PROBLEM 2

a. Dataset:

A $(0, 1.1)$

B $(0, 0)$

C $(1, 0)$

D $(1, 0.9)$

Distance matrix (manhattan distance)

$= |x_1 - x_2| + |y_1 - y_2|$

$d(A,B) = |0-0| + |1.1-0| = 1.1$

$d(A,C) = |0-1| + |1.1-0| = 2.1$

$d(A,D) = |0-1| + |1.1-0.9| = 1.2$

$d(B,C) = |0-1| + |0-0| = 1$

$d(B,D) = |0-1| + |0-0.9| = 1.9$

$d(C,D) = |1-1| + |0-0.9| = 0.9$

Distance matrix

	A	B	C	D
A	0	1.1	2.1	1.2
B	1.1	0	1	1.9
C	2.1	1	0	0.9
D	1.2	1.9	0.9	0

Clustering using d_{max}

The closest pair is (C, D) with $d(C,D) = 0.9$.
Hence, we update the distance matrix using the maximum distance

$d_{max}(\{A\}, \{C, D\}) = \max(d(A, C), d(A, D))$
$= \max(2.1, 1.2) = 2.1$

$d_{max}(\{B\}, \{C, D\}) = \max(d(B, C), d(B, D)) = \max(1, 1.9) = 1.9$

updated distance matrix becomes

	A	B	{C, D}
A	0	1.1	2.1
B	1.1	0	1.9
{C, D}	2.1	1.9	0

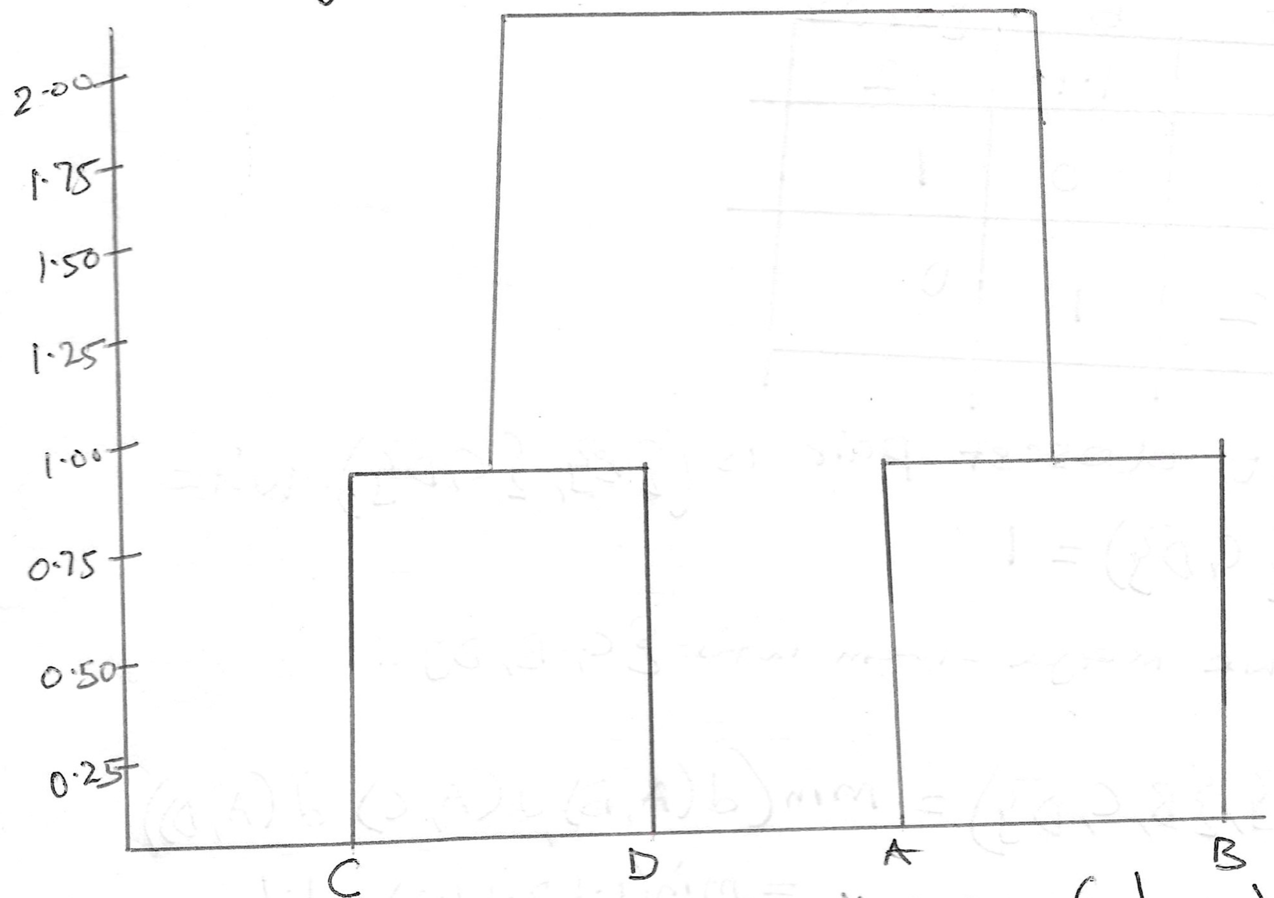
The next pair with smallest distance is (A, B)
$d(A, B) = 1.1$, hence we merge A and B into {A, B}
$d_{max}(\{A, B\}, \{C, D\}) = \max(d(A, C, D), d(B, C, D))$
$= \max(2.1, 1.9) = 2.1$

Updated distance matrix becomes

	{A, B}	{C, D}
{A, B}	0	2.1
{C, D}	2.1	0

We finally merge {A, B} ~~into~~ and {C, D} into one cluster

The dendogram becomes



Dendogram for Complete linkage (dmax)

b. Clustering with single linkage (dmin)
Continuing from (a) above (that using distances already calculated,

