Question – 1

Question:- Six points with the following attributes are given, calculate and find out clustering representations and dendrogram using Single, complete, and average link proximity function in hierarchical clustering technique.

point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

Table : X-Y coordinates of six points.

	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points

Answer:-

Single link:

p3 and p6 have the minimum distance from the above table. So, distance(p3,p6) is minimum.

Now we need to update the distance matrix.

To update the distance matrix MIN[distance((p3,p6),p1)]:

   MIN[distance((p3,p1),(p6,p1))] = MIN[0.2218,0.2347] = 0.2218

To update the distance matrix MIN[distance((p3,p6),p2)]:

   MIN[distance((p3,p2),(p6,p2))] = MIN[0.1483,0.2540] = 0.1483

To update the distance matrix MIN[distance((p3,p6),p4)]:

   MIN[distance((p3,p4),(p6,p4))] = MIN[0.1513,0.2216] = 0.1513

To update the distance matrix MIN[distance((p3,p6),p5)]:

   MIN[distance((p3,p5),(p6,p5))] = MIN[0.2843,0.3921] = 0.2843

Updated distance matrix for cluster p3,p6

	p1	p2	p3,p6	p4	p5
p1	0	0.2357	0.2218	0.3688	0.3421
p2	0.2357	0	0.1483	0.2042	0.1388
p3,p6	0.2218	0.1483	0	0.1513	0.2843
p4	0.3688	0.2042	0.1513	0	0.2932
p5	0.3421	0.1388	0.2843	0.2932	0

p2 and p5 have the minimum distance from the above table. So, distance(p2,p5) is minimum.

Now we need to update the distance matrix.

To update the distance matrix MIN[distance((p2,p5),p1)]:

MIN[distance((p2,p1),(p5,p1))] = MIN[0.2357,0.3421] = 0.2357

To update the distance matrix MIN[distance((p2,p5),(p3,p6))]:

MIN[distance((p2,(p3,p6)),(p5,(p3,p6)))] = MIN[0.1483,0.2843] = 0.1483

To update the distance matrix MIN[distance((p2,p5),p4)]:

MIN[distance((p2,p4),(p5,p4))] = MIN[0.2042,0.2932] = 0.2042

Updated distance matrix for cluster p2,p5

	p1	p2,p5	p3,p6	p4
p1	0	0.2357	0.2218	0.3688
p2,p5	0.2357	0	0.1483	0.2042
p3,p6	0.2218	0.1483	0	0.1513
p4	0.3688	0.2042	0.1513	0

(p2,p5) and (p3,p6) have the minimum distance from the above table. So, distance((p2,p5),(p3,p6)) is minimum.

Now we need to update the distance matrix.

To update the distance matrix MIN[distance(((p2,p5),(p3,p6)),p1)]:

MIN[distance(((p2,p5),p1),((p3,p6),p1))] = MIN[0.2357,0.2218] = 0.2218

To update the distance matrix MIN[distance(((p2,p5),(p3,p6)),p4)]:

MIN[distance(((p2,p5),p4),((p3,p6),p4))] = MIN[0.2042,0.1513] = 0.1513

Updated distance matrix for cluster p2,p5,p3,p6

	p1	p2,p5,p3,p6	p4
p1	0	0.2218	0.3688
p2,p5,p3,p6	0.2218	0	0.1513
p4	0.3688	0.1513	0

