



MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A Constituent unit of MAHE, Manipal)

(I)

1. use k means clustering algorithm to divide the following data into two clusters. we choose $(2,1)$ and $(2,3)$ as initial centroids.

x_1	1	2	3	3	4	5
x_2	1	1	3	2	3	5

(x_1, y_1) (x_2, y_2)

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

Iteration 1

Data point	Distance from $v_1 (2,1)$	Distance from $v_2 (2,3)$	Assigned cluster
$(1,1)$	1	2.24	v_1
$(2,1)$	0	2	v_1
$(2,3)$	2	0	v_2
$(3,2)$	1.41	1.41	v_1
$(4,3)$	2.83	2	v_2
$(5,5)$	5	3.61	v_2

(2)

Step 3:- cluster 1 of $v_1: \{a_1, a_2, a_4\}$

cluster 2 of $v_2: \{a_3, a_5, a_6\}$

$$v_1: (1,1) (2,1) (3,2) \Rightarrow \left\{ \frac{1+2+3}{3}, \frac{1+1+2}{3} \right\}$$

$$\Rightarrow \{2, 1.33\}$$

$$v_2: (2,3) (4,3) (5,5) \Rightarrow \left\{ \frac{2+4+5}{3}, \frac{3+3+5}{3} \right\}$$

$$\Rightarrow \{3.666, 3.666\}$$

$$\Rightarrow \{3.67, 3.67\}$$

Step 4:- Repeat from step 2 until we get same cluster center or same cluster elements as in the previous iteration.



Iteration 2

Data point	Distance from $V_1 (2, 1.33)$	Distance from $V_2 (3.67, 3.67)$	Assigned center
$a_1 (1, 1)$	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
$a_6 (5, 5)$	4.74	1.88	V_2

Cluster 1 of $V_1 = \{a_1, a_2, a_3, a_4\}$

Cluster 2 of $V_2 = \{a_5, a_6\}$

$$V_1 = \left\{ \frac{1+2+2+3}{4}, \frac{1+1+3+2}{4} \right\}$$

$$= \{2, 1.75\}$$

$$V_2 = \left\{ \frac{4+5}{2}, \frac{3+5}{2} \right\} = \{4.5, 4\}$$

Iteration 3

4

Data points	Distance from V_1 (2, 1.75)	Distance from V_2 (4.5, 4)	Assigned center
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)	4.42	1.12	V_2

cluster 1 of $V_1 = \{a_1, a_2, a_3, a_4\}$

cluster 2 of $V_2 = \{a_5, a_6\}$

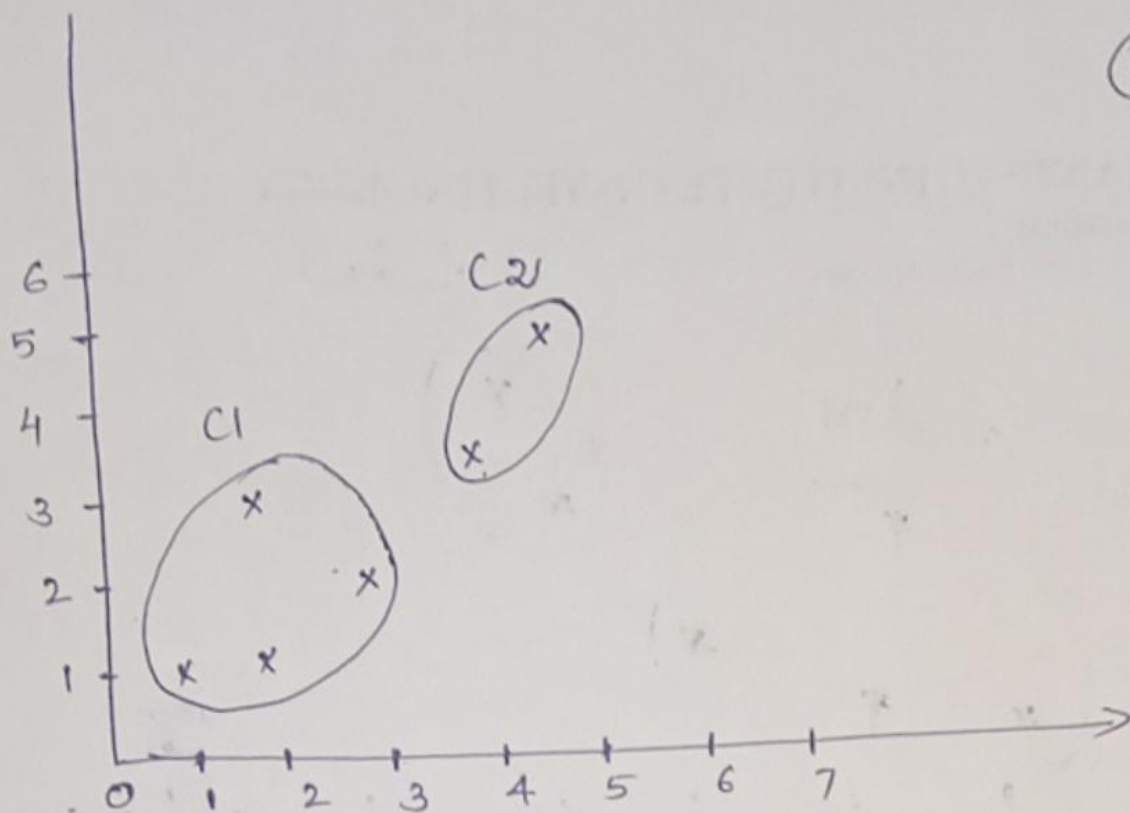
cluster elements are same as in the previous iteration

So the final clusters are

cluster 1 :- $\{(1, 1) (2, 1) (2, 3) (3, 2)\}$

cluster 2 :- $\{(4, 3) (5, 5)\}$

⑤.



Home work k-means

- ① use k means clustering algorithm to divide the following data points into 3 clusters.
- $A_1(2, 10)$, $A_2(2, 5)$, $A_3(8, 4)$, $B_1(5, 8)$
 $B_2(7, 5)$, $B_3(6, 4)$, $C_1(1, 2)$, $C_2(4, 9)$.
- The distance fn is Euclidean distance. initial
- Random clusters are A_1 , B_1 and C_1 as the centers of each cluster respectively.