19

Three two dimensional data points. Let's find two clusters.

```
##      [,1] [,2]
## [1,]    1    1
## [2,]    2    1
## [3,]    4    5
```

x_1, x_2, x_3

Begin with the randomly allocated centres:

```
##      [,1] [,2]
## [1,]    2    2
## [2,]    3    3
```

m_1, m_2

Compute which object belongs to which cluster

① distances:

$$d(x_1, m_1)^2 = (1-2)^2 + (1-2)^2 = 2$$
$$d(x_1, m_2)^2 = (1-3)^2 + (1-3)^2 = 8$$

$$d(x_2, m_1)^2 = (2-2)^2 + (1-2)^2 = 1$$
$$d(x_2, m_2)^2 = (2-3)^2 + (1-3)^2 = 5$$

$$d(x_3, m_1)^2 = (4-2)^2 + (5-2)^2 = 13$$
$$d(x_3, m_2)^2 = (4-3)^2 + (5-3)^2 = 5$$

$\Rightarrow$ x_1 and $x_2 \rightarrow m_1$

$x_3 \rightarrow m_2$

② Recompute cluster centers

$$m_1 = \frac{x_1 + x_2}{2} = \frac{[1\ 1] + [2\ 1]}{2} = \left[\frac{3}{2}\ \ 1\right] = [1.5\ \ 1]$$

$$m_2 = x_3 = [4\ \ 5]$$

③ Compute all distances again

$$d(x_1, m_1)^2 = (1-1.5)^2 + (1-1)^2 = 0.25$$
$$d(x_1, m_2)^2 = (1-4)^2 + (1-5)^2 = 25$$

$$d(x_2, m_1)^2 = (2-1.5)^2 + (1-1)^2 = 0.25$$
$$d(x_2, m_2)^2 = (2-4)^2 + (1-5)^2 = 20$$

$$d(x_3, m_1)^2 = (4-1.5)^2 + (5-1)^2 = 22.5$$
$$d(x_3, m_2)^2 = (4-4)^2 + (5-5)^2 = 0$$

$\Rightarrow$ x_1 and $x_2 \rightarrow m_1$

$x_3 \rightarrow m_2$

31)

Compute SSW, SSB and SST for the given set of points and the previously computed cluster centres.

```
##       [,1] [,2]
## [1,]     1    1
## [2,]     2    1
## [3,]     4    5
```

x_1, x_2, x_3

Compare to the SSW, SSB and SST using the initial cluster centres.

```
##       [,1] [,2]
## [1,]   1.5    1
## [2,]     4    5
```

c_1, c_2

SST = squared distances from all points to the overall mean.

overall mean $\Rightarrow$ $\bar{x} = \left[\frac{1+2+4}{3} \quad \frac{1+1+5}{3} \right] = \left[\frac{7}{3} \quad \frac{7}{3} \right]$

$$SST = \Sigma (x_i - \bar{x})^2 = \left(1-\tfrac{7}{3}\right)^2 + \left(1-\tfrac{7}{3}\right)^2 + \left(2-\tfrac{7}{3}\right)^2 + \left(1-\tfrac{7}{3}\right)^2 + \left(4-\tfrac{7}{3}\right)^2 + \left(5-\tfrac{7}{3}\right)^2$$

$$= \frac{138}{9} = 15.33$$

SSW = squared distances from each point to its cluster centre.

$$SSW = \Sigma \|x - c\|^2 = (1-1.5)^2 + (1-1)^2 + (2-1.5)^2 + (1-1)^2 + (4-4)^2 + (5-5)^2$$