(p2,p5,p3,p6) and p4 have the minimum distance from the above table. So, distance((p2,p5,p3,p6),p4) is minimum.
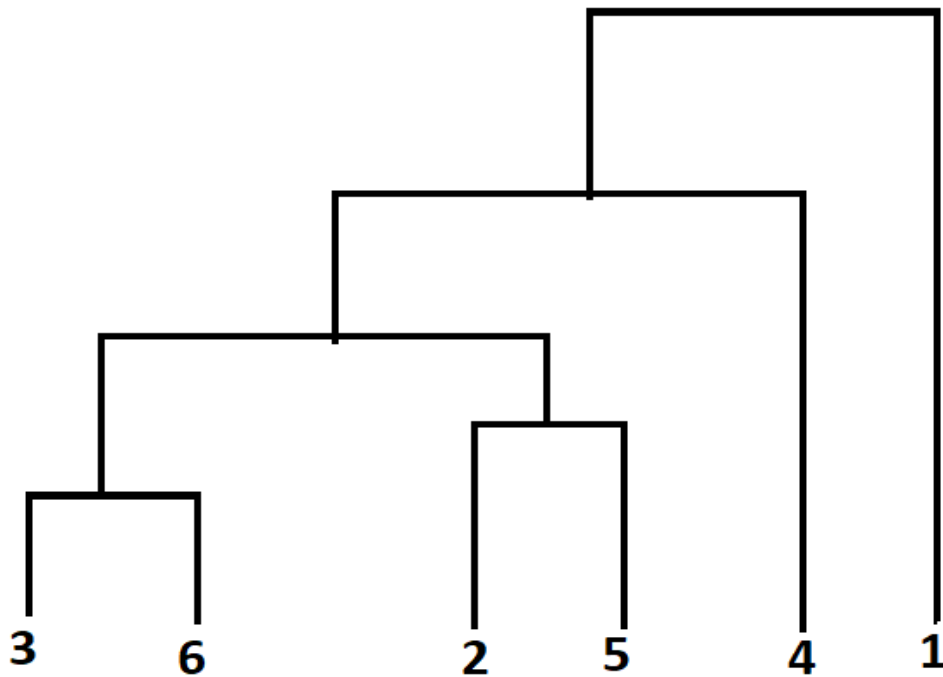
Now we need to update the distance matrix.

To update the distance matrix MIN[distance(((p2,p5,p3,p6),p4),p1)]:

MIN[distance((p2,p5,p3,p6),p1),(p4,p1))] = MIN[0.2218,0.3688] = 0.2218

	p1	p2,p5,p3,p6,p4
p1	0	0.2218
p2,p5,p3,p6,p4	0.2218	0

Dendrogram for Single link:

Complete link:

p3 and p6 have the minimum distance from the above table. So, distance(p3,p6) is minimum.

Now we need to update the distance matrix.

To update the distance matrix MAX[distance((p3,p6),p1)]:

MAX[distance((p3,p1),(p6,p1))] = MAX[0.2218,0.2347] = 0.2347

To update the distance matrix MAX[distance((p3,p6),p2)]:

MAX[distance((p3,p2),(p6,p2))] = MAX[0.1483,0.2540] = 0.2540

To update the distance matrix MAX[distance((p3,p6),p4)]:

MAX[distance((p3,p4),(p6,p4))] = MAX[0.1513,0.2216] = 0.2216

To update the distance matrix MAX[distance((p3,p6),p5)]:

MAX[distance((p3,p5),(p6,p5))] = MAX[0.2843,0.3921] = 0.3921

Updated distance matrix for cluster p3,p6

	p1	p2	p3,p6	p4	p5
p1	0	0.2357	0.2357	0.3688	0.3421
p2	0.2357	0	0.2540	0.2042	0.1388
p3,p6	0.2357	0.2540	0	0.2216	0.3921
p4	0.3688	0.2042	0.2216	0	0.2932
p5	0.3421	0.1388	0.3921	0.2932	0

p2 and p5 have the minimum distance from the above table. So, distance(p2,p5) is minimum.

Now we need to update the distance matrix.

