Plot the observations in a two-dimensional graph.

Perform k-means clustering with *k* = 2 using the Euclidean distance norm. Toss a coin **seven** times to initiate the algorithm.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | i=1 | i=2 | i=3 | i=4 | i=5 | i=6 | i=7 |
| x\_i\_1 | 1 | 1 | 1 | 5 | 2 | 6 | 4 |
| x\_i\_2 | 4 | 3 | 2 | 1 | 3 | 2 | 1 |
| init clusters | 1 | 2 | 2 | 2 | 1 | 2 | 2 |

centroid1\_1 = (x\_1\_1 + x\_5\_1)/2 = (1+2)/2=1.5

centroid1\_2 = (x\_1\_2 + x\_5\_2)/2 = (4 + 3) /2 = 3.5

**centroid1 = (1.5, 3.5)**

centroid2\_1 = (x\_2\_1 + x\_3\_1 + x\_4\_1 +x\_6\_1 + x\_7\_1)/5 = (1 + 1+ 5 + 6 + 4)/5 = 17/5 = 3.4

centroid2\_2 = (x\_2\_2 + x\_3\_2 + x\_4\_2 + x\_6\_2 + x\_7\_2)/5 = (3 + 2 + 1 + 2 + 1) /5 = 1.8

**centroid2 = (3.4, 1.8)**

Calculate the distance to the centroid:

**Centroid1**

centroid1\_1 =1.5

centroid1\_2 = 3.5

dist\_1\_clusters\_c1 = dist((x\_1\_1, x\_1\_2), (centroid1\_1, centroid1\_2))= sqrt ( (x\_1\_1 - centroid1\_1)^2 +

+ (x\_1\_2 - centroid1\_2) ^2 ) = sqrt( (1 - 1.5) ^2 + (4 - 3.5)^2) = 0.7

dist\_2\_clusters\_c1 = dist((x\_2\_1, x\_2\_2), (centroid1\_1, centroid1\_2))= sqrt ( (x\_2\_1 - centroid1\_1)^2 +

+ (x\_2\_2 - centroid1\_2) ^2 ) = sqrt( (1 - 1.5) ^2 + (3 - 3.5)^2) = 0.7

dist\_3\_clusters\_c1 = dist((x\_3\_1, x\_3\_2), (centroid1\_1, centroid1\_2))= sqrt ( (x\_3\_1 - centroid1\_1)^2 +

+ (x\_3\_2 - centroid1\_2) ^2 ) = sqrt( (1 - 1.5) ^2 + (2 - 3.5)^2) = 1.58

dist\_4\_clusters\_c1 = dist((x\_4\_1, x\_4\_2), (centroid1\_1, centroid1\_2))= sqrt ( (x\_4\_1 - centroid1\_1)^2 +

+ (x\_4\_2 - centroid1\_2) ^2 ) = sqrt( (5 - 1.5) ^2 + (1 - 3.5)^2) = 4.3

dist\_5\_clusters\_c1 = dist((x\_5\_1, x\_5\_2), (centroid1\_1, centroid1\_2))= sqrt ( (x\_5\_1 - centroid1\_1)^2 +

+ (x\_5\_2 - centroid1\_2) ^2 ) = sqrt( (2 - 1.5) ^2 + (3 - 3.5)^2) = 0.7

dist\_6\_clusters\_c1 = dist((x\_6\_1, x\_6\_2), (centroid1\_1, centroid1\_2))= sqrt ( (x\_6\_1 - centroid1\_1)^2 +

+ (x\_6\_2 - centroid1\_2) ^2 ) = sqrt( (6 - 1.5) ^2 + (2 - 3.5)^2) = 4.7

dist\_7\_clusters\_c1 = dist((x\_7\_1, x\_7\_2), (centroid1\_1, centroid1\_2))= sqrt ( (x\_7\_1 - centroid1\_1)^2 +

+ (x\_7\_2 - centroid1\_2) ^2 ) = sqrt( (4 - 1.5) ^2 + (1 - 3.5)^2) = 3.5

**Centroid2**

centroid2\_1 = 3.4

centroid2\_2 = 1.8

dist\_1\_clusters\_c2 = dist((x\_1\_1, x\_1\_2), (centroid2\_1, centroid2\_2))= sqrt ( (x\_1\_1 – centroid2\_1)^2 +

+ (x\_2\_2 - centroid1\_2) ^2 ) = sqrt( (1 - 3.4) ^2 + (4 - 1.8)^2) = 3.3

dist\_2\_clusters\_c2 = dist((x\_2\_1, x\_2\_2), (centroid2\_1, centroid2\_2))= sqrt ( (x\_2\_1 – centroid2\_1)^2 +

+ (x\_2\_2 – centroid2\_2) ^2 ) = sqrt( (1 - 3.4) ^2 + (3 - 1.8)^2) = 2.7

dist\_3\_clusters\_c2 = dist((x\_3\_1, x\_3\_2), (centroid2\_1, centroid2\_2))= sqrt ( (x\_3\_1 – centroid2\_1)^2 +

