Homework 7 UG

December 7, 2021

Jose Carlos Munoz

Part 1)
Q6.a)
Q6.b)
Q6.c)
Q6.d)
Q6.e)
Q16)

The Following is for the Single Link
This is the similarity matrix from each of the points.

-	1	2	3	4	5
1	1.00	0.10	0.41	0.55	0.35
2	0.10	1.00	0.64	0.47	0.98
3	0.41	0.64	1.00	0.44	0.85
4	0.55	0.47	0.44	1.00	0.76
5	0.35	0.98	0.85	0.76	1.00

From this set we can see that the value with the highest similarity is 2 and 5, so they are cluster together to form cluster 2,5

-	1	2	3	4	5
1	1.00	0.35	0.41	0.55	0.35
2	0.35	1.00	0.85	0.76	1.00
3	0.41	0.85	1.00	0.44	0.85
4	0.55	0.76	0.44	1.00	0.76
5	0.35	1.00	0.85	0.76	1.00

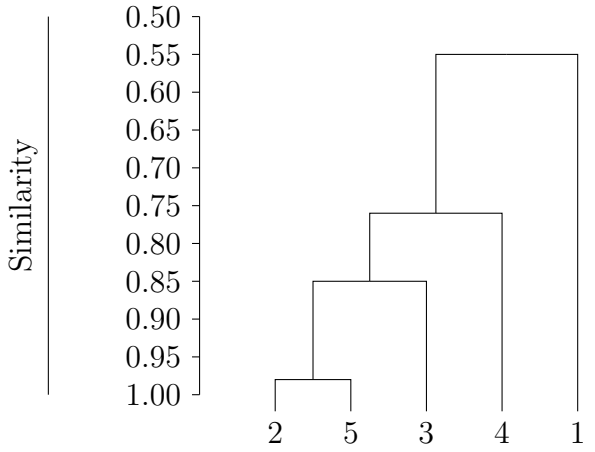
From this set we can see that the value with the highest similarity is cluster 2,5 and 3, so they are cluster together to form cluster 2,5,3

-	1	2	3	4	5
1	1.00	0.41	0.41	0.55	0.41
2	0.41	1.00	1.00	0.76	1.00
3	0.41	1.00	1.00	0.76	1.00
4	0.55	0.76	0.44	1.00	0.76
5	0.41	1.00	1.00	0.76	1.00

From this set we can see that the value with the highest similarity is cluster 2,5,3 and 4, so they are cluster together to form cluster 2,5,3,4

-	1	2	3	4	5
1	1.00	0.55	0.55	0.55	0.55
2	0.55	1.00	1.00	1.00	1.00
3	0.55	1.00	1.00	1.00	1.00
4	0.55	1.00	1.00	1.00	1.00
5	0.55	1.00	1.00	1.00	1.00

From this set we can see that the value with the highest similarity is cluster 2,5,3,4 and 1, so they are cluster together to form cluster 2,5,3,4,1
Now we display the graph made from these clusters.



The Following is for the Complete Link
This is the similarity matrix from each of the points.

-	1	2	3	4	5
1	1.00	0.10	0.41	0.55	0.35
2	0.10	1.00	0.64	0.47	0.98
3	0.41	0.64	1.00	0.44	0.85
4	0.55	0.47	0.44	1.00	0.76
5	0.35	0.98	0.85	0.76	1.00

From this set we can see that the value with the highest similarity is 2 and 5, so they are cluster together to form cluster 2,5

-	1	2	3	4	5
1	1.00	0.10	0.41	0.55	0.10
2	0.10	1.00	0.64	0.47	1.00
3	0.41	0.64	1.00	0.44	0.64
4	0.55	0.47	0.44	1.00	0.47
5	0.10	1.00	0.64	0.47	1.00

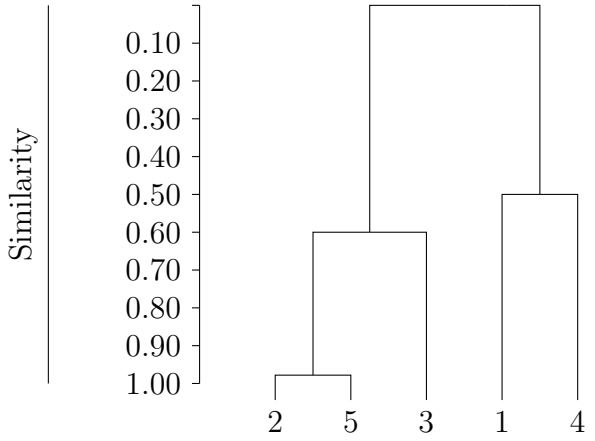
From this set we can see that the value with the highest similarity is cluster 2,5 and 3, so they are cluster together to form cluster 2,5,3

-	1	2	3	4	5
1	1.00	0.10	0.10	0.55	0.10
2	0.10	1.00	1.00	0.44	1.00
3	0.10	1.00	1.00	0.44	1.00
4	0.55	0.44	0.44	1.00	0.44
5	0.10	1.00	1.00	0.44	1.00

From this set we can see that the value with the highest similarity is cluster 1 and 4, so they are cluster together to form cluster 1,4

-	1	2	3	4	5
1	1.00	0.10	0.10	1.00	0.10
2	0.10	1.00	1.00	0.10	1.00
3	0.10	1.00	1.00	0.10	1.00
4	1.00	0.10	0.10	1.00	0.10
5	0.10	1.00	1.00	0.10	1.00

From this set we can see that the value with the highest similarity is cluster 2,5,3 and 1,4 so they are cluster together to form cluster 2,5,3,4,1
Now we display the graph made from these clusters.



Q17.a)
i)

-	Dist. 18	Dist 45	*Class*	*error*
6	12	39	18	144
12	6	33	18	36
18	0	27	18	0
24	6	21	18	36
30	12	15	18	144
42	24	3	45	9
48	30	3	45	9

From the distance Chart above, we can conclude that the first cluster has valuse 6,12,18,24 and 30 with an error of 360. The Second cluster has values 42 and 48 with an error of 18. So the total error for these clusters is 378
ii)

-	Dist. 15	Dist 40	*Class*	*error*
6	9	34	15	81
12	3	28	15	9
18	3	22	15	9
24	9	16	15	81
30	15	10	40	100
42	27	2	40	4
48	33	8	40	64

From the distance Chart above, we can conclude that the first cluster has valuse 6,12,18 and

24 with an error of 180. The Second cluster has values 30,42 and 48 with an error of 168. So the total error for these clusters is 348
Q17.b)
Both centroids are stable solutions
Q17.c)
The Clusters produce by single link is set 6,12,18,24,30 and set 42 and 48
Q17.d)
The K-means produces the most natural clustering.
Q17.e)
K-mean produces cluster that have a higher density in a region. Q17.f)
A problem with k-means is that it is not that well at handeling different cluster size and might break larger clusters.
Part2)
1)
All the Core points are p,q,v,r,s,t,w,b,c,h,k,d,e,i,f,g,j,and n
2)
a is density reachable to d because d is a core point and a is one of its neighboors to d
3)
It is not true becuase it can be that going the oppisite direction wil lead us to a non core point, thereby nullifying the results. So it is not symmetic
4)
n is density reachable from e. the points to get there are as follows;e,b,c,f,g,j and n
5)
The Density-based cluster are as follow
C1 = {p,v,g};C2 = {d,h,k,r,s,t,w};C3 = {i,e,b,c,f,g,j,n,o}
The noise points are a,x,l,u,m