To update the distance matrix MAX[distance((p2,p5),p1)]:

MAX[distance((p2,p1),(p5,p1))] = MAX[0.2357,0.3421] = 0.3421

To update the distance matrix MAX[distance((p2,p5),(p3,p6))]:

MAX[distance((p2,(p3,p6)),(p5,(p3,p6)))] = MAX[0.2540,0.3921] = 0.3921

To update the distance matrix MAX[distance((p2,p5),p4)]:

MAX[distance((p2,p4),(p5,p4))] = MAX[0.2042,0.2932] = 0.2932

Updated distance matrix for cluster p2,p5

	p1	p2,p5	p3,p6	p4
p1	0	0.3421	0.2357	0.3688
p2,p5	0.3421	0	0.3921	0.2932
p3,p6	0.2357	0.3921	0	0.2216
p4	0.3688	0.2932	0.2216	0

(p3,p6) and p4 have the minimum distance from the above table. So, distance((p3,p6),p4) is minimum.

Now we need to update the distance matrix.

To update the distance matrix MAX[distance(((p3,p6),p4),p1)]:

MAX[distance(((p3,p6),p1),(p4,p1))] = MAX[0.2357,0.3688] = 0.3688

To update the distance matrix MAX[distance(((p3,p6),p4),(p2,p5))]:

MAX[distance(((p3,p6),(p2,p5)),(p4,(p2,p5)))] = MAX[0.3921,0.2932] = 0.3921

Updated distance matrix for p3,p6,p4

	p1	p2,p5	p3,p6,p4
p1	0	0.3421	0.3688
p2,p5	0.3421	0	0.3921
p3,p6,p4	0.3688	0.3921	0

(p2,p5) and p1 have the minimum distance from the above table. So, distance((p2,p5),p1) is minimum.
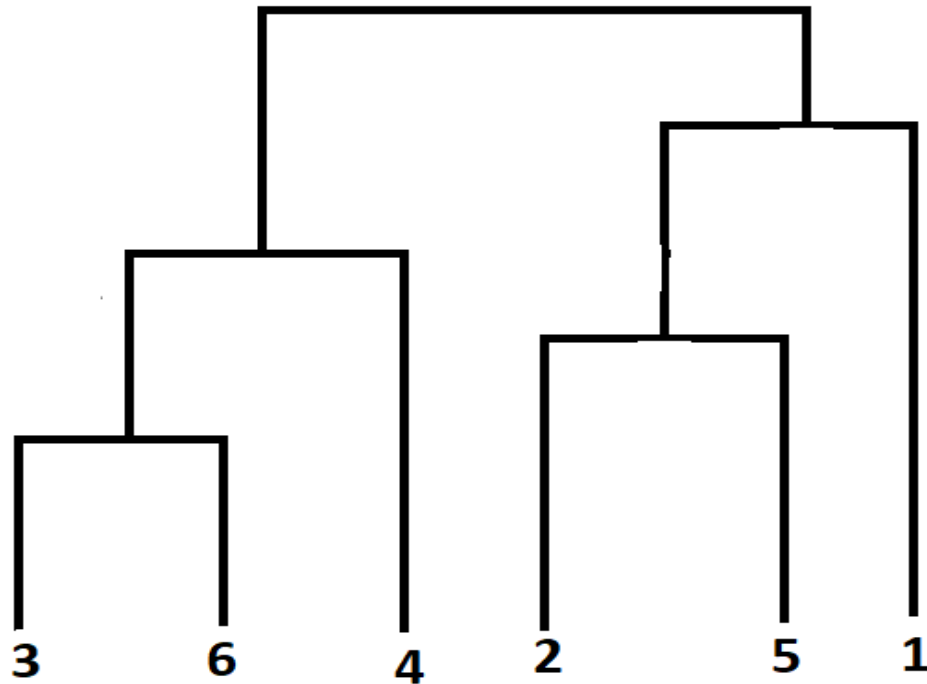
Now we need to update the distance matrix.

To update the distance matrix MAX[distance(((p2,p5),p1),(p3,p6,p4))]:

MAX[distance(((p2,p5),(p3,p6,p4)),(p1,(p3,p6,p4)))] = MAX[0.3921,0.3688] = 0.3921

Updated distance matrix for p2,p5,p1

	p2,p5,p1	p3,p6,p4
p2,p5,p1	0	0.3921
p3,p6,p4	0.3921	0

Dendrogram for Complete link:

Average link:

p3 and p6 have the minimum distance from the above table. So, distance(p3,p6) is minimum.