+ (x\_3\_2 – centroid2\_2) ^2 ) = sqrt( (1 - 3.4) ^2 + (2 - 1.8)^2) = 2.4

dist\_4\_clusters\_c2 = dist((x\_4\_1, x\_4\_2), (centroid2\_1, centroid2\_2))= sqrt ( (x\_4\_1 – centroid2\_1)^2 +

+ (x\_4\_2 – centroid2\_2) ^2 ) = sqrt( (5 - 3.4) ^2 + (1 - 1.8)^2) = 1.8

dist\_5\_clusters\_c2 = dist((x\_5\_1, x\_5\_2), (centroid2\_1, centroid2\_2))= sqrt ( (x\_5\_1 – centroid2\_1)^2 +

+ (x\_5\_2 – centroid2\_2) ^2 ) = sqrt( (2 - 3.4) ^2 + (3 - 1.8)^2) = 1.8

dist\_6\_clusters\_c2 =dist((x\_6\_1, x\_6\_2), (centroid1\_1, centroid1\_2))= sqrt ( (x\_6\_1 – centroid2\_1)^2 +

+ (x\_6\_2 – centroid2\_2) ^2 ) = sqrt( (6 - 3.4) ^2 + (2 - 1.8)^2) = 2.6

dist\_7\_clusters\_c2 =dist((x\_7\_1, x\_7\_2), (centroid1\_1, centroid1\_2))= sqrt ( (x\_7\_1 – centroid2\_1)^2 +

+ (x\_7\_2 – centroid2\_2) ^2 ) = sqrt( (4 - 3.4) ^2 + (1 - 1.8)^2) = 1.0

**Comparison:**

dist\_1\_clusters\_c1 = 0.7 < dist\_1\_clusters\_c2 = 3.3 🡺 (x\_1\_1, x\_1\_2) = (1,4) going to **cluster 1**

dist\_2\_clusters\_c1 = 0.7 < dist\_2\_clusters\_c2 = 2.7 🡺 (x\_2\_1, x\_2\_2) = (1,3) going to **cluster 1**

dist\_3\_clusters\_c1 = 1.58 < dist\_3\_clusters\_c2 = 2.4 🡺 (x\_3\_1, x\_3\_2) = (1,2) going to **cluster 1**

dist\_4\_clusters\_c1 = 4.3 > dist\_4\_clusters\_c2 = 1.8 🡺 (x\_4\_1, x\_4\_2) = (5,1) going to **cluster 2**

dist\_5\_clusters\_c1 = 0.7 < dist\_5\_clusters\_c2 = 1.8 🡺 (x\_5\_1, x\_5\_2) = (2,3) going to **cluster 1**

dist\_6\_clusters\_c1 = 4.7 > dist\_6\_clusters\_c2 = 2.6 🡺 (x\_6\_1, x\_6\_2) = (6,2) going to **cluster 2**

dist\_7\_clusters\_c1 = 3.5 > dist\_7\_clusters\_c2 = 1.0 🡺 (x\_7\_1, x\_7\_2) = (4,1) going to **cluster 2**

Let's recalculate centroids again:

centroid1\_1 = (1+1+1+2)/4 = = 1.25

centroid1\_2 = (4+3+2+3) /4 =3.0

**centroid1 = (1.25, 3.0)**

centroid2\_1 = ( 5 + 6 + 4)/3 = 5.0

centroid2\_2 = (1 + 2 + 1) /3 = 1.33

**centroid2 = (5.0, 1.33)**

**Next,**

Calculate the distance to the centroid:

