

HW 9

Anusha muppalla
16286311

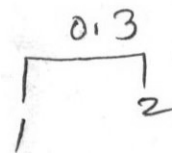
1) given,

	1	2	3	4
1		0.3	0.4	0.7
2	0.3		0.5	0.8
3	0.4	0.5		0.45
4	0.7	0.8	0.45	

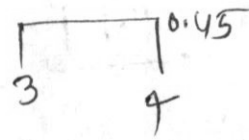
	1	2	3	4
1	0			
2	0.3	0		
3	0.4	0.5	0	
4	0.7	0.8	0.45	0

a) using complete linkage

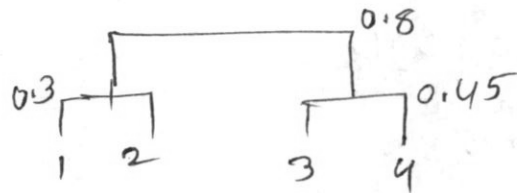
	(1,2)	3	4
(1,2)	0		
3	0.5	0	
4	0.8	0.45	0



	(1,2)	(3,4)
(1,2)	0	
(3,4)	0.8	0

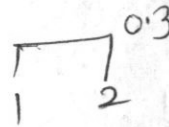


cluster dendrogram



b) single linkage:-

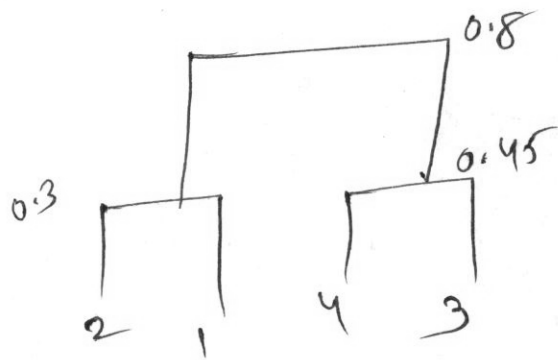
	1	2	3	4
1	0			
2	0.3	0		
3	0.4	0.5	0	
4	0.7	0.8	0.45	0



c) when we cut the dendrogram obtained in (a), the two clusters would be
 $(1,2)$ & $(3,4)$

d) when we cut the dendrogram obtained in (b)
 then, the two clusters would be
 $((1,2), 3)$ & 4

c)



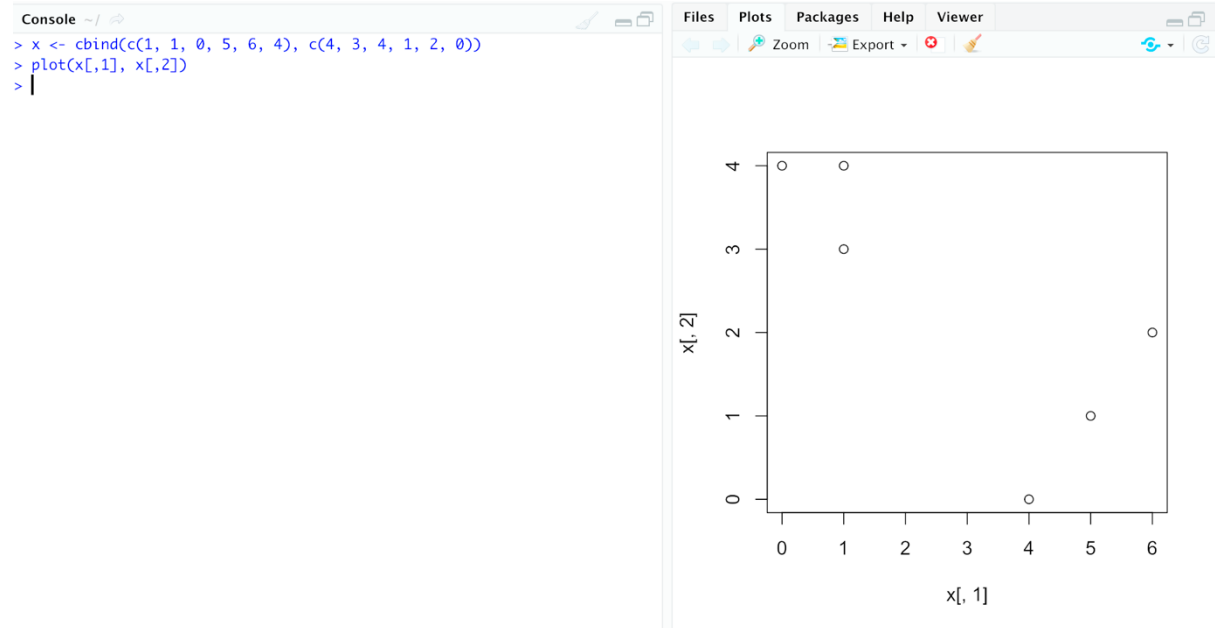
3)

a) there is not enough information for example, if $d(1,4) = 2$, $d(1,5) = 3$, $d(2,4) = 1$, $d(2,5) = 3$, $d(3,4) = 4$ and $d(3,5) = 1$ the single linkage dissimilarity between $\{1,2,3\}$ and $\{4,5\}$ would be equal to 1 and the linkage dissimilarity between $\{1,2,3\}$ and $\{4,5\}$ would be equal to 4 so with single linkage they would fuse at a height of 1, and with complete linkage, they would fuse at a height of 4, but if all inter observations distance are equal to 2.

