K-means clustering (centroid-based technique)

Algorithm

Input:

K: no of clusters

D: Data set containing n objects.

Output:

A = set of K clusters

Method

1) Arbitarily choose k objects from D as the initial cluster centers
2) repeal
3) (re) assign
   (re) assign each object to cluster to which the object is most similar, based on mean values of objects in cluster.
4) update the cluster means i.e calculate the mean value of objects for each cluster.
5) Until no change

Ex 1) {2, 4, 10, 12, 3, 20, 30, 11, 25}

Assume K = 2

{2, 3, 4, 10, 11, 12, 20, 25, 30}

Assume means

$m_1 = 2$ $\quad$ $m_2 = 4$

$K_1 = \{2, 3\}$ $\quad$ $K_2 = \{4, 10, 12, 20, 30, 11, 25\}$

$m_1 = \frac{2+3}{2} = 2.5$ $\quad$ $m_2 = \frac{4+10+12+20+30+11+25}{7} = 16$

$K_1 = \{2, 3, 4\}$ $\quad$ $K_2 = \{10, 11, 12, 20, 30, 25\}$

$m_1 = 3$ $\quad$ $m_2 = 18$

$K_1 = \{2, 3, 4, 10\}$ $\quad$ $K_2 = \{11, 12, 20, 30, 25\}$

$m_1 = 4.75$ $\quad$ $m_2 = 19.6$

$K_1 = \{2, 3, 4, 10, 11, 12\}$ $\quad$ $K_2 = \{20, 30, 25\}$

$m_1 = 7$ $\quad$ $m_2 = 25$

$K_1 = \{2, 3, 4, 10, 11, 12\}$ $\quad$ $K_2 = \{20, 25, 30\}$

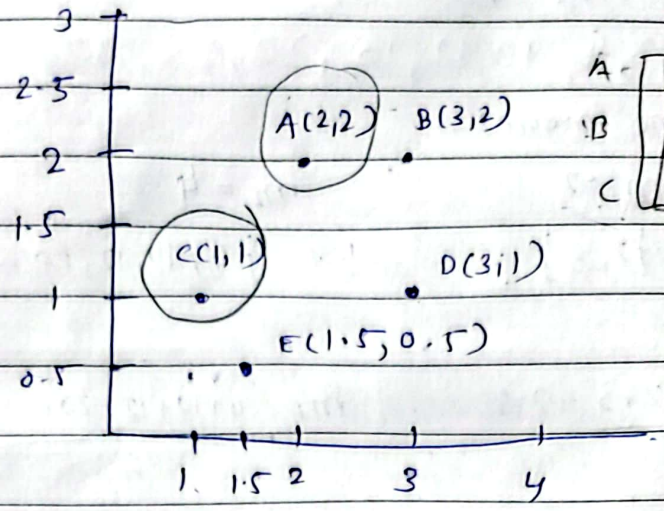
The clusters are

$K_1 = \{2, 3, 4, 10, 11, 12\}$

$K_2 = \{20, 25, 30\}$

create

Ex 2) Use K-means algorithm to ~~check~~ two clusters



	A	B	C
A	0	1	2
B	1	0	3
C	2	3	0

Find adjacency matrise using Euclidean distance

$d(A,A) = 0$

$d(A,B) = \sqrt{(2-3)^2+(2-2)^2} = \sqrt{(-1)^2} = 1$

$d(A,C) = \sqrt{(2-1)^2+(2-1)^2} = \sqrt{2} = 1.41$

$d(A,D) = \sqrt{(2-3)^2+(2-1)^2} = \sqrt{2} = 1.41$

$d(A,E) = \sqrt{(2-1.5)^2+(2-0.5)^2} = 1.58$

$D(B,B) = 0$

$d(B,C) = \sqrt{(3-1)^2+(2-1)^2} = \sqrt{5} = 2.24$

$d(B,D) = \sqrt{(3-3)^2+(2-1)^2} = 1$

$d(B,E) = \sqrt{(3-1.5)^2+(2-0.5)^2} = 2.12$

$d(C,C) = 0$

$d(C,D) = \sqrt{(1-3)^2+(1-1)^2} = 2$

$d(C,E) = \sqrt{(1-1.5)^2+(1-0.5)^2} = 0.71$

$d(D,D) = 0$

$d(D,E) = \sqrt{(3-1.5)^2+(1-0.5)^2} = 1.58$

$d(E,E) = 0$

Let A and C be the centroids A(2,2) & C(1,1)

$d(A,A) = 0$
$d(A,B) = 1$
$d(A,C) = 1.41$
$d(A,D) = 1.41$
$d(A,E) = 1.58$

} group 1

$d(C,A) = 1.41$
$d(C,B) = 2.24$
$d(C,C) = 0$
$d(C,D) = 2$
$d(C,E) = 0.71$

} group 2

Assign each object based on the minimum distance. Thus,

A	B	C	D	E	centroid
0	1	1.41	1.41	1.58	A(2,2)
1.41	2.24	0	2	0.71	C(1,1)

The group Matrix G^0 =

G^0 =	A	B	C	D	E	centroid
G_1	1	1	0	1	0	A(2,2)
G_2	0	0	1	0	1	C(1,1)

group 1 → A(2,2), B(3,2), D(3,1)
group 2 → C(1,1), E(1.5,0.5)

Now, the centroid c_1 is the average of the coordinates of these three members of group 1

$$c1 = \left(\frac{2+3+3}{3}, \frac{2+2+1}{3}\right) = (2.67, 1.67)$$

$$c2 = \left(\frac{1+1.5}{2}, \frac{1+0.5}{2}\right) = (1.25, 0.75)$$

c1 (2.67, 1.67)

$d(c_1, A) = \sqrt{(2.67-2)^2 + ((1.67)-2)^2} = 0.75$

$d(c_1, B) = \sqrt{(2.67-3)^2 + (1.67-2)^2} = 0.47$

$d(c1, C) = \sqrt{(2.67-1)^2 + (1.67-1)^2} = 1.079$

$d(c1, D) = \sqrt{(2.67-3)^2 + (1.67-1)^2} = 0.75$

$d(c1, E) = \sqrt{(2.67-1.5)^2 + (1.67-0.5)^2} = 1.65$

c2 (1.25, 0.75)

$d(c2, A) = \sqrt{(1.25-2)^2 + (0.75-2)^2} = 1.45$

$d(c2, B) = \sqrt{(1.25-3)^2 + (0.75-2)^2} = 2.15$

$d(c2, C) = \sqrt{(1.25-1)^2 + (0.75-1)^2} = 0.35$

$d(c2, D) = \sqrt{(1.25-3)^2 + (0.75-1)^2} = 1.76$

$d(c2, E) = \sqrt{(1.25-1.5)^2 + (0.75-0.5)^2} = 0.35$

A	B	C	D	E	centroid
0.75	0.47	1.079	0.75	1.65	c1(2.67, 1.67)
1.45	2.15	0.35	1.76	0.35	c2(1.25, 0.75)

The group matrix G1 =

G' =

A	B	C	D	E	centroid
1	1	0	1	0	c1(2.67, 1.67)
0	0	1	0	1	c2(1.25, 0.75)

$$G_1' = G_1^0$$

Final clusters are

group 1 = { A(2,2), B(3,2), D(3,1) }

group 2 = { C(1,1), E(1.5, 0.5) }