**Centroid1**

centroid1\_1 = 1.25

centroid1\_2 = 3.0

dist\_1\_clusters\_c1 = sqrt( (1 - 1.25) ^2 + (4 - 3)^2) = 1.0

dist\_2\_clusters\_c1 = sqrt( (1 - 1.25) ^2 + (3 - 3)^2) = 0.3

dist\_3\_clusters\_c1 = sqrt( (1 - 1.25) ^2 + (2 - 3)^2) = 1.0

dist\_4\_clusters\_c1 = sqrt( (5 - 1.25) ^2 + (1 - 3)^2) = 4.3

dist\_5\_clusters\_c1 = sqrt( (2 - 1.25) ^2 + (3 - 3)^2) = 0.8

dist\_6\_clusters\_c1 = sqrt( (6 - 1.25) ^2 + (2 - 3)^2) = 4.9

dist\_7\_clusters\_c1 = sqrt( (4 - 1.25) ^2 + (1 - 3)^2) = 3.4

**Next,**

**Centroid2**

centroid2\_1 = 5.0

centroid2\_2 = 1.33

dist\_1\_clusters\_c2 = sqrt( (1 - 5) ^2 + (4 - 1.33)^2) = 4.8

dist\_2\_clusters\_c2 = sqrt( (1 - 5) ^2 + (3 - 1.33)^2) = 4.3

dist\_3\_clusters\_c2 = sqrt( (1 - 5) ^2 + (2 - 1.33)^2) = 4.1

dist\_4\_clusters\_c2 = sqrt( (5 - 5) ^2 + (1 - 1.33)^2) = 0.3

dist\_5\_clusters\_c2 = sqrt( (2 - 5) ^2 + (3 - 1.33)^2) = 3.4

dist\_6\_clusters\_c2 = sqrt( (6 - 5) ^2 + (2 - 1.33)^2) = 1.2

dist\_7\_clusters\_c2 = sqrt( (4 - 5) ^2 + (1 - 1.33)^2) = 1.1

**Comparison:**

dist\_1\_clusters\_c1 = 1 < dist\_1\_clusters\_c2 = 4.8 🡺 (x\_1\_1, x\_1\_2) = (1,4) going to **cluster 1**

dist\_2\_clusters\_c1 = 0.3 < dist\_2\_clusters\_c2 = 4.3 🡺 (x\_2\_1, x\_2\_2) = (1,3) going to **cluster 1**

dist\_3\_clusters\_c1 = 1 < dist\_3\_clusters\_c2 = 4.1 🡺 (x\_3\_1, x\_3\_2) = (1,2) going to **cluster 1**

dist\_4\_clusters\_c1 = 4.3 > dist\_4\_clusters\_c2 = 0.3 🡺 (x\_4\_1, x\_4\_2) = (5,1) going to **cluster 2**

dist\_5\_clusters\_c1 = 0.8 < dist\_5\_clusters\_c2 = 3.4 🡺 (x\_5\_1, x\_5\_2) = (2,3) going to **cluster 1**

dist\_6\_clusters\_c1 = 4.9 > dist\_6\_clusters\_c2 = 1.2 🡺 (x\_6\_1, x\_6\_2) = (6,2) going to **cluster 2**

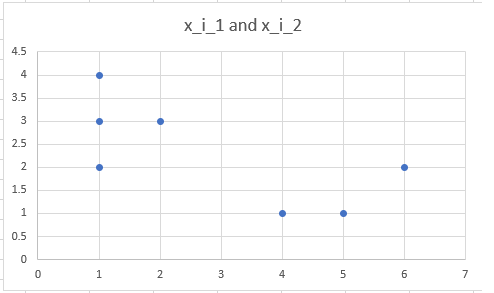
dist\_7\_clusters\_c1 = 3.4 > dist\_7\_clusters\_c2 = 1.1 🡺 (x\_7\_1, x\_7\_2) = (4,1) going to **cluster 2**

**We have to stop here, since our points do not change the clusters they hit.**

Cluster the data using hierarchical clustering with complete linkage and the Euclidean norm. Draw the resulting dendrogram.

Minimum Distance “Single Linkage”

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | i=1 | i=2 | i=3 | i=4 | i=5 | i=6 | i=7 |
| x\_i\_1 | 1 | 1 | 1 | 5 | 2 | 6 | 4 |
| x\_i\_2 | 4 | 3 | 2 | 1 | 3 | 2 | 1 |



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **(1,4)** | **(1,3)** | **(1,2)** | **(5,1)** | **(2, 3)** | **(6, 2)** | **(4, 1)** |
| **(1,4)** | 0 | 1 | 2 | 5 | 1.4 | 5.4 | 4.2 |
| **(1,3)** | 1 | 0 | 1 | 4.5 | 1 | 5.1 | 3.6 |
| **(1,2)** | 2 | 1 | 0 | 4.1 | 1.4 | 5 | 3.2 |
| **(5,1)** | 5 | 4.5 | 4.1 | 0 | 3.6 | 1.4 | 1 |
| **(2, 3)** | 1.4 | 1 | 1.4 | 3.6 | 0 | 4.1 | 2.8 |
| **(6, 2)** | 5.4 | 5.1 | 5 | 1.4 | 4.1 | 0 | 2.2 |
| **(4, 1)** | 4.2 | 3.6 | 3.2 | 1 | 2.8 | 2.2 | 0 |