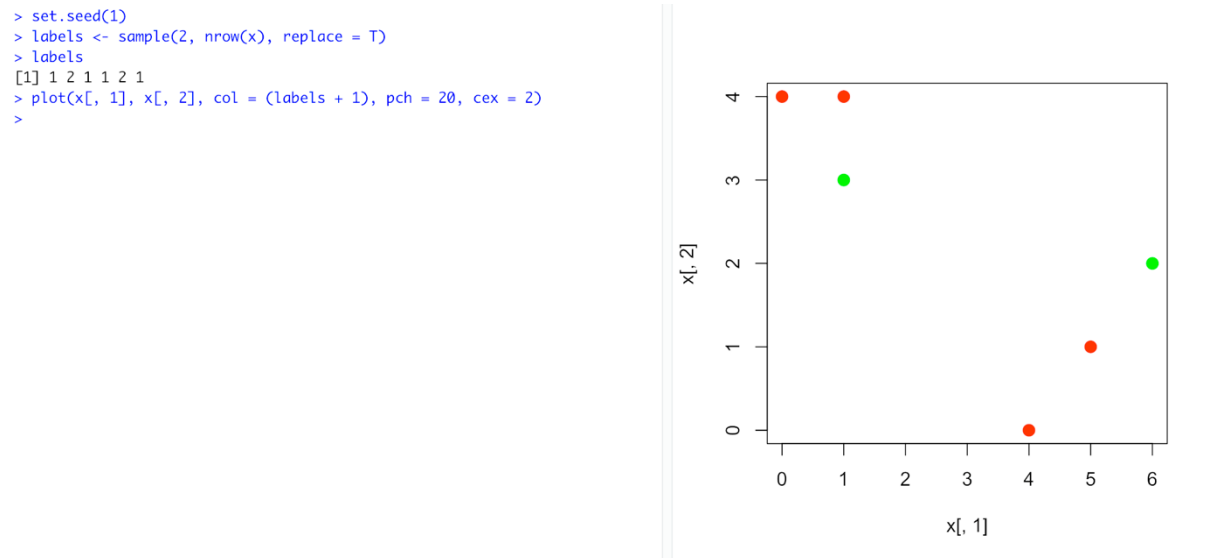
b) they would fuse at same height for example, if $d(5,6) = 2$ the single and complete linkage dendrogram, the clusters $\{5\}$ and $\{6\}$ also fuse at a certain point. which linkage dissimilarities $\{5\}$ and $\{6\}$ would be equal to 2. so, they would fuse at a height of 2 for single and complete linkage

2)

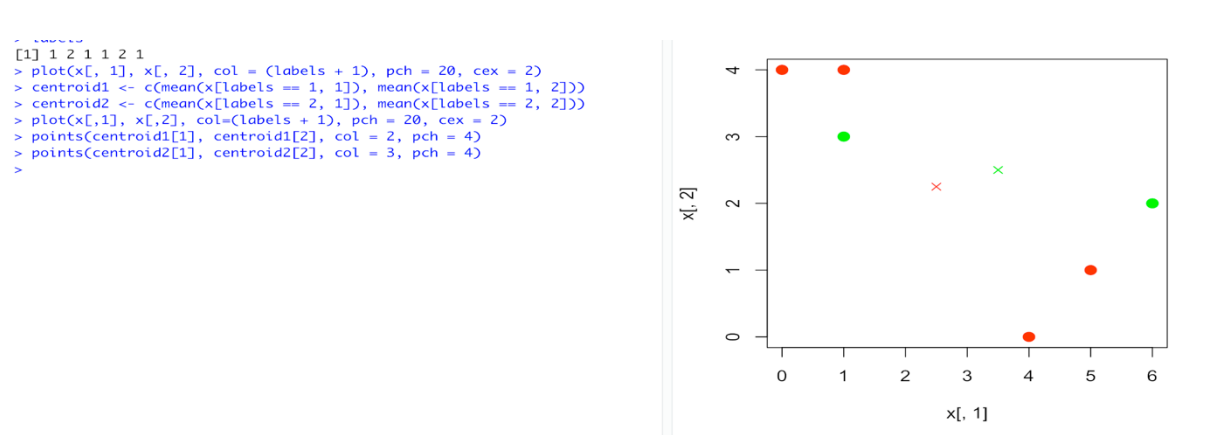
a.



b.



c.

