Big Data Assignment 3B

By - Navaneeth Venugopala Rao

NET ID - nbv140130

1.

I	1	2	3	4	5	6	7	8
U1	4	5	0	5	1	0	3	2
U2	0	3	4	3	1	2	1	0
U3	2	0	1	3	0	4	5	3

a) U1,U2: 4/8
U2,U3: 4/8
U1,U3: 4/8

b) U1,U2: = 1 - 34 / 56.5685424949
= 1 - 0.601040764009 = **0.39895**
U2,U3: = 1 - 26 / 50.5964425627
= 1 - 0.513870119777 = **0.48612**
U1,U3: = 1 - 44 / 71.55417528
= 1 - 0.614918693812 = **0.38508**

c) U1,U2: = 1 - 4 / 6
= **1 / 3**
U2,U3: = 1 - 4 / 6
= **1 / 3**
U1,U3: = 1 - 4 / 6
= **1 / 3**

d)

I	1	2	3	4	5	6	7	8
U1	1	1	0	1	0	0	1	0
U2	0	1	1	1	0	0	0	0
U3	0	0	0	1	0	1	1	1

U1,U2: 1 – 2 / 5 = **3 / 5**
U2,U3: 1 – 1 / 6 = **5 / 6**
U1,U3: 1 – 2 / 6 = **2 / 3**

e) U1,U2: = 1 - 2 / 3.46410161514
= 1 – 0.57735026919 = **0.42264**
U2,U3: = 1 - 1 / 3.46410161514
= 1 – 0.28867513459 = **0.71132**
U1,U3: = 2 / 4
= 1 -0.5 = **0.5**

f)

I	1	2	3	4	5	6	7	8	
U1	0.67	1.67	0	1.67	-2.33	0	-0.33	-1.33	avg = 20/6
U2	0	0.67	1.67	0.67	-1.33	-0.33	-1.33	0	avg = 14/6
U3	-1	0	-2	0	0	1	2	0	avg = 18/6 = 3

g) U1,U2: = 1 - 5.7756 / 9.88833431676
= 1 - 0.584082193723 = **0.41591**
U2,U3: =1 - (-6.33 / 8.56352731064)
= **1.739181387574**
U1,U3: = 1 - (-1.33 / 11.5470342513)
= **1.115181090751**

2.

a)

Distances::

(1,2)	(1,3)	(1,4)	(1,5)	(1,6)	(1,7)	(1,8)
0.5	1	0.66	1	1	0.5	1

(2,3)	(2,4)	(2,5)	(2,6)	(2,7)	(2,8)
0.5	0.33	1	1	0.66	1

(3,4)	(3,5)	(3,6)	(3,7)	(3,8)
0.66	1	1	1	1

(4,5)	(4,6)	(4,7)	(4,8)
1	0.66	0.33	0.66

(5,6)	(5,7)	(5,8)
1	1	1

(6,7)	(6,8)
0.5	0

(7,8)
0.5

Joining (6,8) to form a cluster (68).

Distances::

(1,2)	(1,3)	(1,4)	(1,5)	(1,68)	(1,7)
0.5	1	0.66	1	1	0.5

(2,3)	(2,4)	(2,5)	(2,68)	(2,7)
0.5	0.33	1	1	0.66

(3,4)	(3,5)	(3,68)	(3,7)
0.66	1	1	1

(4,5)	(4,68)	(4,7)
1	0.66	0.33

(5,68)	(5,7)
1	1

(7,68)
0.5

Joining (2,4) to form a cluster (24).

Distances::

(1,24)	(1,3)	(1,5)	(1,68)	(1,7)
0.5	1	1	1	0.5

(24,3)	(24,5)	(24,68)	(24,7)
0.5	1	0.66	0.33

(3,5)	(3,68)	(3,7)
1	1	1

(5,68)	(5,7)
1	1

(7,68)
0.5

Joining (24,7) to form a cluster (247).

Distances::

(1,247)	(1,3)	(1,5)	(1,68)
0.5	1	1	1

(247,3)	(247,5)	(247,68)
0.5	1	0.5

(3,5)	(3,68)
1	1

(5,68)
1

Joining (1,247) to form a cluster (1247).