dist((1, 4), (1, 3)) = sqrt( (1 - 1) ^2 + (3 - 4)^2) = 1

dist((1, 4), (1, 2)) = sqrt( (1 - 1) ^2 + (2 - 4)^2) = 2

dist((1, 4), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 4)^2) = 5

dist((1, 4), (2, 3)) = sqrt( (2 - 1) ^2 + (3 - 4)^2) = 1.4

dist((1, 4), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 4)^2) = 5.4

dist((1, 4), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 4)^2) = 4.2

dist((1, 3), (1, 2)) = sqrt( (1 - 1) ^2 + (2 - 3)^2) = 1

dist((1, 3), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 3)^2) = 4.5

dist((1, 3), (2, 3)) = sqrt( (2 - 1) ^2 + (3 - 3)^2) = 1

dist((1, 3), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 3)^2) = 5.1

dist((1, 3), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 3)^2) = 3.6

dist((1, 2), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 2)^2) = 4.1

dist((1, 2), (2, 3)) = sqrt( (2 - 1) ^2 + (3 - 2)^2) = 1.4

dist((1, 2), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 2)^2) = 5

dist((1, 2), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 2)^2) = 3.2

dist((2, 3), (5, 1)) = sqrt( (5 - 2) ^2 + (1 - 3)^2) = 3.6

dist((6, 2), (5, 1)) = sqrt( (5 - 6) ^2 + (1 - 2)^2) = 1.4

dist((4, 1), (5, 1)) = sqrt( (5 - 4) ^2 + (1 - 1)^2) = 1

dist((6, 2), (2, 3)) = sqrt( (2 - 6) ^2 + (3 - 2)^2) = 4.1

dist((4, 1), (2, 3)) = sqrt( (2 - 4) ^2 + (3 - 1)^2) = 2.8

dist((4, 1), (6, 2)) = sqrt( (6 - 4) ^2 + (2 - 1)^2) = 2.2

The smallest distance between points

**(1,3) and (1,4) is 1**

Converting the table:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **(1,4), (1,3) = (1, 3.5)** | **(1,2)** | **(5,1)** | **(2, 3)** | **(6, 2)** | **(4, 1)** |
| **(1,4), (1,3) = (1, 3.5)** | 0 | 1.5 | 4.7 | 1.1 | 5.2 | 3.9 |
| **(1,2)** | 1.5 | 0 | 4.1 | 1.4 | 5 | 3.2 |
| **(5,1)** | 4.7 | 4.1 | 0 | 3.6 | 1.4 | 1 |
| **(2, 3)** | 1.1 | 1.4 | 3.6 | 0 | 4.1 | 2.8 |
| **(6, 2)** | 5.2 | 5 | 1.4 | 4.1 | 0 | 2.2 |
| **(4, 1)** | 3.9 | 3.2 | 1 | 2.8 | 2.2 | 0 |

dist((1, 3.5), (1, 2)) = sqrt( (1 - 1) ^2 + (2 - 3.5)^2) = 1.5

dist((1, 3.5), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 3.5)^2) = 4.7

dist((1, 3.5), (2, 3)) = sqrt( (2 - 1) ^2 + (3 - 3.5)^2) = 1.1

dist((1, 3.5), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 3.5)^2) = 5.2

dist((1, 3.5), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 3.5)^2) = 3.9

dist((1, 2), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 2)^2) = 4.1

dist((1, 2), (2, 3)) = sqrt( (2 - 1) ^2 + (3 - 2)^2) = 1.4

dist((1, 2), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 2)^2) = 5

dist((1, 2), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 2)^2) = 3.2

dist((2, 3), (5, 1)) = sqrt( (5 - 2) ^2 + (1 - 3)^2) = 3.6

dist((6, 2), (5, 1)) = sqrt( (5 - 6) ^2 + (1 - 2)^2) = 1.4

dist((4, 1), (5, 1)) = sqrt( (5 - 4) ^2 + (1 - 1)^2) = 1

dist((6, 2), (2, 3)) = sqrt( (2 - 6) ^2 + (3 - 2)^2) = 4.1

dist((4, 1), (2, 3)) = sqrt( (2 - 4) ^2 + (3 - 1)^2) = 2.8

dist((4, 1), (6, 2)) = sqrt( (6 - 4) ^2 + (2 - 1)^2) = 2.2

Converting the table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **(1,4), (1,3) = (1, 3.5)** | **(1,2)** | **(5,1), (4,1) = (4.5, 1)** | **(2, 3)** | **(6, 2)** |
| **(1,4), (1,3) = (1, 3.5)** | 0 | 1.5 | 4.3 | 1.1 | 5.2 |
| **(1,2)** | 1.5 | 0 | 3.6 | 1.4 | 5 |
| **(5,1), (4.1) = (4.5, 1)** | 4.3 | 3.6 | 0 | 3.2 | 1.8 |
| **(2, 3)** | 1.1 | 1.4 | 3.2 | 0 | 4.1 |
| **(6, 2)** | 5.2 | 5 | 1.8 | 4.1 | 0 |

dist((1, 3.5), (4.5, 1)) = sqrt( (4.5 - 1) ^2 + (1 - 3.5)^2) = 4.3

dist((1, 2), (4.5, 1)) = sqrt( (1 – 4.5) ^2 + (2 - 1)^2) = 3.6

dist((4.5, 1), (2, 3)) = sqrt( (2 – 4.5) ^2 + (3 - 1)^2) = 3.2

dist((4.5, 1), (6, 2)) = sqrt( (6 – 4.5) ^2 + (2 - 1)^2) = 1.8

The smallest distance between points **(1, 3.5) and (2, 3) is 1.1 ,** Converting the table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **(2, 3), (1, 3.5) =**  **(1.5, 3.25)** | **(1,2)** | **(5,1), (4,1) = (4.5, 1)** | **(6, 2)** |
| **(2, 3), (1, 3.5) =(1.5, 3.25)** | 0 | 1.35 | 3.75 | 4.7 |
| **(1,2)** | 1.35 | 0 | 3.6 | 5 |
| **(5,1), (4.1) = (4.5, 1)** | 3.75 | 3.6 | 0 | 1.8 |
| **(6, 2)** | 4.7 | 5 | 1.8 | 0 |

dist((1, 2), (1.5, 3.25)) = sqrt( (1 – 1.5) ^2 + (2 – 3.25)^2) = 1.35

dist((4.5, 1), (1.5, 3.25)) = sqrt( (4.5 - 1.5) ^2 + (1 - 3.25)^2) = 3.75

dist((6, 2), (1.5, 3.25)) = sqrt( (6 - 1.5) ^2 + (2 - 3.25)^2) = 4.7

The smallest distance between points **(1.5, 3.25) and (1, 2) is 1.35 ,** Converting the table:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **(1.5, 3.25), (1, 2) = (1.25, 2.6)** | **(5,1), (4,1) = (4.5, 1)** | **(6, 2)** |
| **(1.5, 3.25), (1, 2) = (1.25, 2.6)** | 0 | 3.6 | 4.8 |
| **(5,1), (4.1) = (4.5, 1)** | 3.6 | 0 | 1.8 |
| **(6, 2)** | 4.8 | 1.8 | 0 |

dist((4.5, 1), (1.25, 2.6)) = sqrt( (4.5 - 1.25) ^2 + (1 - 2.6)^2) = 3.6

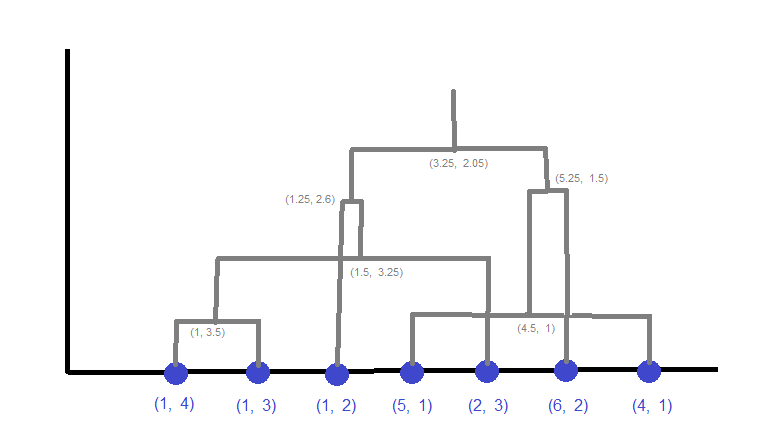
dist((6, 2), (1.25, 2.6)) = sqrt( (6 - 1.25) ^2 + (2 - 2.6)^2) = 4.8

The smallest distance between points **(4.5, 1) and (6, 2) is 1.8 ,** Converting the table:

|  |  |  |
| --- | --- | --- |
|  | **(1.5, 3.25), (1, 2) = (1.25, 2.6)** | **(4.5, 1), (6, 2) = (5.25, 1.5)** |
| **(1.5, 3.25), (1, 2) = (1.25, 2.6)** | 0 | 4.1 |
| **(4.5, 1), (6, 2) = (5.25, 1.5)** | 4.1 | 0 |

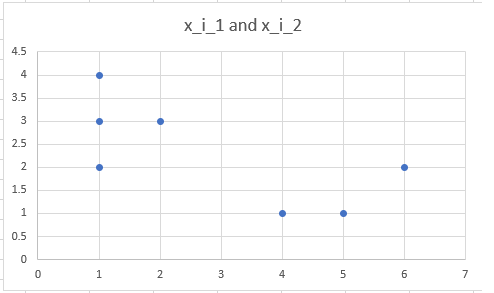
dist((1.25, 2.6), (5.25, 1.5)) = sqrt( (1.25 - 5.25) ^2 + (2.6 - 1.5)^2) = 4.1

|  |  |
| --- | --- |
|  | **(1.25, 2.6), (5.25, 1.5) = (3.25, 2.05)** |
| **(1.25, 2.6), (5.25, 1.5) = (3.25, 2.05)** | 0 |



Maximum Distance “Complete Linkage”

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | i=1 | i=2 | i=3 | i=4 | i=5 | i=6 | i=7 |
| x\_i\_1 | 1 | 1 | 1 | 5 | 2 | 6 | 4 |
| x\_i\_2 | 4 | 3 | 2 | 1 | 3 | 2 | 1 |



