17-Nov-2023

Fuzzy C-Means Clustering

clustering

- Hard clustering Algorithm
  - data point assigned only a single cluster example
  - K-Means., k-Medoid
- Soft clustering Algorithm
  - each data point belongs to a cluster with a certain probability also known as membership value example.
  - Fuzzy C-Means algorithm

Fuzzy C-Means steps:

Step-1: Given data points based on the number of clusters required initialize then membership table with random values

Suppose the given data points are

$\{(1,3), (2,5), (6,8), (7,9)\}$

cluster	(1,3)	(2,5)	(4,8)	(7,8)
	D_1	D_2	D_3	D_4
1	0.8	0.7	0.2	0.1
2	0.2	0.3	0.8	0.9

Step-2: Find out the centroid

the formula for finding out the centroid (v) is

$$V_{ij} = \frac{\sum_{K=1}^{n} \gamma_{ik}^{m} \cdot x_{ik}}{\sum_{K=1}^{n} \gamma_{ik}^{m}}$$

γ = Fuzzy membership value

m = Fuziness parameter generally taken as 2

x_k = is the data point.

$$V_{11} = \frac{(0.8^2 \times 1 + 0.7^2 \times 2 + 0.2^2 \times 4 + 0.1^2 \times 7)}{(0.8^2 + 0.7^2 + 0.2^2 + 0.1^2)} = 1.568$$

$$V_{12} = \frac{(0.8^2 \times 3 + 0.7^2 \times 5 + 0.2^2 \times 8 + 0.1^2 \times 9)}{(0.8^2 + 0.7^2 + 0.2^2 + 0.1^2)} = 4.051$$

$$V_{21} = \frac{0.2^2 \times 1 + 0.3^2 \times 2 + 0.8^2 \times 4 + 0.9^2 \times 7}{0.2^2 + 0.3^2 + 0.8^2 + 0.9^2} = 5.35$$

$$V_{22} = \frac{0.2^2 \times 3 + 0.3^2 \times 5 + 0.8^2 \times 8 + 0.9^2 \times 9}{0.2^2 + 0.3^2 + 0.8^2 + 0.9^2} = 8.215$$

Hence centroids are (1.568, 4.051) and (5.35, 8.215)

Step-3: Find out the distance of each point from the centroid.

$$D_{11} = \sqrt{(1-1.568)^2 + (3-4.051)^2} = 1.2$$

$D_{12} = 6.79$ $\quad D_{31} = 4.63$

$D_{21} = 1.04$ $\quad D_{32} = 1.36$

$D_{22} = 1.64$ $\quad D_{41} = 7.34$

$D_{42} = 1.82$

cluster	$D_1(1,3)$	$D_2(2,5)$	$D_3(4,8)$	$D_4(7,9)$
1st	0.8	0.7	0.2	0.1
2nd	0.2	0.3	0.8	0.9
cluster→	1st	1st	2nd	2nd

Step: 4. Updating the membership values:

$$\gamma_{ki} = \left(\sum_{j=1}^{n} \left\{ \frac{d_{ki}^2}{d_{kj}^2} \right\}^{\left(\frac{1}{m-1}\right)} \right)^{-1}$$

for point 1 new membership values are:

$$\gamma_{11} = \left(\left\{ \frac{(1.2)^2}{(1.2)^2} + \frac{(1.2)^2}{(6.79)^2} \right\}^{\left(\frac{1}{2-1}\right)} \right)^{-1} = 0.97$$

$$\gamma_{12} = \left(\left\{ \frac{(6.79)^2}{(1.2)^2} + \frac{(6.79)^2}{(6.79)^2} \right\}^{\left(\frac{1}{2-1}\right)} \right)^{-1} = 0.03$$

$\gamma_{21} = 0.95$ $\gamma_{31} = 0.08$ $\gamma_{41} = 0.06$

$\gamma_{22} = 0.05$ $\gamma_{32} = 0.92$ $\gamma_{42} = 0.94$

Hence updated membership value

cluster	(1,3)	(2,5)	(4,8)	(7,9)
1	0.97	0.95	0.08	0.06
2	0.03	0.05	0.92	0.94

step-5: Repeat the step-2 to step-4 until the constant values are obtained for the membership values or the difference is less than the tolerence value

$$\epsilon = 0.01$$