

K-means clustering

2] a. $\phi(x_1) = [10, 0]$ $\phi(x_3) = [10, 20]$
 $\phi(x_2) = [30, 0]$ $\phi(x_4) = [20, 20]$

i. Initial centers: $u_1 = [20, 30]$ $u_2 = [20, -10]$

Distance from ^{initial} centers -

Point $x_1 \rightarrow d_{x_1 u_1} = \sqrt{(20-20)^2 + (30-30)^2} = \sqrt{100+900} = 31.62$
 $d_{x_1 u_2} = \sqrt{(20-20)^2 + (30-(-10))^2} = \sqrt{0+400} =$
 $d_{x_1 u_2} = \sqrt{(10-20)^2 + (0-(-10))^2} = \sqrt{100+100} = 14.14$

Point $x_2 \rightarrow d_{x_2 u_1} = \sqrt{(30-20)^2 + (0-30)^2} = \sqrt{100+900} = 31.62$
 $d_{x_2 u_2} = \sqrt{(30-20)^2 + (0-(-10))^2} = \sqrt{100+100} = 14.14$

Point $x_3 \rightarrow d_{x_3 u_1} = \sqrt{(10-20)^2 + (20-30)^2} = \sqrt{100+100} = 14.14$
 $d_{x_3 u_2} = \sqrt{(10-20)^2 + (20-(-10))^2} = \sqrt{100+900} = 31.62$

Point $x_4 \rightarrow d_{x_4 u_1} = \sqrt{(20-20)^2 + (20-30)^2} = \sqrt{0+100} = 10$
 $d_{x_4 u_2} = \sqrt{(20-20)^2 + (20-(-10))^2} = \sqrt{900} = 30$

Assignments -

Point x_1 and x_2 assigned to cluster 2 (u_2)

Point x_3 and x_4 assigned to cluster 1 (u_1)

Update cluster centers -

$u_1 = \text{mean}(x_3, x_4) = \left[\frac{10+20}{2}, \frac{20+20}{2} \right] = [15, 20] \rightarrow \text{New } u_1$

$u_2 = \text{mean}(x_1, x_2) = \left[\frac{10+30}{2}, \frac{0+0}{2} \right] = [20, 0] \rightarrow \text{New } u_2$

Distance from new centers - $u_1' = [15, 20]$ $u_2' = [20, 0]$

Point $x_1 \rightarrow d_{x_1 u_1'} = \sqrt{(10-15)^2 + (0-20)^2} = \sqrt{25+400} = 20.62$
 $d_{x_1 u_2'} = \sqrt{(10-20)^2 + (0-0)^2} = \sqrt{100} = 10$

Point $x_2 \rightarrow d_{x_2 u_1'} = \sqrt{(30-15)^2 + (0-20)^2} = \sqrt{225+400} = 25$
 $d_{x_2 u_2'} = \sqrt{(30-20)^2 + (0-0)^2} = \sqrt{100+0} = 10$

Point $x_3 \rightarrow d_{x_3 u_1'} = \sqrt{(10-15)^2 + (20-20)^2} = \sqrt{25+0} = 5$
 $d_{x_3 u_2'} = \sqrt{(10-20)^2 + (20-0)^2} = \sqrt{100+400} = 22.3$

Point $x_4 \rightarrow d_{x_4 u_1'} = \sqrt{(20-15)^2 + (20-20)^2} = \sqrt{25+0} = 5$
 $d_{x_4 u_2'} = \sqrt{(20-20)^2 + (20-0)^2} = \sqrt{0+400} = 20$

Assignments -

Point x_1 and x_2 are still assigned to cluster 2 (u_2)

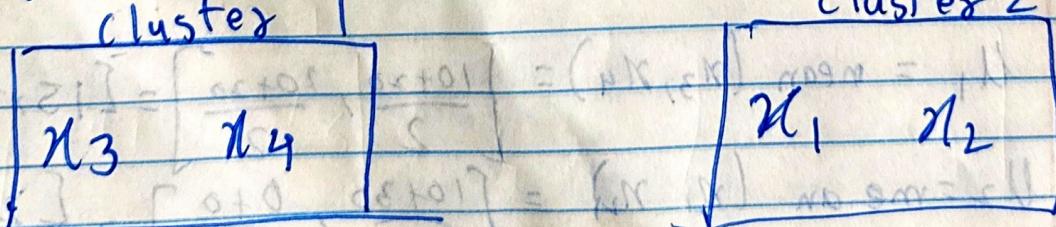
Point x_3 and x_4 are still assigned to cluster 1 (u_1)

Convergence -

Since, the assignments didn't change after the second iteration, we do not need to update the centers.