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **(1,4)** | **(1,3)** | **(1,2)** | **(5,1)** | **(2, 3)** | **(6, 2)** | **(4, 1)** |
| (1,4) | 0 | 1 | 2 | 5 | 1.4 | 5.4 | 4.2 |
| (1,3) | 1 | 0 | 1 | 4.5 | 1 | 5.1 | 3.6 |
| (1,2) | 2 | 1 | 0 | 4.1 | 1.4 | 5 | 3.2 |
| (5,1) | 5 | 4.5 | 4.1 | 0 | 3.6 | 1.4 | 1 |
| (2, 3) | 1.4 | 1 | 1.4 | 3.6 | 0 | 4.1 | 2.8 |
| (6, 2) | 5.4 | 5.1 | 5 | 1.4 | 4.1 | 0 | 2.2 |
| (4, 1) | 4.2 | 3.6 | 3.2 | 1 | 2.8 | 2.2 | 0 |

dist((1, 4), (1, 3)) = sqrt( (1 - 1) ^2 + (3 - 4)^2) = 1

dist((1, 4), (1, 2)) = sqrt( (1 - 1) ^2 + (2 - 4)^2) = 2

dist((1, 4), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 4)^2) = 5

dist((1, 4), (2, 3)) = sqrt( (2 - 1) ^2 + (3 - 4)^2) = 1.4

dist((1, 4), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 4)^2) = 5.4

dist((1, 4), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 4)^2) = 4.2

dist((1, 3), (1, 2)) = sqrt( (1 - 1) ^2 + (2 - 3)^2) = 1

dist((1, 3), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 3)^2) = 4.5

dist((1, 3), (2, 3)) = sqrt( (2 - 1) ^2 + (3 - 3)^2) = 1

dist((1, 3), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 3)^2) = 5.1

dist((1, 3), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 3)^2) = 3.6

dist((1, 2), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 2)^2) = 4.1

dist((1, 2), (2, 3)) = sqrt( (2 - 1) ^2 + (3 - 2)^2) = 1.4

dist((1, 2), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 2)^2) = 5

dist((1, 2), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 2)^2) = 3.2

dist((2, 3), (5, 1)) = sqrt( (5 - 2) ^2 + (1 - 3)^2) = 3.6

dist((6, 2), (5, 1)) = sqrt( (5 - 6) ^2 + (1 - 2)^2) = 1.4

dist((4, 1), (5, 1)) = sqrt( (5 - 4) ^2 + (1 - 1)^2) = 1

dist((6, 2), (2, 3)) = sqrt( (2 - 6) ^2 + (3 - 2)^2) = 4.1

dist((4, 1), (2, 3)) = sqrt( (2 - 4) ^2 + (3 - 1)^2) = 2.8

dist((4, 1), (6, 2)) = sqrt( (6 - 4) ^2 + (2 - 1)^2) = 2.2

The smallest distance between points (clusters)

**(1,3) and (1,4) is 1**

Combine columns (1,3) and (1,4) into one, forming a new cluster :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **(1,4), (1,3)** | **(1,2)** | **(5,1)** | **(2, 3)** | **(6, 2)** | **(4, 1)** |
| **(1,4), (1,3)** | 0 | 1.5 | 4.7 | 1.1 | 5.2 | 3.9 |
| **(1,2)** | 1.5 | 0 | 4.1 | 1.4 | 5 | 3.2 |
| **(5,1)** | 4.7 | 4.1 | 0 | 3.6 | 1.4 | 1 |
| **(2, 3)** | 1.1 | 1.4 | 3.6 | 0 | 4.1 | 2.8 |
| **(6, 2)** | 5.2 | 5 | 1.4 | 4.1 | 0 | 2.2 |
| **(4, 1)** | 3.9 | 3.2 | 1 | 2.8 | 2.2 | 0 |

We need to find the distance from this cluster to all other points. We take in turn every other point from the remaining 5, and for this selected point we find the distance to two points from the cluster, and choose the one that is larger.

dist((1, 4), (…))

dist((1, 4), (1, 2)) = sqrt( (1 - 1) ^2 + (2 - 4)^2) = 2.0

dist((1, 4), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 4)^2) = 5.0

dist((1, 4), (2, 3)) = sqrt( (2 - 1) ^2 + (3 - 4)^2) = 1.4

dist((1, 4), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 4)^2) = 5.4

dist((1, 4), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 4)^2) = 4.2

dist((1, 3), (…))

dist((1, 3), (1, 2)) = sqrt( (1 - 1) ^2 + (2 - 3)^2) = 1.0

dist((1, 3), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 3)^2) = 4.5

dist((1, 3), (2, 3)) = sqrt( (2 - 1) ^2 + (3 - 3)^2) = 1.0

dist((1, 3), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 3)^2) = 5.1

dist((1, 3), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 3)^2) = 3.6

dist((1, 4), (1, 2)) = 2.0 > dist((1, 3), (1, 2)) = 1.0

* The distance between our cluster ((1.4), (1,2)) and the cluster (1,2) (which is point) should be 2.0 as a maximum distance.