$$= 0.5$$

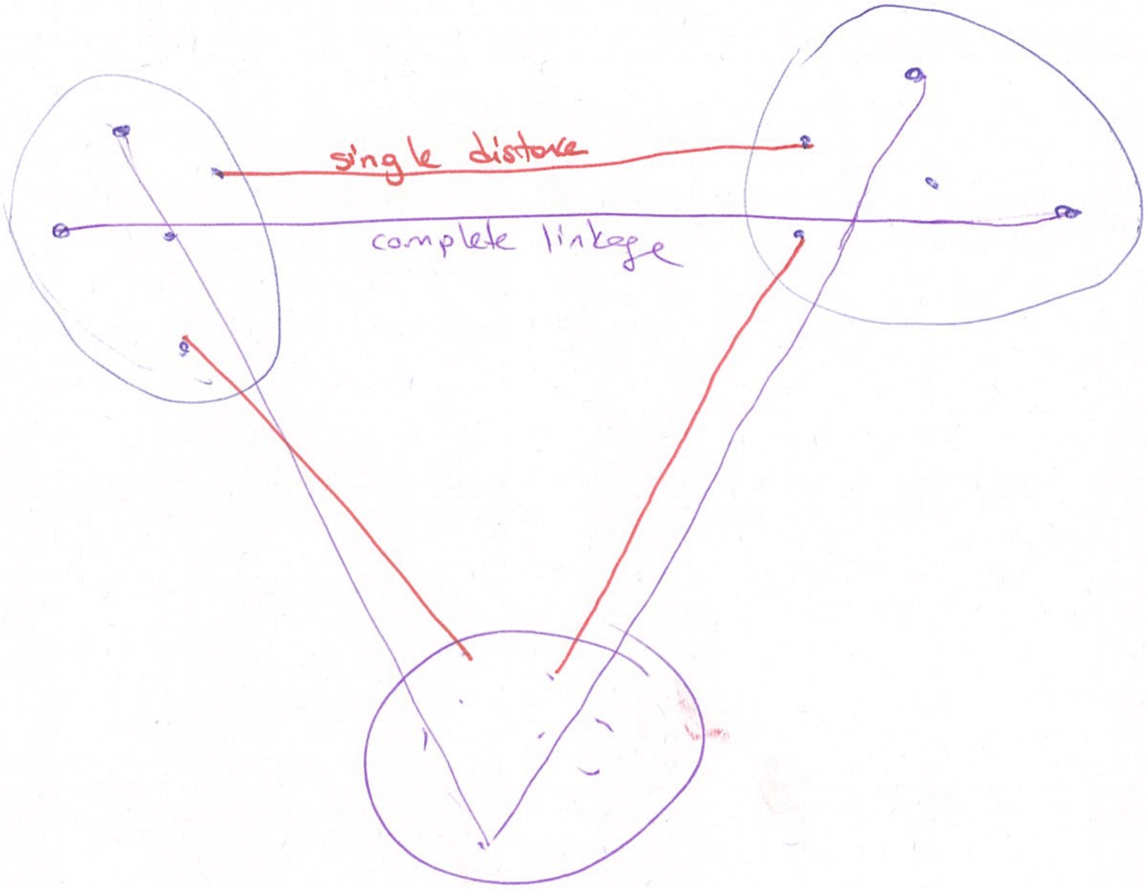
SSB = squared distances from each point's cluster centre to the overall mean.

$$SSB = \Sigma |c(x)| \, \|c - \bar{x}\|^2 = 2\left[\left(1.5-\tfrac{7}{3}\right)^2 + \left(1-\tfrac{7}{3}\right)^2\right] + \left[\left(4-\tfrac{7}{3}\right)^2 + \left(5-\tfrac{7}{3}\right)^2\right]$$