This means, convergence has been reached.

Clusters -



ii] Initial Centers - $U_1 = [10, 20]$ $U_2 = [20, 30]$

distances from initial centers -

$$\text{Point } x_1 - d_{x_1 U_1} = \sqrt{(10-0)^2 + (0-10)^2} = \sqrt{100+100} = 14.14$$

$$d_{x_1 U_2} = \sqrt{(10-30)^2 + (0-20)^2} = \sqrt{400+400} = 28.28$$

$$\text{Point } x_2 - d_{x_2 U_1} = \sqrt{(30-0)^2 + (0-10)^2} = \sqrt{900+100} = 31.62$$

$$d_{x_2 U_2} = \sqrt{(30-30)^2 + (0-20)^2} = \sqrt{0+400} = 20$$

$$\text{Point } x_3 - d_{x_3 U_1} = \sqrt{(10-0)^2 + (20-10)^2} = \sqrt{100+100} = 14.14$$

$$d_{x_3 U_2} = \sqrt{(10-30)^2 + (20-20)^2} = \sqrt{400+0} = 20.00$$

$$\text{Point } x_4 - d_{x_4 U_1} = \sqrt{(20-0)^2 + (20-10)^2} = \sqrt{400+100} = 22.36$$

$$d_{x_4 U_2} = \sqrt{(20-30)^2 + (20-20)^2} = \sqrt{100+0} = 10$$

Assignments -

Point x_1 and Point x_3 assigned to cluster 1 (U_1)

Point x_2 and Point x_4 assigned to cluster 2 (U_2)

Update Centers -

$$U_1' = \text{mean}(x_1, x_3) = \left[\frac{10+10}{2}, \frac{0+20}{2} \right] = [10, 10] \rightarrow \text{New } U_1$$

$$U_2' = \text{mean}(x_2, x_4) = \left[\frac{30+20}{2}, \frac{0+20}{2} \right] = [25, 10] \rightarrow \text{New } U_2$$

Distance from new centers - $u_1' = [10, 10]$ $u_2' = [25, 10]$

$$\text{Point } x_1 - d_{x_1 u_1'} = \sqrt{(10-10)^2 + (0-10)^2} = \sqrt{0+100} = 10$$

$$d_{x_1 u_2'} = \sqrt{(10-25)^2 + (0-10)^2} = \sqrt{225+100} = 18.03$$

$$\text{Point } x_2 - d_{x_2 u_1'} = \sqrt{(30-10)^2 + (0-10)^2} = \sqrt{400+100} = 22.36$$

$$d_{x_2 u_2'} = \sqrt{(30-25)^2 + (0-10)^2} = \sqrt{25+100} = 11.18$$

$$\text{Point } x_3 - d_{x_3 u_1'} = \sqrt{(10-10)^2 + (20-10)^2} = \sqrt{0+100} = 10$$

$$d_{x_3 u_2'} = \sqrt{(10-25)^2 + (20-10)^2} = \sqrt{225+100} = 18.03$$

$$\text{Point } x_4 - d_{x_4 u_1'} = \sqrt{(20-10)^2 + (20-10)^2} = \sqrt{100+100} = 14.14$$

$$d_{x_4 u_2'} = \sqrt{(20-25)^2 + (20-10)^2} = \sqrt{25+100} = 11.18$$

Assignments -

Point x_1 and x_3 are still assigned to cluster 1 (u_1')
 Point x_2 and x_4 are still assigned to cluster 2 (u_2')

Convergence -

Since the assignments didn't change after the 2nd iteration, we do not need to update the centers

This means, convergence has been reached.

Clusters -

