Example 1:

- Dataset shown below consists of seven customers' ratings of two products A and B.
- Ranking represents the number of points that each customer has given to a product. The more points given, the higher the product is ranked.
- Use a k-means algorithm, where $k = 2$, to partition the dataset into two clusters C_1 and C_2.

Example 1:

- Two randomly selected centroids (also called cluster representatives) are Customers 1 and 5 (i.e., data point $\boldsymbol{p}_1 = (2, 2)$ and $\boldsymbol{p}_5 = (10, 14)$).
- We have $C_1 = \{\boldsymbol{p}_1\}$, $C_2 = \{\boldsymbol{p}_5\}$, $\boldsymbol{c}_1 = (2, 2)$, $\boldsymbol{c}_2 = (10, 14)$.

	Ratings of A	*Ratings of B*
1	**2**	**2**
2	3	4
3	6	8
4	7	10
5	**10**	**14**
6	9	10
7	7	9

Example 1: Iteration 1

$d(\boldsymbol{p}_1, \boldsymbol{c}_1) = \mathbf{0}$, $d(\boldsymbol{p}_1, \boldsymbol{c}_2) = 14.42$

No data point movement: $C_1 = \{\boldsymbol{p}_1\}$, $C_2 = \{\boldsymbol{p}_5\}$

No centroids update: $\boldsymbol{c}_1 = (2, 2)$, $\boldsymbol{c}_2 = (10, 14)$

$d(\boldsymbol{p}_2, \boldsymbol{c}_1) = \mathbf{2.24}$, $d(\boldsymbol{p}_2, \boldsymbol{c}_2) = 12.21$

Move $\boldsymbol{p}_2$ to C_1: $C_1 = \{\boldsymbol{p}_1, \boldsymbol{p}_2\}$, $C_2 = \{\boldsymbol{p}_5\}$

Centroids updated:

$\boldsymbol{c}_1 = \text{mean}(\boldsymbol{p}_1, \boldsymbol{p}_2) = (2.5, 3)$, $\boldsymbol{c}_2 = (10, 14)$

Example 1: Iteration 1

$d(\boldsymbol{p}_3, \boldsymbol{c}_1) = \mathbf{6.10}$, $d(\boldsymbol{p}_3, \boldsymbol{c}_2) = 7.21$

Move $\boldsymbol{p}_3$ to C_1: $C_1 = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3\}$, $C_2 = \{\boldsymbol{p}_5\}$,

Centroids updated

$\boldsymbol{c}_1 = \text{mean}(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3) = (3.67, 4.67)$, $\boldsymbol{c}_2 = (10, 14)$

$d(\boldsymbol{p}_4, \boldsymbol{c}_1) = 6.29$, $d(\boldsymbol{p}_4, \boldsymbol{c}_2) = \mathbf{5}$

Move $\boldsymbol{p}_4$ to C_2: $C_1 = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3\}$, $C_2 = \{\boldsymbol{p}_4, \boldsymbol{p}_5\}$,

Centroids updated

$\boldsymbol{c}_1 = \text{mean}(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3) = (3.67, 4.67)$, $\boldsymbol{c}_2 = \text{mean}(\boldsymbol{p}_4, \boldsymbol{p}_5) = (8.5, 12)$

Example 1: Iteration 1

$d(\boldsymbol{p_5}, \boldsymbol{c_1}) = 11.28$, $d(\boldsymbol{p_5}, \boldsymbol{c_2}) =$ **2.5**

No data point movement

$C_1 = \{\boldsymbol{p_1}, \boldsymbol{p_2}, \boldsymbol{p_3}\}$, $C_2 = \{\boldsymbol{p_4}, \boldsymbol{p_5}\}$,

No centroids update

$\boldsymbol{c_1} = \text{mean}(\boldsymbol{p_1}, \boldsymbol{p_2}, \boldsymbol{p_3}) = (3.67, 4.67)$, $\boldsymbol{c_2} = \text{mean}(\boldsymbol{p_4}, \boldsymbol{p_5}) = (8.5, 12)$

