

K-mean clustering: calculation example

id	x_1	x_2
d1	2	1
d2	-3	8
d3	0	10
d4	3	2
d5	-2	8
d6	3	0
d7	4	0
d8	-2	6
d9	-3	9
d10	6	1

Step1: Choose the number of clusters K.

K = ...2.....

Step2: Select K random datapoints from the data as centroids.

centroid1 = ...(-2, 6).....

centroid2 = ...(3, 2).....

Step3 (ครั้งที่1): Assign all the datapoints to the closet cluster centroid.

distance(d1, centroid1) = ...6.40.....

distance(d1, centroid2) = ...1.41.....

distance(d2, centroid1) = ...2.23.....

distance(d2, centroid2) = ...8.48.....

distance(d3, centroid1) = 4.47

distance(d3, centroid2) = 8.54

distance(d4, centroid1) = 6.40

distance(d4, centroid2) = 0

distance(d5, centroid1) = 2

distance(d5, centroid2) = 7.81

distance(d6, centroid1) = 7.81

distance(d6, centroid2) = 2

distance(d7, centroid1) = 8.48

distance(d7, centroid2) = 2.23

distance(d8, centroid1) = 0

distance(d8, centroid2) = 6.40

distance(d9, centroid1) = 3.16

distance(d9, centroid2) = 9.21

distance(d10, centroid1) = 9.43

distance(d10, centroid2) = 3.16

$$2+3+3+4+6 = \frac{18}{5} = 3.6$$

$$1+2+0+0+1 = \frac{4}{5} = 0.8$$

Step4 (ครั้งที่1): Recompute the centroids of newly form cluster.

Centroid1 = average of $(-3, 8), (0, 10), (-2, 8), (-2, 6), (-3, 9)$
 = $(-2.0, 8.2)$

Centroid2 = average of $(2, 1), (3, 2), (3, 0), (4, 0), (6, 1)$
 = $(3.6, 0.8)$

Step5: Repeat step 3 and 4 until meet stopping criteria

Step3 (ครั้งที่2): Assign all the datapoints to the closet cluster centroid.

distance(d1, centroid1) = 1.61

distance(d1, centroid2) = 8.23

distance(d2, centroid1) = 9.76

distance(d2, centroid2) = 1.01

distance(d3, centroid1) = 9.87

distance(d3, centroid2) = 2.69

distance(d4, centroid1) = 1.34

distance(d4, centroid2) = 7.69

distance(d5, centroid1) = 9.12

distance(d5, centroid2) = 0.2

distance(d6, centroid1) = 1

distance(d6, centroid2) = 9.60

$$\text{distance}(d7, \text{centroid1}) = 0.89$$

$$\text{distance}(d7, \text{centroid2}) = 10.16$$

$$\text{distance}(d8, \text{centroid1}) = 7.64$$

$$\text{distance}(d8, \text{centroid2}) = 2.2$$

$$\text{distance}(d9, \text{centroid1}) = 10.52$$

$$\text{distance}(d9, \text{centroid2}) = 1.28$$

$$\text{distance}(d10, \text{centroid1}) = 2.40$$

$$\text{distance}(d10, \text{centroid2}) = 10.76$$

Step5: Repeat step 3 and 4 until meet stopping criteria

Step4 (ครั้งที่2): Recompute the centroids of newly form cluster.

$$\begin{aligned} \text{Centroid1} &= \text{average of } (-3, 8), (0, 10), (-2, 8), (-2, 6), (-3, 9) \\ &= (-2.0, 8.2) \end{aligned}$$

$$\begin{aligned} \text{Centroid2} &= \text{average of } (3, 1), (3, 2), (3, 0), (4, 0), (6, 1) \\ &= (3.6, 0.8) \end{aligned}$$

⋮