Hence the 4 clusters are
(1247)(3)(5)(68)

b)

I	1247	3	5	68
U1	4.25	0	1	1
U2	2.33	4	1	1
U3	3.33	1	0	3.5

c) U1,U2: = 1 - 11.9025 / 21.6804590876
= 1 - 0.548996677233 = **0.45100**
U2,U3: = 1 - 15.2589 / 23.8795656202
= 1 - 0.638994035431 = **0.36100**
U1,U3: = 1 - 17.6525 / 22.0974926462
= 1 - 0.798846289152 = **0.20115**

3.
rt implies root( )

a)	P	1	2	3	4	5	6	7	8
	x	2	3	4	5	6	8	8	9
	y	2	4	7	3	7	7	1	3

Distances::

(1,5)	(2,5)	(3,5)	(4,5)	(6,5)	(7,5)	(8,5)
rt(4^2 + 5^2)	|rt(3^2 + 3^2)	|rt(2^2 + 0^2)	|rt(1^2 + 4^2)	|rt(2^2 + 0^2)	|rt(2^2 + 6^2)	|rt(3^2 + 4^2)
6.40	|4.24	|2	|4.12	|2	|6.32	|5

(1,6)	(2,6)	(3,6)	(4,6)	(6,5)	(7,6)	(8,6)
rt(6^2 + 5^2)	|rt(5^2 + 3^2)	|rt(4^2 + 0^2)	|rt(3^2 + 4^2)	|rt(2^2 + 0^2)	|rt(0^2 + 6^2)	|rt(1^2 + 4^2)
7.81	|5.83	|4	|5	|2	|6	|4.12

cluster 1: {1,2,3,4,5}
cluster 2: {6,7,8}

Iteration 2::
New Centroids::
M1 = (4,4.6)
M2 = (8.33,3.66)

Distances::

M	1	2
P1	rt(2^2 + 2.6^2)=3.28	rt(6.33^2 + 1.66^2)=6.51
P2	rt(1^2 + 0.6^2)=1.16	rt(5.33^2 + 0.33^2)=5.31
P3	rt(0^2 + 2.4^2)=2.4	rt(4.33^2 + 3.33^2)=5.43
P4	rt(1^2 + 1.6^2)=1.886	rt(3.33^2 + 0.66^2)=3.36
P5	rt(2^2 + 2.4^2)=3.12	rt(2.33^2 + 3.33^2)=4.047
P6	rt(4^2 + 2.4^2)=4.66	rt(0.33^2 + 3.33^2)=3.343
P7	rt(4^2 + 3.6^2)=5.38	rt(0.33^2 + 2.66^2)=2.686
P8	rt(5^2 + 1.6^2)=5.259	rt(0.66^2 + 0.66^2)=0.97

Cluster formed is the same, hence centroid values are same. Hence:
cluster 1: {1,2,3,4,5}
cluster 2: {6,7,8}

b)

Distances::

(1,3)	(2,3)	(4,3)	(5,3)	(6,3)	(7,3)	(8,3)
rt(2^2 + 5^2)	|rt(1^2 + 3^2)	|rt(1^2 + 4^2)	|rt(2^2 + 0^2)	|rt(4^2 + 0^2)	|rt(4^2 + 6^2)	|rt(5^2 + 4^2)
5.385	|3.16	|4.12	|2	|4	|7.211	|6.403

(1,7)	(2,7)	(3,7)	(4,7)	(5,7)	(6,7)	(8,7)
rt(6^2 + 5^2)	|rt(5^2 + 3^2)	|rt(4^2 + 6^2)	|rt(3^2 + 2^2)	|rt(2^2 + 6^2)	|rt(0^2 + 6^2)	|rt(1^2 + 2^2)
6.083	|5.831	|7.211	|3.605	|6.324	|6	|2.236

cluster 1: {1,2,3,5,6}
cluster 2: {4,7,8}

Iteration 2::
New Centroids::
M1 = (4.6,5.4)
M2 = (7.33,2.33)

Distances::

M	1	2
P1	rt(2.6^2 + 3.4^2)=4.280	rt(5.33^2 + 0.33^2)=5.340
P2	rt(1.6^2 + 1.4^2)=2.126	rt(4.33^2 + 1.67^2)=4.640
P3	rt(0.6^2 + 1.6^2)=1.709	rt(3.33^2 + 4.67^2)=5.735
P4	rt(0.4^2 + 2.4^2)=2.433	rt(2.33^2 + 0.67^2)=2.424
P5	rt(1.4^2 + 1.6^2)=2.126	rt(1.33^2 + 4.67^2)=4.855
P6	rt(3.4^2 + 1.6^2)=3.757	rt(0.67^2 + 4.67^2)=4.727
P7	rt(3.4^2 + 4.4^2)=5.560	rt(0.67^2 + 1.33^2)=1.489
P8	rt(4.4^2 + 2.4^2)=5.012	rt(1.67^2 + 0.67^2)=1.799

