

CS 5525

Biplav Choudhury (906089826)

19 October 2020

1 Solutions to Assignment 5

1.1 Problem 1

1.1.1 Part 1

After 3 iterations, the the centroids are 0.1, 0.3 and 0.7 where the assignments are -

- 0.1 - 0.1
- 0.3 - 0.2, 0.4
- 0.7 - 0.5, 0.6, 0.8, 0.9

1.1.2 Part 2

The SSE is 0.12

1.1.3 Part 3

Using bisecting k-means, the new centroids are 0.267, 0.0.55 and 0.85. The clusters are -

- 0.267 - 0.1, 0.2, 0.4
- 0.55 - 0.5, 0.6
- 0.85 - 0.8, 0.9

The SSE here is 0.060067

1.1.4 Part 3

As the SSE in bisecting k means is larger, bisecting k means is better.

1.2 Problem 2

The dendrogram in case of single linkage is

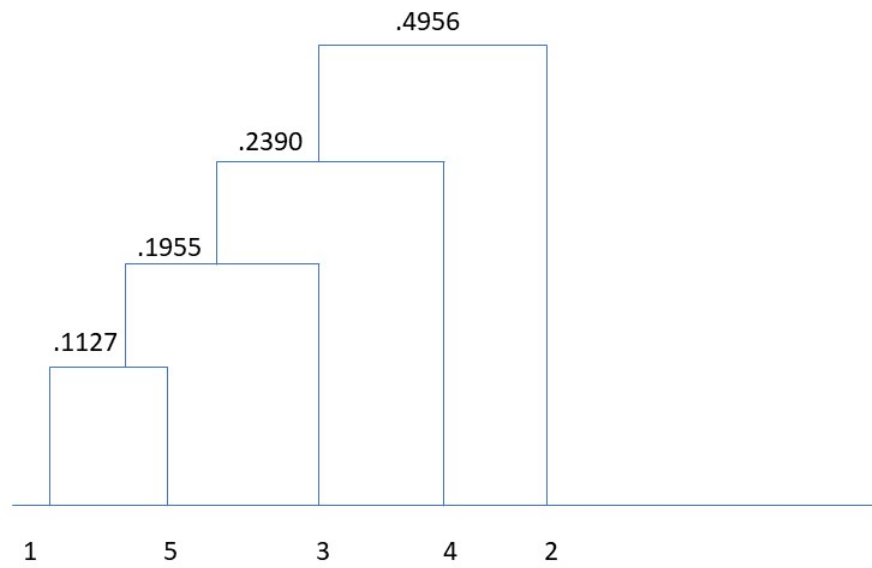


Figure 1: Single Linkage

- Merge 1 and 5
- Merge (1,5) and 3
- Merge (1,5,3) and 4
- Merge (1,5,3,4) and 2.

