K - Means clustering:

Data points:

	a_1	a_2	a_3	a_4	a_5	a_6
x_1	1	2	2	3	4	5
x_2	1	1	3	2	3	5

Step 1: Choose no. of clusters i.e k

$K = 2$

Step 2: Choose random 2 initial cluster centers:

Example $V_1 = (2, 1)$, $V_2 (2, 3)$

Step 3: Find the distance between cluster centers and data points

DISTANCE TABLE:

Distance Formula = Euclidian distance $= d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$

Data point	Distance from V_1 (2,1)	Distance from V_2 (2,3)	Assigned cluster
a_1 (1,1)	1	2.24	V_1
a_2 (2,1)	0	2	V_1
a_3 (2,3)	2	0	V_2
a_4 (3,2)	1.41	1.41	V_1 [Either V_1 or V_2]
a_5 (4,3)	2.83	2	V_2
a_6 (5,5)	5	3.61	V_2

Example distance Calculation:

a_1 x_1 x_2

a_1 (1, 1) → (x_1, x_2)

y_1 y_2

V_1 (2, 1) → (y_1, y_2)

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

$$d = \sqrt{(1-2)^2 + (1-1)^2}$$

$$= \sqrt{(-1)^2 + (0)^2}$$

$$= \sqrt{1+0} = \sqrt{1} = 1$$

Step 4:

cluster 1 : $(V_1) = \{a_1, a_2, a_4\}$

cluster 2 : $(V_2) = \{a_3, a_5, a_6\}$

Step 5:

Recalculate the cluster center

$$V_1 = \frac{1}{3}\left[\overset{a_1}{(1,1)} + \overset{a_2}{(2,1)} + \overset{a_4}{(3,2)}\right]$$

$$= \frac{1}{3}[6, 4] = (2, 1.33)$$

$$V_2 = \frac{1}{3}\left[\overset{a_3}{(2,3)} + \overset{a_5}{(4,3)} + \overset{a_6}{(5,5)}\right]$$

$$= \frac{1}{3}[11, 11] = (3.67, 3.67)$$

Repeat the steps from step 3. until we get same cluster center or no change in cluster data points.

Distance Table:

Data point	Distance from V_1 (2, 1.33)	Distance from V_2 (3.67, 3.67)	Assigned cluster
a_1 (1,1)	1.05	3.78	V_1
a_2 (2,1)	0.33	3.15	V_1
a_3 (2,3)	1.67	1.8	V_1
a_4 (3,2)	1.204	1.8	V_1
a_5 (4,3)	2.605	0.75	V_2
a_6 (5,5)	4.75	1.88	V_2

Clusters

$V_1 = \{a_1, a_2, a_3, a_4\}$

$V_2 = \{a_5, a_6\}$

Step: Recalculate the cluster center

$V_1 = \frac{1}{4}[(1,1) + (2,1) + (2,3) + (3,2)]$

$= \frac{1}{4}[8,7] = (2, 1.75)$

$V_2 = \frac{1}{2}[(4,3) + (5,5)]$

$= \frac{1}{2}[9,8] = (4.5, 4)$

Step : Repeat from Step 3.

Distance Table:

Data point	Distance from V_1 (2, 1.75)	Distance from V_2 (4.5, 4)	Assigned cluster
a_1 (1,1)	1.25	4.61	V_1
a_2 (2,1)	0.75	3.9	V_1
a_3 (2,3)	1.25	2.69	V_1
a_4 (3,2)	1.03	2.5	V_1
a_5 (4,3)	2.36	1.12	V_2
a_6 (5,5)	4.48	1.12	V_2

Clusters

$V_1 = \{a_1, a_2, a_3, a_4\}$

$V_2 = \{a_5, a_6\}$

Hence, no change in the cluster's data points, we can stop the procedure.

— x —