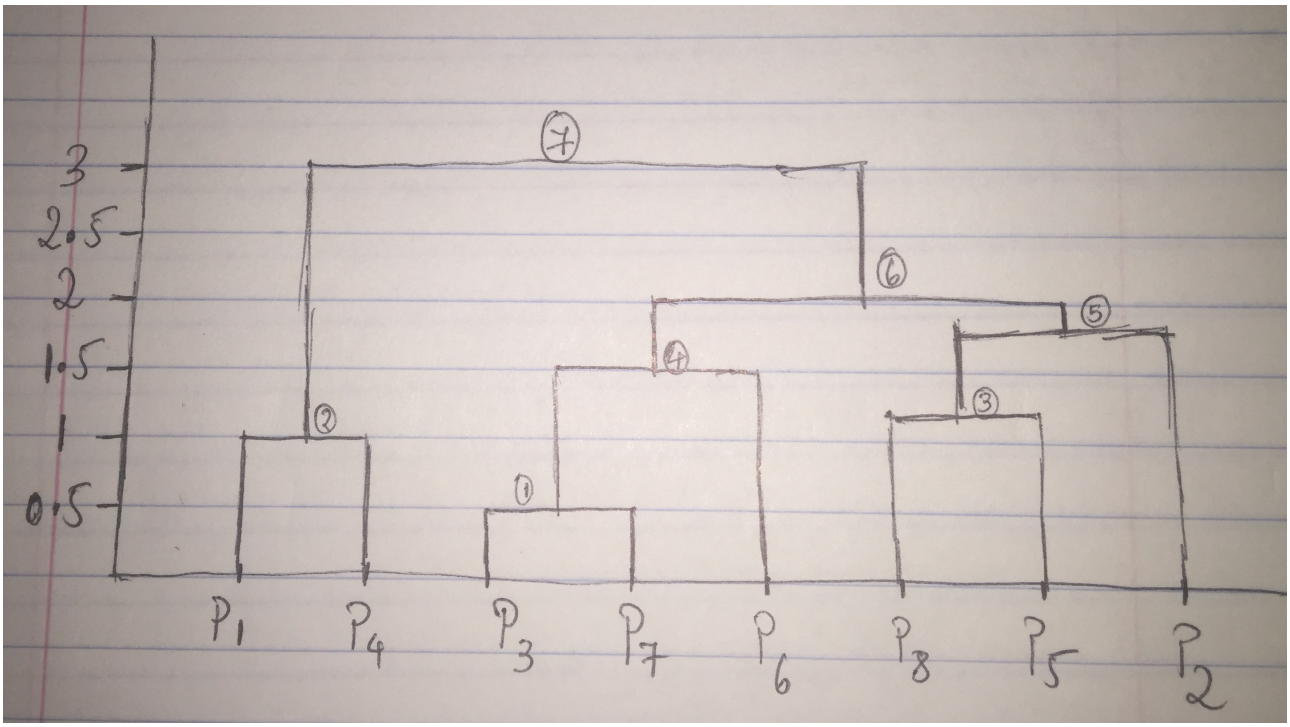
Cluster formed is the same, hence centroid values are same. Hence:
cluster 1: {1,2,3,5,6}
cluster 2: {4,7,8}

Yes the clustering assignments changed.