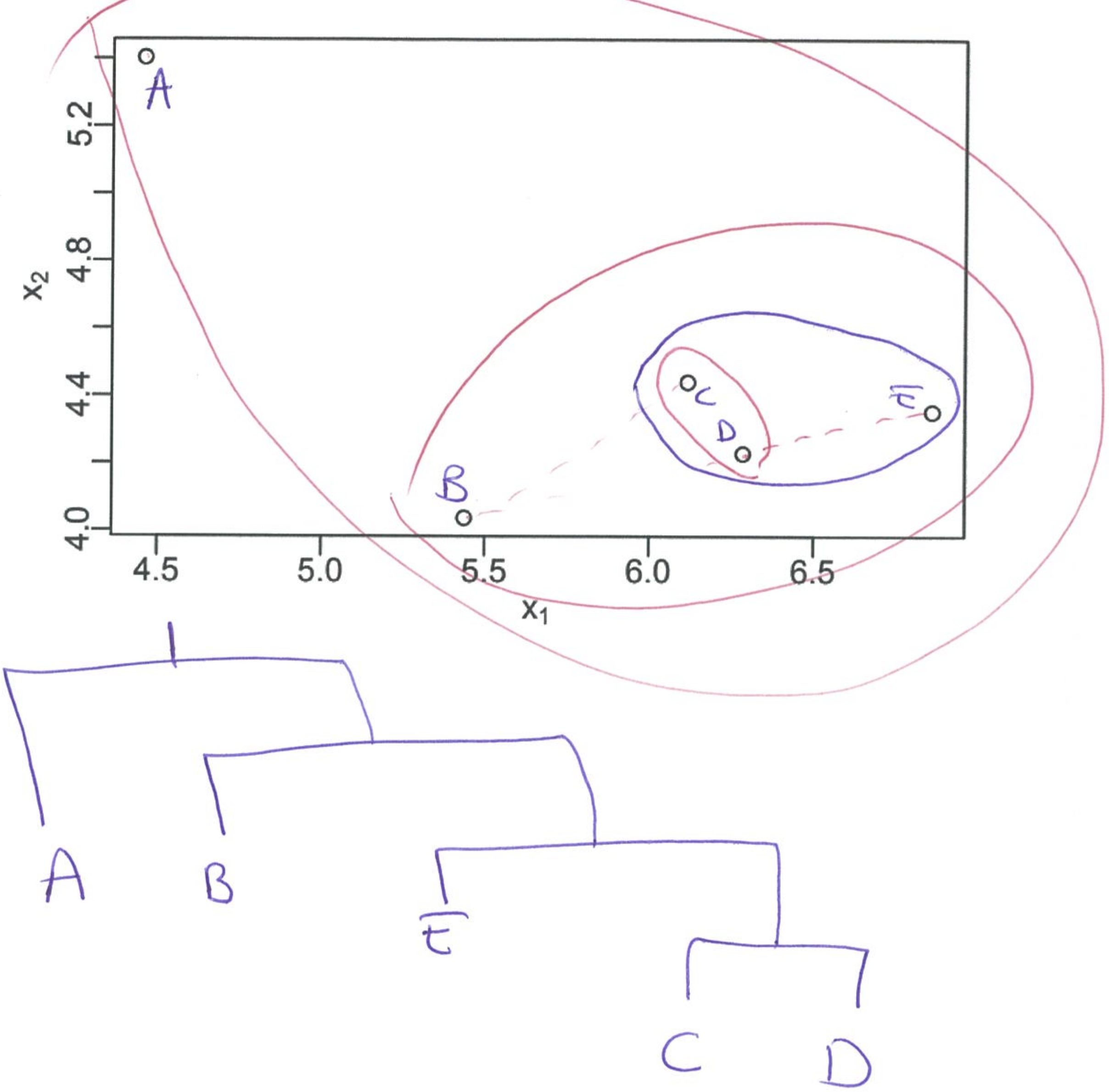
$$= 14.83$$

Note that SST = SSW + SSB

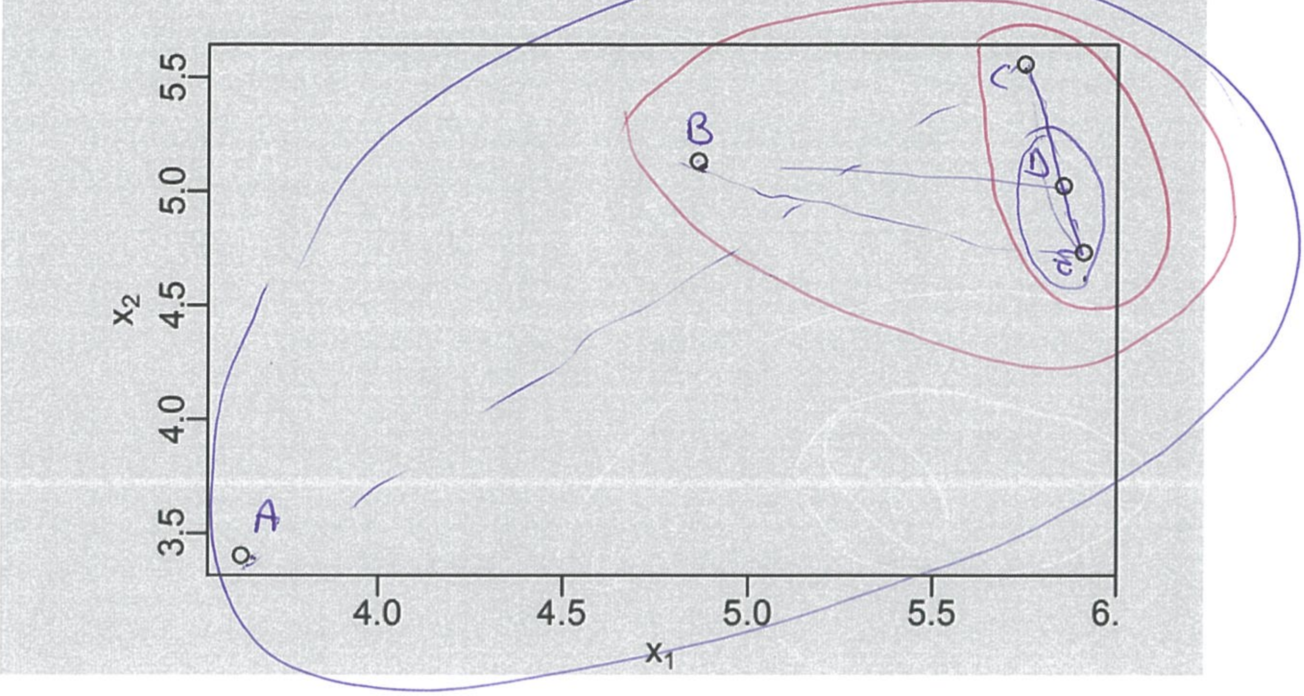
SSB can also be found by $SSB = SST - SSW$
$= 15.33 - 0.5$
$= 14.83$

