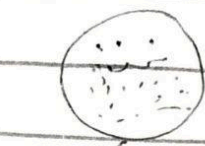


Abdul Raffy

W25-PAK-INP-AI-16

K-Mean clustering

Make clusters of data with 2 clusters
similarities.



cluster
of circle.

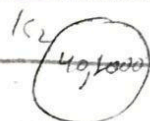
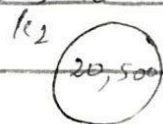


cluster
of cross

Sr-no	age	Amount
C1	20	500
C2	40	1000
C3	30	800
C4	18	300
C5	28	200
C6	35	1400
C7	45	1800

Steps

1) How many clusters to make
lets take two clusters.



2) Centroid calculation

which value from dataset will
lie in which cluster.

(Randomly assign datasets to clusters)

→ lets associate C1 with k1 and
C2 with k2. (This is initialization)

→ These value are called centroid
initialization, and the remaining values
will be observed values.

lets see → it is calculated on basis of distance nearest to
(k1 or k2)

$$C_3 = d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \rightarrow \text{Euclidean Distance}$$

Applying to C_3

For k_1

$$C_3 = \sqrt{(30-20)^2 + (800-500)^2}$$

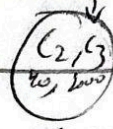
$$= \sqrt{(10)^2 + (300)^2} = 300.5$$

For k_2

$$C_3 = \sqrt{(30-40)^2 + (800-1000)^2}$$

$$= 200.0$$

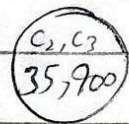
↳ 200 is smallest value and it is for k_2 so, C_3 will go to k_2 cluster.



update to new Mean
by calculating C_2

C_2 <u>Mean</u>	C_3	C_2
$= 40 + 30$		$\frac{800+1000}{2}$
$= 35$		$= 900$

So updated cluster k_2 is now.



→ new centroid value

k_1

Now, if we calculate $C_4 = \sqrt{(18-20)^2 + (300-500)^2}$

$$= 200.0$$

According to updated k_2 cluster

↳ $C_4 = \sqrt{(18-35)^2 + (200-900)^2}$

→ after calculation it will become centroid value for k_1 and update it by calculated Mean of C_2 and C_4

↓
as was used