Now we need to update the distance matrix.

To update the distance matrix AVG[distance((p3,p6),p1)]:

   AVG[distance((p3,p1),(p6,p1))] = ½ * [0.2218+0.2347] = 0.2282

To update the distance matrix MAX[distance((p3,p6),p2)]:

   AVG[distance((p3,p2),(p6,p2))] = ½ * [0.1483+0.2540] = 0.2011

To update the distance matrix MAX[distance((p3,p6),p4)]:

   AVG[distance((p3,p4),(p6,p4))] = ½ * [0.1513+0.2216] = 0.1864

To update the distance matrix MAX[distance((p3,p6),p5)]:

   AVG[distance((p3,p5),(p6,p5))] = ½ * [0.2843+0.3921] = 0.3382

Updated distance matrix for cluster p3,p6

	p1	p2	p3,p6	p4	p5
p1	0	0.2357	0.2282	0.3688	0.3421
p2	0.2357	0	0.2011	0.2042	0.1388
p3,p6	0.2282	0.2011	0	0.1864	0.3382
p4	0.3688	0.2042	0.1864	0	0.2932
p5	0.3421	0.1388	0.3382	0.2932	0

p2 and p5 have the minimum distance from the above table. So, distance(p2,p5) is minimum.

Now we need to update the distance matrix.

To update the distance matrix AVG[distance((p2,p5),p1)]:

AVG[distance((p2,p1),(p5,p1))] = ½ * [0.2357+0.3421] = 0.2889

To update the distance matrix MAX[distance((p2,p5),(p3,p6))]:

AVG[distance((p2,(p3,p6)),(p5,(p3,p6)))] = ½ * [0.2011+0.3382] = 0.2696

To update the distance matrix MAX[distance((p2,p5),p4)]:

AVG[distance((p2,p4),(p5,p4))] = ½ * [0.2042+0.2932] = 0.2487

Updated distance matrix for cluster p2,p5

	p1	p2,p5	p3,p6	p4
p1	0	0.2889	0.3688	0.3688
p2,p5	0.2889	0	0.2696	0.2487
p3,p6	0.2282	0.2696	0	0.1864
p4	0.3688	0.2487	0.1864	0

(p3,p6) and p4 have the minimum distance from the above table. So, distance((p3,p6),p4) is minimum.

Now we need to update the distance matrix.

To update the distance matrix AVG[distance(((p3,p6),p4),p1)]:

AVG[distance(((p3,p6),p1),(p4,p1))] = ½ * [0.2282+0.3688] = 0.2985

To update the distance matrix AVG[distance(((p3,p6),p4),(p2,p5))]:

AVG[distance(((p3,p6),(p2,p5)),(p4,(p2,p5)))] = ½ * [0.2696+0.2487] = 0.2591

Updated distance matrix for cluster p3,p6,p4

	p1	p2,p5	p3,p6,p4
p1	0	0.2899	0.2985
p2,p5	0.2889	0	0.2591
p3,p6,p4	0.2985	0.2591	0

(p3,p6,p4) and (p2,p5) have the minimum distance from the above table. So, distance((p3,p6,p4),(p2,p5)) is minimum.

Now we need to update the distance matrix.

To update the distance matrix AVG[distance(((p3,p6,p4),(p2,p5)),p1)]:

AVG[distance(((p3,p6,p4),p1),((p2,p5),p1))] = ½ * [0.2985+0.2889] = 0.2937

Updated distance matrix for cluster p3,p6,p4,p2,p5

	p1	p3,p6,p4,p2,p5
p1	0	0.2937
p3,p6,p4,p2,p5	0.2937	0

Dendrogram for Average link: