

[Example K-Means Clustering]

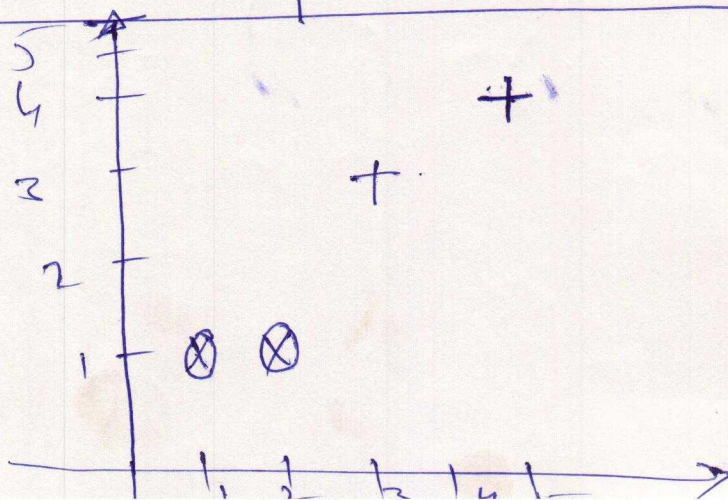
①

Suppose we have 4 medicines as our training data point objects and each medicine has 2 attributes. Each attribute represents coordinate of the object. We have to determine which medicine belongs to cluster 1 and which medicine belongs to the other cluster.

Object	Weight index: (x) (attribute 1)	pH: (y) (attribute 2)
Medicine A	1	1
Medicine B	2	1
Medicine C	3	3
Medicine D	5	4

$$C_1: (1, 1)$$

$$C_2: (2, 1)$$



Iteration 1 -

Again calculate the distance of all objects ② to the new centroids

$$D' = \begin{bmatrix} 0 & 1 & 3.61 & 5 \\ 3.14 & 2.36 & 0.47 & 1.89 \end{bmatrix}$$

$$C_1 = (1, 1)$$

$$C_2 = \left(\frac{2+4+5}{3}, \frac{1+3+4}{3} \right) = \left(\frac{11}{3}, \frac{8}{3} \right)$$

$$\begin{bmatrix} \text{A} & \text{B} & \text{C} & \text{D} \\ 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix} \begin{matrix} X \\ Y \end{matrix}$$

Again assign the Medicine objects to appropriate group based on minimum distance.

$$G' = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{matrix} \text{Group-1} \\ \text{Group-2} \end{matrix}$$

Determine the new Centroid

$$C_1 = \left(\frac{1+2}{2}, \frac{1+1}{2} \right) = \left(\frac{3}{2}, 1 \right)$$

$$C_2 = \left(\frac{4+5}{2}, \frac{3+4}{2} \right) = \left(\frac{9}{2}, \frac{7}{2} \right)$$

$$\begin{matrix} & A & B & C & D \\ \begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 1 & 3 & 4 \end{bmatrix} & x & & & \\ & & & & y \end{matrix}$$

③

Step 1: We calculate the distance between cluster centroid to each object. Let us use Euclidean distance.

$$d(x, y) = \sum_{i=1}^p |x_i - y_i|$$

$$D^0 = \begin{bmatrix} 0 & 1 & 3.61 & 5 \\ 1 & 0 & 2.83 & 4.24 \end{bmatrix} \quad \begin{matrix} c1 = (1, 1) \text{ Group 1} \\ c2 = (2, 1) \text{ Group 2} \end{matrix}$$

for example for Medicine C,

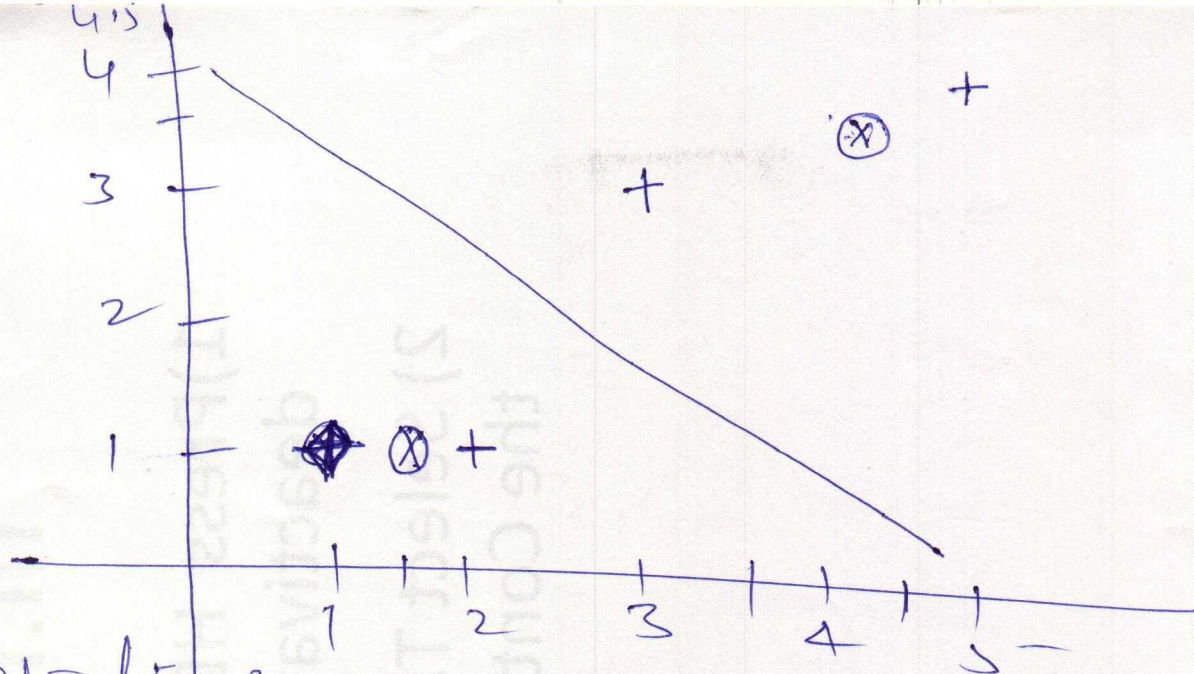
$$c1 = (1, 1) = \sqrt{(4-1)^2 + (3-1)^2} = 3.61$$

$$c2 = (2, 1) = \sqrt{(4-2)^2 + (3-1)^2} = 2.83$$

Step 2: We assign each object based on the minimum distance.

$$G^0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \quad \begin{matrix} \text{Group 1} \\ \text{Group 2} \end{matrix}$$

④



Iteration-2.

Again calculate the distance

$$D^2 = \begin{bmatrix} 0.5 & 0.5 & 3.20 & 4.61 \\ 4.30 & 3.54 & 0.71 & 0.71 \end{bmatrix} \begin{matrix} C1 = (3/2, 1) \\ C2 = (9/2, 7/2) \end{matrix}$$

	A	B	C	D	
[1	2	4	5	X
	1	1	3	4	Y

Based on minimum distance again assign the object to appropriate cluster

$$G^2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{matrix} \text{Group 1} \\ \text{Group 2} \end{matrix}$$

We obtained $G^2 = G^1$ Convergence point.