Fuzzy c-Mean

15 November 2024   18:47

Hard clustering? K-means. Each data point = single cluster.

↓

FUZZY C-MEANS? Each data point can belong to multiple clusters (probability / likelihood)

- c
  - hyperparam
  - no. of clusters
- m
  - fuzziness param
  - $[1.25, 2]$
- μ
  - fuzzy membership value
  - $[0, 1]$
- ϵ
  - hyperparameter
  - threshold criteria

General flow:

1. Define hyperparameters (c, ϵ)
2. Random init of membership matrix
3. Calculate centroid of each cluster
4. Find distance of each sample from each centroid
5. Update membership value

repeat until membership value $< \epsilon$

PROS

- better than k-means for overlapped dataset (datapoint belongs to ≥ 1 cluster)
- doesn't assume spherical cluster like k-means does (can handle diff shapes)
- less sensitive to noise compared to k-means

CONS

- Apriori specification of no. of clusters (like kmeans)
- performance depends on initial cluster select^n (–〃–)
- low m ⇒ better convergence w/ large no. of iterations

Math:

$$V_{ij} = \frac{\sum_{k=1}^{n} \gamma_{ik}^{m} \times x_k}{\sum_{k=1}^{n} \gamma_{ik}^{m}}$$

centroid

i^{th} cluster

k^{th} coordinate of sample x.

n = no. of datapoints

γ = fuzzy membership value

m = fuzziness parameter (usually 2)

x_k = data point.

Example:

Step 1

Given: clusters, datapoints.

To do: randomly initialize membership table w/ probabilities

cluster	(1,3)	(2,5)	(4,8)	(7,9)
1	0.8	0.7	0.2	0.1
2	0.2	0.3	0.8	0.9
	1.0	1.0	1.0	1.0

→ (7,9) belongs to cluster 1 w/ a 0.1 probability

→ (7,9) belongs to cluster 2 w/ a 0.9 probability

Step 2

Find centroid using $V_{ij} = \sum \gamma_{ij}^{m} \cdot x_j$

prob of 1st cluster, 1st point

m=2 usually

1st datapoint

1st cluster

$$V_{11} = \frac{\gamma_{11}^{2} \cdot x_1^{(1)} + \gamma_{12}^{2} \cdot x_1^{(2)} + \gamma_{13}^{2} x_1^{(3)} + \gamma_{14}^{2} x_1^{(4)}}{\gamma_{11}^{2} + \gamma_{12}^{2} + \gamma_{13}^{2} + \gamma_{14}^{2}}$$

1st cluster

x coord.

→ prob. of datapt in cluster^2 × x coordinate of each data point

$$= \frac{(0.8)^2(1) + (0.7)^2(2) + (0.2)^2(4) + (0.1)^2(7)}{\ldots}$$

Find centroid using
$$V_{ij} = \frac{\sum \gamma_{ij}^{m} \cdot x_j}{\sum \gamma_{ij}^{m}}$$

cluster x coord.

x coordinate of each data point

$$\gamma_{11}^2 + \gamma_{12}^2 + \gamma_{13}^2 + \gamma_{14}^2$$

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

$$= 1.568$$

$$V_{12} = \frac{\gamma_{11}^2 \cdot x_1^{(1)} + \gamma_{12}^2 \cdot x_1^{(2)} + \gamma_{13}^2 \cdot x_1^{(3)} + \gamma_{14}^2 \cdot x_1^{(4)}}{\gamma_{11}^2 + \gamma_{12}^2 + \gamma_{13}^2 + \gamma_{14}^2}$$

1st cluster, y coord

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

$$= 4.051$$

centroid coords

$\therefore$ for cluster 1 : $\underbrace{(1.568, 4.051)}$

probability of 1st data pt in cluster 2

1st data point cluster 2

$$V_{21} = \frac{\gamma_{21}^2 \cdot x_2^{(1)} + \gamma_{22}^2 x_2^{(2)} + \gamma_{23}^2 x_2^{(3)} + \gamma_{24}^2 x_2^{(4)}}{\gamma_{21}^2 + \gamma_{22}^2 + \gamma_{23}^2 + \gamma_{24}^2}$$

cluster 2, x coord

$$= \frac{(0.2)^2(1) + (0.3)^2(2) + (0.8)^2(4) + (0.9)^2(7)}{0.2^2 + 0.3^2 + 0.8^2 + 0.9^2}$$

$$= 5.35$$

$$V_{22} = \frac{\gamma_{21}^2 x_2^{(1)} + \gamma_{22}^2 x_2^{(2)} + \gamma_{23}^2 x_2^{(3)} + \gamma_{24}^2 x_2^{(4)}}{\gamma_{21}^2 + \gamma_{22}^2 + \gamma_{23}^2 + \gamma_{24}^2}$$

cluster 2, y coord

$$= \frac{(0.2)^2(3) + (0.3)^2(5) + (0.8)^2(8) + (0.9)^2(9)}{0.2^2 + 0.3^2 + 0.8^2 + 0.9^2}$$

$$= 8.215$$

centroids of cluster 2

$\therefore$ for cluster 2 : $\underbrace{(5.35, 8.215)}$

Step 3

Find euclidean distance of each point from cluster centroids.

Cluster 1: (1.568, 4.051)   Cluster 2: (5.35, 8.25)

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

i = point

j = cluster

Point 1: (1,3)

$D_{11} = \sqrt{(1-1.568)^2 + (3-4.051)^2} = \boxed{1.2}$ → minimum distance → assign to cluster 1.

$D_{12} = \sqrt{(1-5.35)^2 + (3-8.215)^2} = 6.79$

Point 2: (2,5)

$D_{21} = \sqrt{(2-1.568)^2 + (5-4.051)^2} = \boxed{1.04}$ → minimum distance → assign to cluster 1.

$D_{22} = \sqrt{(2-5.35)^2 + (5-8.215)^2} = 4.64$

Point 3 : (4,8)

$D_{31} = \sqrt{(4-1.568)^2 + (8-4.051)^2} = 4.63$

Point 3 : (4,8)

$$D_{31} = \sqrt{(4-1.568)^2 + (8-4.051)^2} = 4.63$$

$$D_{32} = \sqrt{(4-5.35)^2 + (8-8.215)^2} = \boxed{1.36} \rightarrow \text{minimum distance}$$

→ assign to cluster 2

Point 4: (7,9)

$$D_{41} = \sqrt{(7-1.568)^2 + (9-4.051)^2} = 7.43$$

$$D_{42} = \sqrt{(7-5.35)^2 + (9-8.215)^2} = \boxed{1.82} \rightarrow \text{minimum distance}$$

→ assign to cluster 2

Step 4

Update membership table with formula:

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

Point 1: $(D_{11}, D_{12}) = (1.2, 6.79)$

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

datapt 1 · cluster 1 · dist d_{11} (1st cluster, 1st coord) · dist d_{12} (1st cluster, 2nd coord).

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

datapt 1 · cluster 2

Point 2: $(D_{21}, D_{22}) = (1.04, 4.64)$

$$\gamma_{21} = \left[\left\{\frac{(1.04)^2}{(1.04)^2} + \frac{(1.04)^2}{(4.64)^2}\right\}^{\frac{1}{2-1}}\right]^{-1} = 0.95$$

$$\gamma_{22} = \left[\left\{\frac{(4.64)^2}{(1.04)^2} + \frac{(4.64)^2}{(4.64)^2}\right\}^{\frac{1}{2-1}}\right]^{-1} = 0.05$$

Point 3: $(D_{31}, D_{32}) = (4.63, 1.36)$

$$\gamma_{31} = \left[\left\{\frac{(4.63)^2}{(4.63)^2} + \frac{(4.63)^2}{(1.36)^2}\right\}^{\frac{1}{2-1}}\right]^{-1} = 0.08$$

$$\gamma_{32} = \left[\left\{\frac{(1.36)^2}{(4.63)^2} + \frac{(1.36)^2}{(1.36)^2}\right\}^{\frac{1}{2-1}}\right]^{-1} = 0.92$$

Point 4: $(D_{41}, D_{42}) = (7.34, 1.82)$

$$\gamma_{41} = \left[\left\{\frac{(7.34)^2}{(7.34)^2} + \frac{(7.34)^2}{(1.82)^2}\right\}^{\frac{1}{2-1}}\right]^{-1} = 0.06$$

$$\gamma_{42} = \left[\left\{\frac{(1.82)^2}{(7.34)^2} + \frac{(7.34)^2}{(7.34)^2}\right\}^{\frac{1}{2-1}}\right]^{-1} = 0.94$$

UPDATE TABLE

cluster	(1,3)	(2,5)	(4,8)	(7,9)
1	0.97 ~~0.8~~	0.95 ~~0.7~~	0.08 ~~0.2~~	0.06 ~~0.1~~

1	0.97 ~~0.8~~	0.95 ~~0.7~~	0.08 ~~0.2~~	0.06 ~~0.1~~
2	0.03 ~~0.2~~	0.05 ~~0.3~~	0.92 ~~0.8~~	0.94 ~~0.9~~
	1.0	1.0	1.0	1.0

Step 5

Repeat until membership value < ϵ