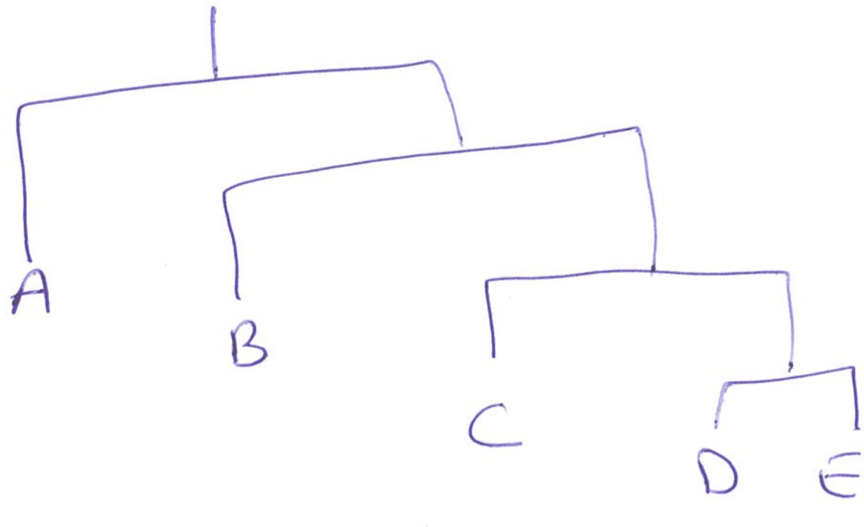

single distance
complete linkage

Let's find the hierarchy of clusters using the following set of points and Euclidean distance.



Find the hierarchy of clusters from the following plot using complete linkage.





51)

Single linkage clustering

	x_1	x_2	x_3	x_4	x_5
x_1	0	1	7	5	6
x_2	1	0	4	8	6
x_3	7	4	0	2	8
x_4	5	8	2	0	3
x_5	6	6	8	3	0

$\Rightarrow$ merge x_1 and x_2

	x_1-x_2	x_3	x_4	x_5
x_1-x_2	0	4	5	6
x_3	4	0	2	8
x_4	5	2	0	3
x_5	6	8	3	0

$\Rightarrow$ merge x_3 and x_4

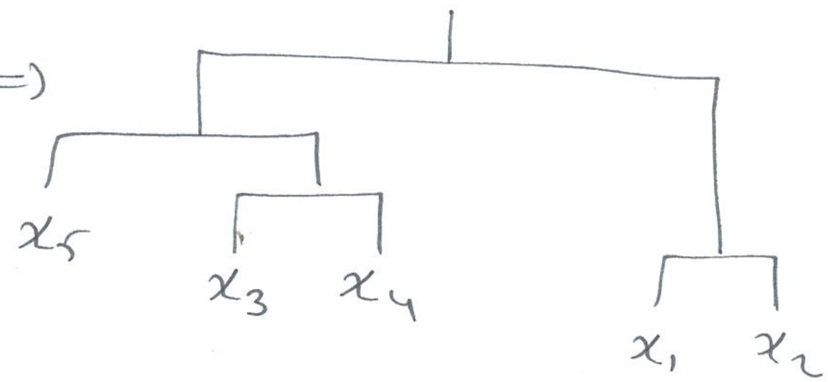
	$x_1 x_2$	$x_3 x_4$	x_5
$x_1 x_2$	0	4	6
$x_3 x_4$	4	0	3
x_5	6	3	0

$\Rightarrow$ merge $x_3 x_4$ and x_5

	$x_1 x_2$	$x_3 x_4 x_5$
$x_1 x_2$	0	4
$x_3 x_4 x_5$	4	0

$\Rightarrow$ then merge them all!

dendrogram $\Rightarrow$



52

Find the hierarchy of clusters from the following dissimilarity matrix using complete linkage clustering.

	x_1	x_2	x_3	x_4	x_5
x_1	0	3	7	8	9
x_2	3	0	5	7	6
x_3	7	5	0	2	1
x_4	8	7	2	0	4
x_5	9	6	1	4	0

$\Rightarrow$ merge x_3 and x_5

	x_1	x_2	$x_3 x_5$	x_4
x_1	0	3	9	8
x_2	3	0	6	7
$x_3 x_5$	9	6	0	4
x_4	8	7	4	0

$\Rightarrow$ merge x_1 and x_2

	$x_1 x_2$	$x_3 x_5$	x_4
$x_1 x_2$	0	9	8
$x_3 x_5$	9	0	4
x_4	8	4	0

$\Rightarrow$ merge $x_3 x_5$ and x_4

	$x_1 x_2$	$x_3 x_4 x_5$
$x_1 x_2$	0	9
$x_3 \cancel{x_4} x_4 x_5$	9	0

Dendrogram