Example 1: Iteration 1

$d(\boldsymbol{p}_6, \boldsymbol{c}_1) = 7.54$, $d(\boldsymbol{p}_6, \boldsymbol{c}_2) =$ **2.06**
Move $\boldsymbol{p}_6$ to C_2: $C_1 = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3\}$, $C_2 = \{\boldsymbol{p}_4, \boldsymbol{p}_5, \boldsymbol{p}_6\}$,
Centroids updated: $\boldsymbol{c}_1 = \text{mean}(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3) = (3.67, 4.67)$, $\boldsymbol{c}_2 = \text{mean}(\boldsymbol{p}_4, \boldsymbol{p}_5, \boldsymbol{p}_6) = (8.67, 11.33)$

$d(\boldsymbol{p}_7, \boldsymbol{c}_1) = 5.47$, $d(\boldsymbol{p}_7, \boldsymbol{c}_2) =$ **1.37**
Move $\boldsymbol{p}_7$ to C_2: $C_1 = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3\}$, $C_2 = \{\boldsymbol{p}_4, \boldsymbol{p}_5, \boldsymbol{p}_6, \boldsymbol{p}_7\}$,
Centroids updated: $\boldsymbol{c}_1 = \text{mean}(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3) = (3.67, 4.67)$, $\boldsymbol{c}_2 = \text{mean}(\boldsymbol{p}_4, \boldsymbol{p}_5, \boldsymbol{p}_6, \boldsymbol{p}_7) = (8.25, 10.75)$

Example 1: Iteration 2

$C_1 = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3\}$, $C_2 = \{\boldsymbol{p}_4, \boldsymbol{p}_5, \boldsymbol{p}_6, \boldsymbol{p}_7\}$,
$\boldsymbol{c}_1 = \text{mean}(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3) = (3.67, 4.67)$,
$\boldsymbol{c}_2 = \text{mean}(\boldsymbol{p}_4, \boldsymbol{p}_5, \boldsymbol{p}_6, \boldsymbol{p}_7) = (8.25, 10.75)$

Similarly, compute $d(\boldsymbol{p}_1, \boldsymbol{c}_1)$, $d(\boldsymbol{p}_1, \boldsymbol{c}_2)$, $d(\boldsymbol{p}_2, \boldsymbol{c}_1)$, $d(\boldsymbol{p}_2, \boldsymbol{c}_2)$, // no change

Example 1: Iteration 2

$d(\boldsymbol{p}_3, \boldsymbol{c}_1) = 4.07$, $d(\boldsymbol{p}_3, \boldsymbol{c}_2) =$ **3.55**
Move $\boldsymbol{p}_3$ to C_2: $C_1 = \{\boldsymbol{p}_1, \boldsymbol{p}_2\}$, $C_2 = \{\boldsymbol{p}_3, \boldsymbol{p}_4, \boldsymbol{p}_5, \boldsymbol{p}_6, \boldsymbol{p}_7\}$
Centroids updated: $\boldsymbol{c}_1 = \text{mean}(\boldsymbol{p}_1, \boldsymbol{p}_2) = (2.5, 3)$, $\boldsymbol{c}_2 = \text{mean}(\boldsymbol{p}_3, \boldsymbol{p}_4, \boldsymbol{p}_5, \boldsymbol{p}_6, \boldsymbol{p}_7) = (7.8, 10.2)$

$d(\boldsymbol{p}_4, \boldsymbol{c}_1)$, $d(\boldsymbol{p}_4, \boldsymbol{c}_2)$, $d(\boldsymbol{p}_5, \boldsymbol{c}_1)$, $d(\boldsymbol{p}_5, \boldsymbol{c}_2)$, $d(\boldsymbol{p}_6, \boldsymbol{c}_1)$, $d(\boldsymbol{p}_6, \boldsymbol{c}_2)$, $d(\boldsymbol{p}_7, \boldsymbol{c}_1)$, $d(\boldsymbol{p}_7, \boldsymbol{c}_2)$ // no change

Example 1: Iteration 3

$C_1 = \{\boldsymbol{p_1}, \boldsymbol{p_2}\}$, $C_2 = \{\boldsymbol{p_3}, \boldsymbol{p_4}, \boldsymbol{p_5}, \boldsymbol{p_6}, \boldsymbol{p_7}\}$

$\boldsymbol{c_1}$ = mean($\boldsymbol{p_1}, \boldsymbol{p_2}$) = (2.5, 3), $\boldsymbol{c_2}$ = mean($\boldsymbol{p_3}, \boldsymbol{p_4}, \boldsymbol{p_5}, \boldsymbol{p_6}, \boldsymbol{p_7}$) = (7.8, 10.2)

- No data point movement and centroids update so STOP