The dendrogram in case of complete linkage is

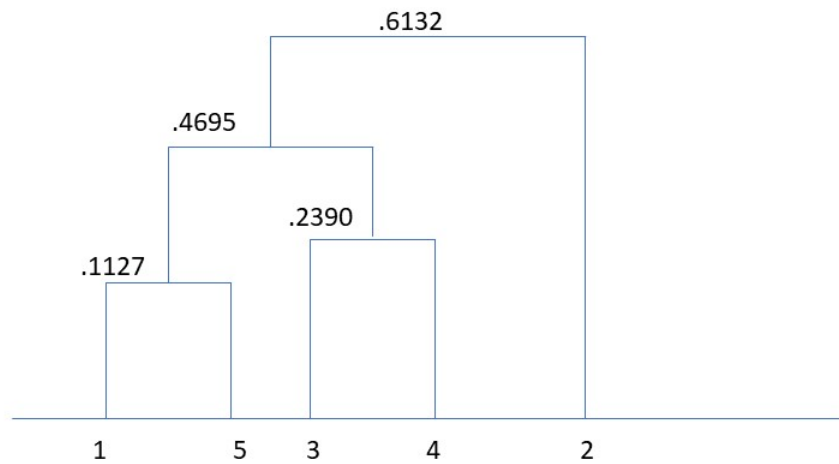


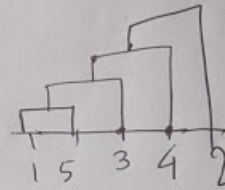
Figure 2: Complete Linkage

- Merge 1 and 5
- Merge 3 and 4
- Merge (1,5) and (3,4)
- Merge (1,5,3,4) and 2

the calculation were done by hand and pictures have been added here

MIN / SINGLE LINK HIERARCHIAL CLUSTERING

	1	2	3	4	5
P1	0	.5840	.1955	.3815	.1127
P2	.5840	0	.6132	.4956	.5733
P3	.1955	.6132	0	.2390	.4694
P4	.3815	.4956	.2390	0	.4694
P5	.1127	.5733	.4694	.4694	0



	P1, P5	P2	P3	P4
P1, P5	0	.5733	.1955	.3815
P2	.5733	0	.6132	.4956
P3	.1955	.6132	0	.2390
P4	.3815	.4956	.2390	0

$$P1, P5 - P2 = \min(.5840, .5733) = .5733$$

$$P1, P5 - P3 = \min(.1955, .3067) = .1955$$

$$P1, P5 - P4 = \min(.3815, .4694) = .3815$$

	P1, P5, P3	P2	P4
P1, P5, P3	0	.5733	.2390
P2	.5733	0	.4956
P4	.2390	.4956	0

$$P1, P5, P3 - P2 = \min(.5733, .6132)$$

$$P1, P5, P3 - P4 = \min(.3815, .2390) = .2390$$

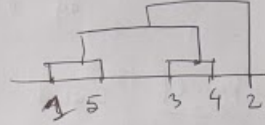
	1, 5, 3, 4	2
1, 5, 3, 4	0	.4956
2	.4956	0

$$\begin{aligned} 2-1 &\rightarrow .5840 \\ 2-5 &\rightarrow .5733 \\ 2-3 &\rightarrow .6132 \\ 2-4 &\rightarrow .4956 \end{aligned}$$

Figure 3: Single Linkage

MAX linkage:

	p1	p2	p3	p4	p5
p1	0	.5840	.1955	.3815	.1127
p2	.5840	0	.6132	.4956	.5733
p3	.1955	.6132	0	.2390	.3067
p4	.3815	.4956	.2390	0	.4694
p5	.1127	.5733	.3067	.4694	0



	p1, p5	p2	p3	p4
p1, p5	0	.5840	.3067	.4694
p2	.5840	0	.6132	.4956
p3	.3067	.6132	0	.2390
p4	.4694	.4956	.2390	0

$$* p1, p5 - p2 = \max(.5840, .5733) = .5840$$

$$* p1, p5 - p3 = \max(.3067, .1955) = .3067$$

$$* p1, p5 - p4 = \max(.3815, .4694) = .4694$$

↓

	p1, p5	p2	p3, p4
p1, p5	0	.5840	.4694
p2	.5840	0	.6132
p3, p4	.4694	.6132	0

$$p1, p5 - p3, p4 = \max(.3815, .3067, .4694) = .4694$$

$$p2 - p3, p4 = \max(.6132, .4956) = .6132$$

	1, 5, 4	2
1, 5, 4	0	.6132
2	.6132	0

$$\begin{aligned} 2-1 &\rightarrow .5840 \\ 2-5 &\rightarrow .5733 \\ 2-3 &\rightarrow .6132 \\ 2-4 &\rightarrow .4956 \end{aligned}$$

Figure 4: Complete Linkage

1.3 Problem 3

1.3.1 Part 1

Core points are a,b,c,d,e,f,g,h,i,j,k,l,q,r,s,t,x (17 points)

1.3.2 Part 2

Boundary points are p,u,z,y,m,w,v (7 points)

1.3.3 Part 3

Noise points are n,o (2 points)

1.3.4 Part 4

3 clusters

1.4 Problem 4

1.4.1 Part 1

For pure, entropy = 0, purity = 1, NMI = 1

1.4.2 Part 2

Entropy for both are 0.6068 and 0.5989

1.4.3 Part 3

Purity for both are 0.7

1.4.4 Part 4

NMI for both are