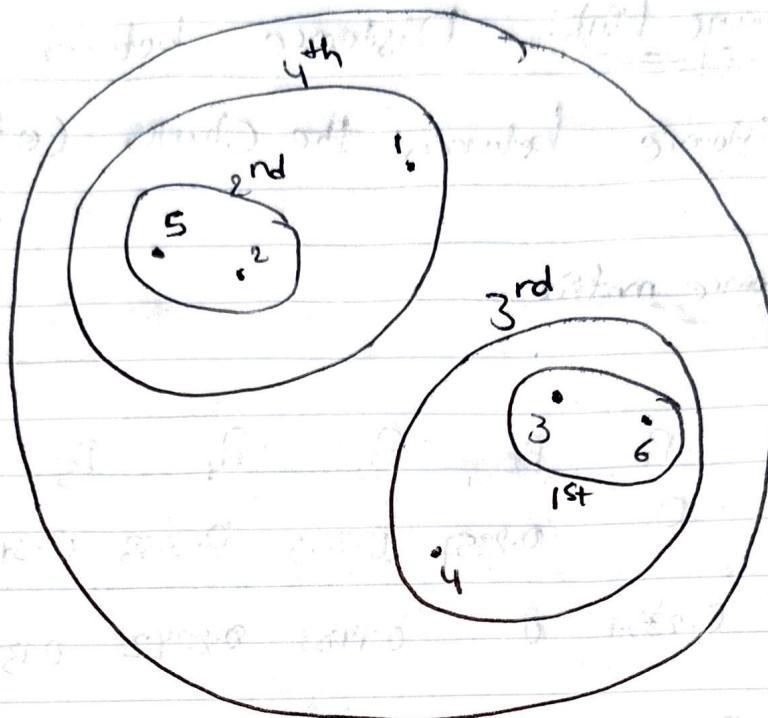
:  $P_2P_5P_3P_6$  and  $P_1$  forms the 4<sup>th</sup> cluster.

$P_1 P_2 P_3 P_4 P_5 P_6$   
 $P_1 P_2 P_3 P_4 P_5 P_6$       0.1483  
 0.3688



Complete linkage Dendrogram

Representation



Complete Linkage Clustering Representation

Average Link: Distance between Clusters is distance between the Cluster Centroid.

Distance matrix:

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	P <sub>6</sub>
P <sub>1</sub>	0	0.2357	0.2218	0.3688	0.3421	0.2347
P <sub>2</sub>	0.2357	0	0.1483	0.2042	0.1388	0.254
P <sub>3</sub>	0.2218	0.1483	0	0.1513	0.2843	0.11
P <sub>4</sub>	0.3688	0.2042	0.1513	0	0.2932	0.2216
P <sub>5</sub>	0.3421	0.1388	0.2843	0.2932	0	0.3921
P <sub>6</sub>	0.2347	0.254	0.11	0.7816	0.3921	0

The Smaller distance from table-1 is 0.11

∴ P<sub>3</sub> and P<sub>6</sub> form the 1<sup>st</sup> Cluster.

	$P_1$	$P_2$	$P_3 P_6$	$P_4$	$P_5$
$P_1$	0.8108	0.2357	0.22825	0.3688	0.3421
$P_2$	0.2357	0	0.2015	0.2042	0.1388
$P_3 P_6$	0.22825	0.20115	0	0.18645	0.3382
$P_4$	0.3688	0.2042	0.18645	0	0.2932
$P_5$	0.3421	0.1388	0.3382	0.2932	0

table-2

The Smallest distance from table-2 is 0.1388

\*  $P_2$  and  $P_5$  forms the 2<sup>nd</sup> Cluster.

	$P_1$	$P_2 P_5$	$P_3 P_6$	$P_4$
$P_1$	0	0.2889	0.2347	0.3688
$P_2 P_5$	0.2889	0	0.269675	0.2487
$P_3 P_6$	0.2347	0.269675	0	0.18645
$P_4$	0.3688	0.2487	0.18645	0

table-3

The Smallest distance from table-3 is 0.18645

\*  $P_2 P_5$  and  $P_3 P_6$  forms 3<sup>rd</sup> Cluster.