Let's fill the table according to this algorithm (yellow color).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **(1,4), (1,3)** | **(1,2)** | **(5,1)** | **(2, 3)** | **(6, 2)** | **(4, 1)** |
| **(1,4), (1,3)** | 0 | 2 | 5.0 | 1.4 | 5.4 | 4.2 |
| **(1,2)** | 2 | 0 | 4.1 | 1.4 | 5 | 3.2 |
| **(5,1)** | 5.0 | 4.1 | 0 | 3.6 | 1.4 | 1 |
| **(2, 3)** | 1.4 | 1.4 | 3.6 | 0 | 4.1 | 2.8 |
| **(6, 2)** | 5.4 | 5 | 1.4 | 4.1 | 0 | 2.2 |
| **(4, 1)** | 4.2 | 3.2 | 1 | 2.8 | 2.2 | 0 |

The smallest distance between clusters

**(4,1) and (5,1) is 1**

Combine columns (4,1) and (5,1) into one, forming a new cluster :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **(1,4), (1,3)** | **(1,2)** | **(5,1), (4,1)** | **(2, 3)** | **(6, 2)** |
| **(1,4), (1,3)** | 0 | 2 | 5.0 | 1.4 | 5.4 |
| **(1,2)** | 2 | 0 | 4.1 | 1.4 | 5 |
| **(5,1), (4,1)** | 5.0 | 4.1 | 0 | 3.6 | 2.2 |
| **(2, 3)** | 1.4 | 1.4 | 3.6 | 0 | 4.1 |
| **(6, 2)** | 5.4 | 5 | 2.2 | 4.1 | 0 |

Let's fill the table according to the algorithm as a maximum distance (yellow color).

dist((4, 1), (…)) dist((5, 1), (…))

dist((4, 1), (1, 2)) = sqrt( (1 - 4) ^2 + (2 - 1)^2) = 3.2

dist((4, 1), (2, 3)) = sqrt( (2 - 4) ^2 + (3 - 1)^2) = 2.8

dist((4, 1), (6, 2)) = sqrt( (6 - 4) ^2 + (2 - 1)^2) = 2.2

dist((5, 1), (1, 2)) = sqrt( (1 - 5) ^2 + (2 - 1)^2) = 4.1

dist((5, 1), (2, 3)) = sqrt( (2 - 5) ^2 + (3 - 1)^2) = 3.6

dist((5, 1), (6, 2)) = sqrt( (6 - 5) ^2 + (2 - 1)^2) = 1.4

dist((4, 1), (1, 4)) = sqrt( (1 - 4) ^2 + (4 - 1)^2) = 4.2

dist((4, 1), (1, 3)) = sqrt( (1 - 4) ^2 + (3 - 1)^2) = 3.6

dist((5, 1), (1, 4)) = sqrt( (1 - 5) ^2 + (4 - 1)^2) = 5.0

dist((5, 1), (1, 3)) = sqrt( (1 - 5) ^2 + (3 - 1)^2) = 4.5

The smallest distance between clusters

**(1,2) and (2,3) is 1.4**

Combine columns (1,2) and (2,3) into one, forming a new cluster :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **(1,4), (1,3)** | **(1,2), (2, 3)** | **(5,1), (4,1)** | **(6, 2)** |
| **(1,4), (1,3)** | 0 | 2.0 | 5.0 | 5.4 |
| **(1,2), (2,3)** | 2.0 | 0 | 4.1 | 5.0 |
| **(5,1), (4,1)** | 5.0 | 4.1 | 0 | 2.2 |
| **(6, 2)** | 5.4 | 5.0 | 2.2 | 0 |

dist((1, 2), (6,2)) dist((2, 3), (6,2))

dist((1, 2), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 2)^2) = 5.0

dist((2, 3), (6, 2)) = sqrt( (6 - 2) ^2 + (2 - 3)^2) = 4.1

**dist(((1,2), (2, 3)) and ((1,4), (1,3)))**

dist((1, 2), (1, 4)) = sqrt( (1 - 1) ^2 + (4 - 2)^2) = 2.0

dist((1, 2), (1, 3)) = sqrt( (1 - 1) ^2 + (3 - 2)^2) = 1.0

dist((2, 3), (1, 4)) = sqrt( (1 - 2) ^2 + (4 - 3)^2) = 1.4

dist((2, 3), (1, 3)) = sqrt( (1 - 2) ^2 + (3 - 3)^2) = 1.0

**dist(((1,2), (2, 3)) and ((5,1), (4,1)))**

dist((1, 2), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 2)^2) = 4.1

dist((1, 2), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 2)^2) = 3.1

dist((2, 3), (5, 1)) = sqrt( (5 - 2) ^2 + (1 - 3)^2) = 3.6

dist((2, 3), (4, 1)) = sqrt( (4 - 2) ^2 + (1 - 3)^2) = 2.8

The smallest distance between clusters

**((1,2) , (2,3)) and ((1,4),(1,3)) is 2.0**

Combine columns into one, forming a new cluster :