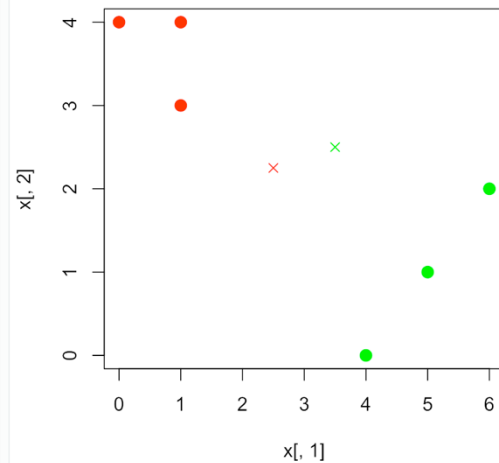


d.

```

# labels
[1] 1 2 1 1 2 1
> plot(x[, 1], x[, 2], col = (labels + 1), pch = 20, cex = 2)
> centroid1 <- c(mean(x[labels == 1, 1]), mean(x[labels == 1, 2]))
> centroid2 <- c(mean(x[labels == 2, 1]), mean(x[labels == 2, 2]))
> plot(x[,1], x[,2], col=(labels + 1), pch = 20, cex = 2)
> points(centroid1[1], centroid1[2], col = 2, pch = 4)
> points(centroid2[1], centroid2[2], col = 3, pch = 4)
> labels <- c(1, 1, 1, 2, 2, 2)
> plot(x[, 1], x[, 2], col = (labels + 1), pch = 20, cex = 2)
> points(centroid1[1], centroid1[2], col = 2, pch = 4)
> points(centroid2[1], centroid2[2], col = 3, pch = 4)
> |

```

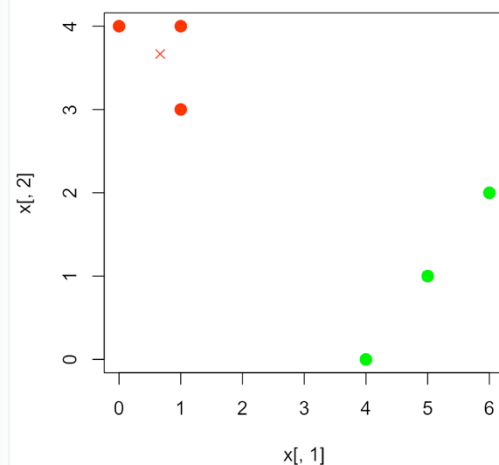


e.

```

# labels
[1] 1 2 1 1 2 1
> plot(x[, 1], x[, 2], col = (labels + 1), pch = 20, cex = 2)
> centroid1 <- c(mean(x[labels == 1, 1]), mean(x[labels == 1, 2]))
> centroid2 <- c(mean(x[labels == 2, 1]), mean(x[labels == 2, 2]))
> plot(x[,1], x[,2], col=(labels + 1), pch = 20, cex = 2)
> points(centroid1[1], centroid1[2], col = 2, pch = 4)
> points(centroid2[1], centroid2[2], col = 3, pch = 4)
> labels <- c(1, 1, 1, 2, 2, 2)
> plot(x[, 1], x[, 2], col = (labels + 1), pch = 20, cex = 2)
> points(centroid1[1], centroid1[2], col = 2, pch = 4)
> points(centroid2[1], centroid2[2], col = 3, pch = 4)
> centroid1 <- c(mean(x[labels == 1, 1]), mean(x[labels == 1, 2]))
> centroid2 <- c(mean(x[labels == 2, 1]), mean(x[labels == 2, 2]))
> plot(x[,1], x[,2], col=(labels + 1), pch = 20, cex = 2)
> points(centroid1[1], centroid1[2], col = 2, pch = 4)
> points(centroid2[1], centroid2[2], col = 3, pch = 4)
> |

```



f.

```

# labels
[1] 1 2 1 1 2 1
> plot(x[, 1], x[, 2], col = (labels + 1), pch = 20, cex = 2)
> centroid1 <- c(mean(x[labels == 1, 1]), mean(x[labels == 1, 2]))
> centroid2 <- c(mean(x[labels == 2, 1]), mean(x[labels == 2, 2]))
> plot(x[,1], x[,2], col=(labels + 1), pch = 20, cex = 2)
> points(centroid1[1], centroid1[2], col = 2, pch = 4)
> points(centroid2[1], centroid2[2], col = 3, pch = 4)
> labels <- c(1, 1, 1, 2, 2, 2)
> plot(x[, 1], x[, 2], col = (labels + 1), pch = 20, cex = 2)
> points(centroid1[1], centroid1[2], col = 2, pch = 4)
> points(centroid2[1], centroid2[2], col = 3, pch = 4)
> centroid1 <- c(mean(x[labels == 1, 1]), mean(x[labels == 1, 2]))
> centroid2 <- c(mean(x[labels == 2, 1]), mean(x[labels == 2, 2]))
> plot(x[,1], x[,2], col=(labels + 1), pch = 20, cex = 2)
> points(centroid1[1], centroid1[2], col = 2, pch = 4)
> points(centroid2[1], centroid2[2], col = 3, pch = 4)
> plot(x[, 1], x[, 2], col=(labels + 1), pch = 20, cex = 2)
> |

```

