

## k-mean clustering

Algorithm:

example

Apply the k-mean clustering in the following data set for 2 clusters.

Sample No	X	Y
→ 1	185	72
→ 2	170	56
✓ 3	168	60
✓ 4	179	68
✓ 5	182	72
✓ 6	188	77

$$K = 2$$

initial centroid.

cluster	X	Y
K1	185	72
K2	170	56

Step 1: calculate the ~~Euclidean~~ distance between two cluster

$$\text{distance} = (x_1, y_1), (a, b) = \sqrt{(x-a)^2 + (y-b)^2}$$

distance from cluster 1  $\Rightarrow \sqrt{(185-185)^2 + (72-72)^2} = 0$   
 Initialise the distance from cluster 1 to its self  $\Rightarrow 0$

similarly distance from cluster 2  $\Rightarrow$

$$\sqrt{(185-170)^2 + (72-56)^2} = \sqrt{225 + 256} = \sqrt{481} = 22.06$$

similarly distance from

$$\text{cluster 1} : \sqrt{(185-170)^2 + (72-56)^2}$$

$$185, 72 = \cancel{22.60} \quad 22.60$$

$$\text{cluster 2} : \sqrt{(170-170)^2 + (56-56)^2}$$

$$170, 56 = 0$$

cluster	Centroid		Assignment
	X	Y	
K1	0	22.65	1
K2	22.65	0	2
K3			
K4			
K5			
K6			

next data (168, 60)

distance from cluster 1 :  $\sqrt{(185-168)^2 + (72-60)^2}$

$$= 20.80$$

distance from K2  $\Rightarrow \sqrt{(170-168)^2 + (56-60)^2}$

$$= 4.47$$

so this date is close to cluster 2  
 now we need to re calculate  
 the cluster centroid.

near Centroid

Page 3

Cluster	X	Y
K1	185	72
K2	$\frac{170+168}{2} = 169$	$56+68/2 = 58$

next data point  
(179, 68)

$$\text{distance from cluster 1} \Rightarrow \sqrt{(179-185)^2 + (68-72)^2} = 7.2$$

$$\text{distance from cluster 2} \Rightarrow \sqrt{(179-169)^2 + (68-58)^2} = 14.14$$

\* the data point is close to  $x_1$  so, it belongs to  $x_1$  now need to recalculate the centroid.

cluster	X	Y
$x_1$	$\frac{185+179}{2} \Rightarrow 182$	$\frac{72+68}{2} \Rightarrow 70$
$x_2$	169	58

next data point's

next data

Point  $(182, 72)$

$$\text{distance from cluster 1} \Rightarrow \sqrt{(182 - 182)^2 + (72 - 70)^2}$$

$$= \sqrt{4}$$

$$= 2$$

$$\text{distance from cluster 2} \Rightarrow \sqrt{(182 - 169)^2 + (72 - 58)^2}$$

$$= \sqrt{365}$$

$$= 19.10$$

\* so this data point belongs to

cluster 1 now calculate the centroid

Cluster	X	Y
K1	$\frac{182+182}{2} = 182$	$\frac{70+72}{2} = 71$
K2	169	58

next data

point  $(188, 77)$

$$\text{distance from } K_1 \Rightarrow \sqrt{(188 - 182)^2 + (77 - 71)^2}$$

$$= 8.48$$

$$\text{distance from } K_2 = \sqrt{(188 - 169)^2 + (77 - 58)^2}$$

$$= 26.87$$

belongs to  $K_1$  is the new centroid is



cluster	X	Y
K1	$\frac{188+182}{2}$ = 185	$\frac{58+77}{2}$ = 74
K2	169	58

this is the final Centroid of the data set of cluster 1 & cluster 2

### Final tabulation

dataset no	X	Y	cluster
1	185	72	K1
2	170	56	K2
3	168	60	K2
4	179	68	K1
5	182	72	K1
6	188	77	K1