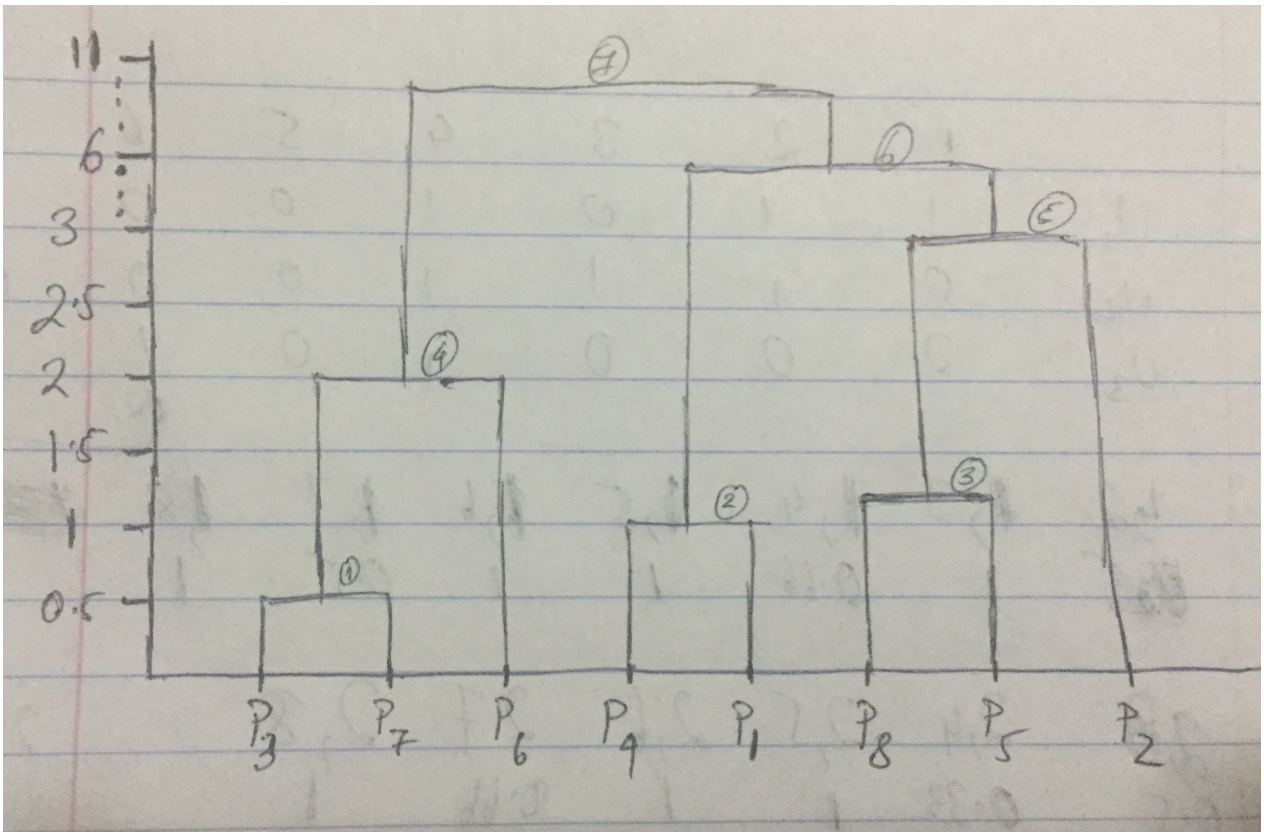
4.

a)

i)


3
2.5
2
1.5
1
0.5
7
6
5
4
3
2
1
P1
P4
P3
P7
P6
P8
P5
P2

ii)


11
6
3
2.5
2
1.5
1
0.5
7
6
5
4
3
2
1
P3
P7
P6
P4
P1
P8
P5
P2

b)
i) Single Link:

d({2,3,5,7,8}, {1,4,6})
= min({2,1},{2,4},{2,6},
{3,1},{3,4},{3,6},
{5,1},{5,4},{5,6},
{7,1},{7,4},{7,6},
{8,1},{8,4},{8,6}}
= min {7 , 8 , 4 , 4.5 , 5.5 , 1.5 , 8.8 , 9.8 , 5.8 , 5 , 6 , 2 , 10 , 11 , 7 }
=1.5
ii) Complete Link:
= max({2,1},{2,4},{2,6},
{3,1},{3,4},{3,6},
{5,1},{5,4},{5,6},
{7,1},{7,4},{7,6},
{8,1},{8,4},{8,6}}
= max {7 , 8 , 4 , 4.5 , 5.5 , 1.5 , 8.8 , 9.8 , 5.8 , 5 , 6 , 2 , 10 , 11 , 7 }
= 11
iii) Average Link:
= avg({2,1},{2,4},{2,6},
{3,1},{3,4},{3,6},
{5,1},{5,4},{5,6},
{7,1},{7,4},{7,6},
{8,1},{8,4},{8,6}}
= avg {7 , 8 , 4 , 4.5 , 5.5 , 1.5 , 8.8 , 9.8 , 5.8 , 5 , 6 , 2 , 10 , 11 , 7 }
= 6.393