|  |  |  |  |
| --- | --- | --- | --- |
|  | **(1,4), (1,3), (1,2), (2, 3)** | **(5,1), (4,1)** | **(6, 2)** |
| **(1,4), (1,3), (1,2), (2, 3)** | 0 | 5.0 | 5.3 |
| **(5,1), (4,1)** | 5.0 | 0 | 2.2 |
| **(6, 2)** | 5.3 | 2.2 | 0 |

**dist(((1,4), (1,3), (1,2), (2, 3)) and ((5,1),(4,1)))**

dist((1, 2), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 2)^2) = 4.1

dist((1, 2), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 2)^2) = 3.1

dist((2, 3), (5, 1)) = sqrt( (5 - 2) ^2 + (1 - 3)^2) = 3.6

dist((2, 3), (4, 1)) = sqrt( (4 - 2) ^2 + (1 - 3)^2) = 2.8

dist((1, 4), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 4)^2) = 5.0

dist((1, 4), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 4)^2) = 4.2

dist((1, 3), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 3)^2) = 4.5

dist((1, 3), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 3)^2) = 3.6

**dist(((1,4), (1,3), (1,2), (2, 3)) and (6, 2))**

dist((1, 2), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 2)^2) = 5.0

dist((2, 3), (6, 2)) = sqrt( (6 - 2) ^2 + (2 - 3)^2) = 4.1

dist((1, 4), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 4)^2) = 5.3

dist((1, 3), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 3)^2) = 5.1

The smallest distance between clusters

**((5,1) , (4,1)) and (6, 2)) is 2.2**

Combine columns into one, forming a new cluster :

|  |  |  |
| --- | --- | --- |
|  | **(1,4), (1,3), (1,2), (2, 3)** | **(5,1), (4,1), (6,2)** |
| **(1,4), (1,3), (1,2), (2, 3)** | 0 | 5.1 |
| **(5,1), (4,1), (6,2)** | 5.1 | 0 |

dist((1, 2), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 2)^2) = 5.0

dist((2, 3), (6, 2)) = sqrt( (6 - 2) ^2 + (2 - 3)^2) = 4.1

dist((1, 4), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 4)^2) = 5.3

dist((1, 3), (6, 2)) = sqrt( (6 - 1) ^2 + (2 - 3)^2) = 5.1

dist((1, 2), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 2)^2) = 3.2

dist((2, 3), (4, 1)) = sqrt( (4 - 2) ^2 + (1 - 3)^2) = 2.8

dist((1, 4), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 4)^2) = 4.2

dist((1, 3), (4, 1)) = sqrt( (4 - 1) ^2 + (1 - 3)^2) = 3.6

dist((1, 2), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 2)^2) = 4.1

dist((2, 3), (5, 1)) = sqrt( (5 - 2) ^2 + (1 - 3)^2) = 3.6

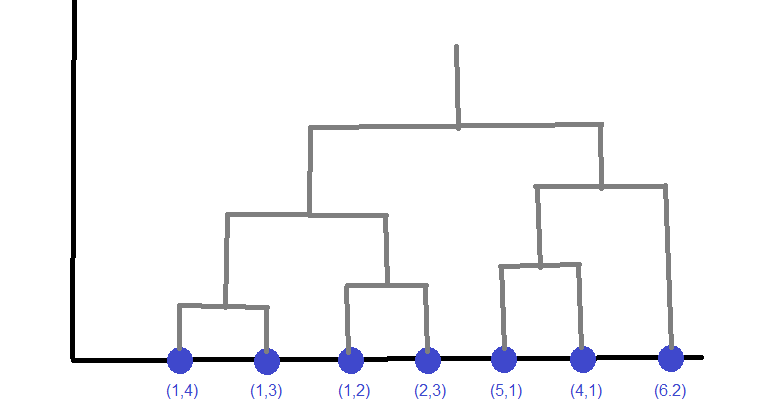
dist((1, 4), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 4)^2) = 5.0

dist((1, 3), (5, 1)) = sqrt( (5 - 1) ^2 + (1 - 3)^2) = 4.5

**Finally, we have one cluster!**

|  |  |
| --- | --- |
|  | **(1,4), (1,3), (1,2), (2, 3) ,**  **(5,1), (4,1), (6,2)** |
| **(1,4), (1,3), (1,2), (2, 3) ,**  **(5,1), (4,1), (6,2)** | 0 |

**Dendrogram** according to the algorithm as a maximum distance

****