The smallest distance is $d(C,D) = 0.9$
Hence, we merge the cluster C and D into $\{C,D\}$. We update the distance matrix using

$$d_{min}(\{A\},\{C,D\}) = \min(d(A,C), d(A,D))$$
$$= \min(2.1, 1.2) = 1.2$$

$$d_{min}(\{B\},\{C,D\}) = \min(d(B,C), d(B,D))$$
$$= \min(1, 1.9) = 1$$

updated matrix becomes

	A	B	{C, D}
A	0	1.1	1.2
B	1.1	0	1
{C, D}	1.2	1	0

The next closest pair is $(\{B\}, \{C, D\})$ with $d(\{B\}, \{C, D\}) = 1$

hence we merge them into $\{C, B, D\}$

$$d_{min}(\{A\}, \{B, C, D\}) = \min(d(A,B), d(A,C), d(A,D))$$
$$= \min(1.1, 2.1, 1.2) = 1.1$$

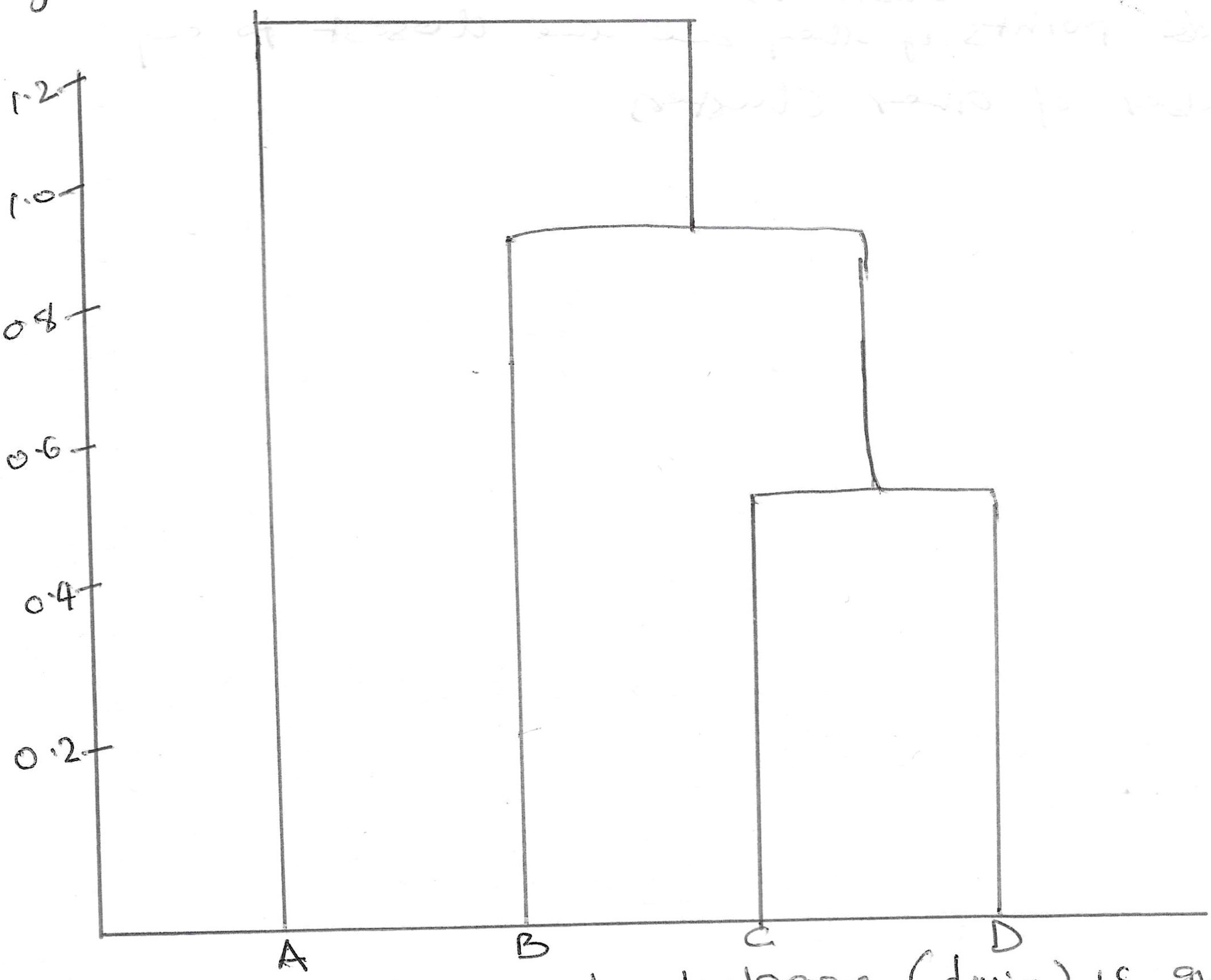
The updated distance matrix now becomes

	{A}	{B, C, D}
{A}	0	1.1
{B, C, D}	1.1	0

{A} and {B, C, D} can now be merged into one cluster

The dendogram for single linkage (dmin) is given below

(5)



<u>Dendogram for single linkage (dmin) is given below</u>

C. Complete linkage tends to create more compact clusters because it looks for maximum distance (dmax) between clusters, often leading to more balanced ~~clusters~~ and spherical shapes.

Single linkage on the other hand, can create long and elaborate clusters since it only considers minimum distance (dmin). this can result in chaining effects where clusters can merge over ~~too~~ a long distance

In addition, single linkage is generally sensitive to noise and outliers because it can connect distant points if they are the closest to any member of other clusters