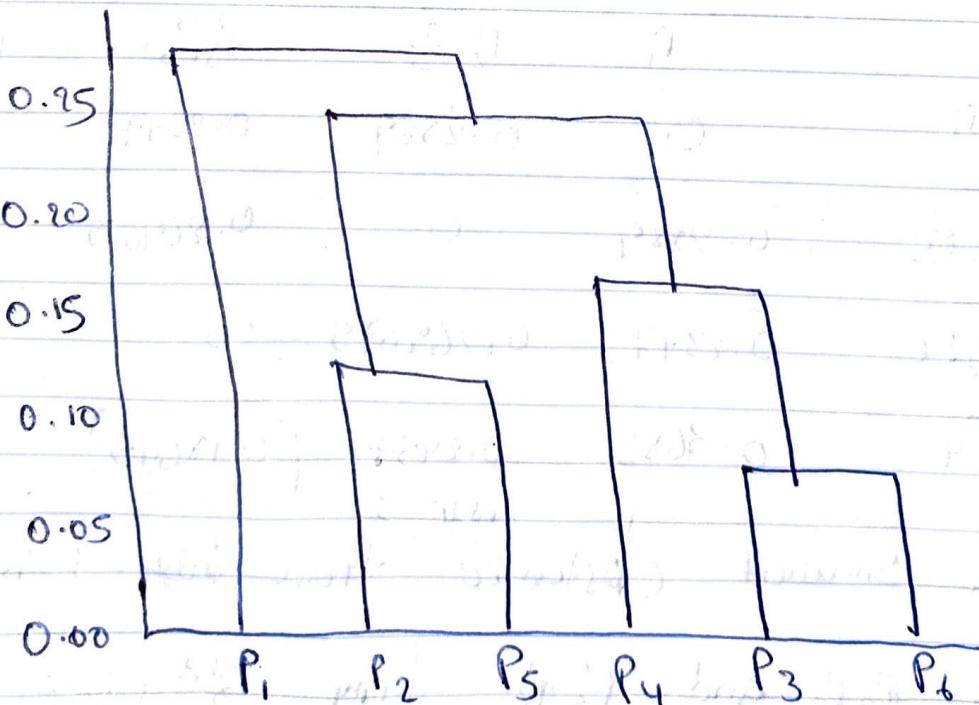
	$P_1$	$P_2P_5P_3P_6$	$P_4$
$P_1$	0	0.2618	0.3688
$P_2P_5P_3P_6$	0.2618	0	0.217575
$P_4$	0.3688	0.217575	0

Table - 4

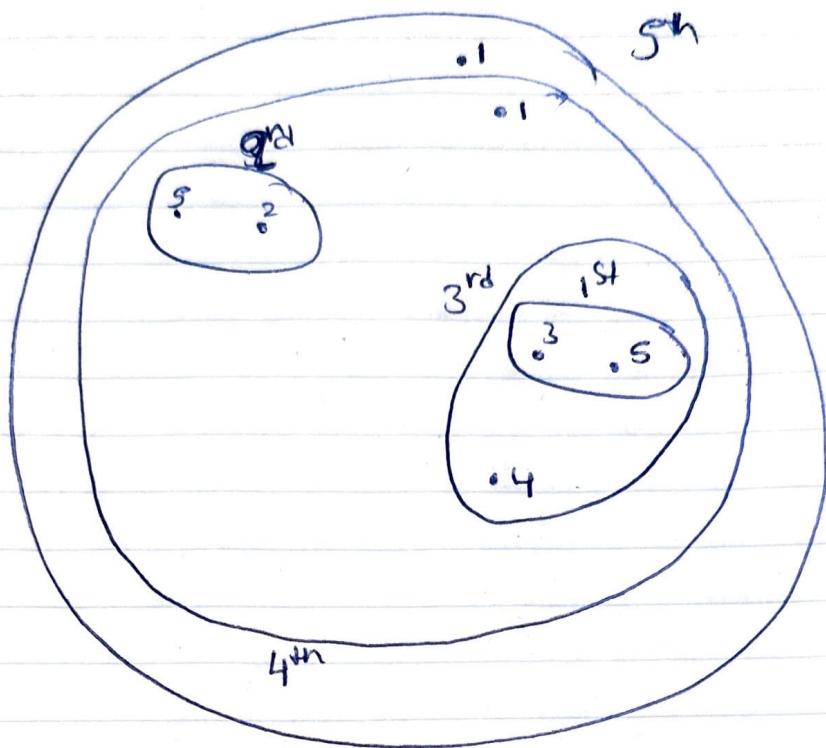
The Smallest distance from table - 4 is 0.217575

$P_2P_5P_3P_6P_4$ , form the 4<sup>th</sup> cluster.

	$P_2P_5P_3P_6$	$P_1$
$P_2P_5P_3P_6$	0	0.3153
$P_1$	0.3153	0



Average Link Dendrogram Representation



Average link Clustering Repetition