**FUZZY C-MEANS ALGORITHM**

1. **What is Fuzzy c-means clustering?**

Fuzzy clustering (also referred to as **soft clustering** or **soft k-means**) is a form of clustering in which each data point can belong to more than one cluster.

Clustering or cluster analysis involves assigning data points to clusters such that items in the same cluster are as similar as possible, while items belonging to different clusters are as dissimilar as possible. Clusters are identified via similarity measures. These similarity measures include distance, connectivity, and intensity. Different similarity measures may be chosen based

on the data or the application.

1. **How to perform**

**Step 1**: Initialize the data points into desired number of clusters randomly.

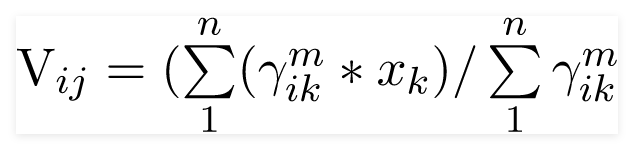
Assume there are 2 clusters in which the data is to be divided, initializing the data point randomly. Each data point lies in both the clusters with some membership value which can be assumed anything in the initial state.

The table below represents the values of the data points along with their membership (gamma) in each of the cluster.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

**Step 2**: Find out the centroid.

The formula for finding out the centroid (V) is:



Where, µ is fuzzy membership value of the data point, m is the fuzziness parameter (generally taken as 2), and xk is the data point.

V11 = (0.82 \*1 + 0.72 \* 2 + 0.22 \* 4 + 0.12 \* 7) / (0.82 + 0.72 + 0.22 + 0.12) = 1.568

V12 = (0.82 \*3 + 0.72 \* 5 + 0.22 \* 8 + 0.12 \* 9) / (0.82 + 0.72 + 0.22 + 0.12) = 4.051

V11 = (0.22 \*1 + 0.32 \* 2 + 0.82 \* 4 + 0.92 \* 7) / (0.22 + 0.32 + 0.82 + 0.92) = 5.35

V11 = (0.22 \*3 + 0.32 \* 5 + 0.82 \* 8 + 0.92 \* 9) / (0.22 + 0.32 + 0.82 + 0.92) = 8.215

Centroids are: **(1.568, 4.051)** and **(5.35, 8.215)**

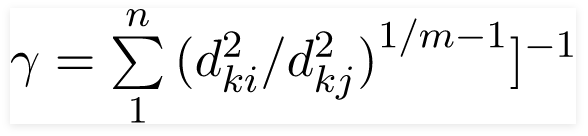
**Step 3:** Find out the distance of each point from centroid.

D11 = ((1 - 1.568)2 + (3 - 4.051)2)0.5 = 1.2

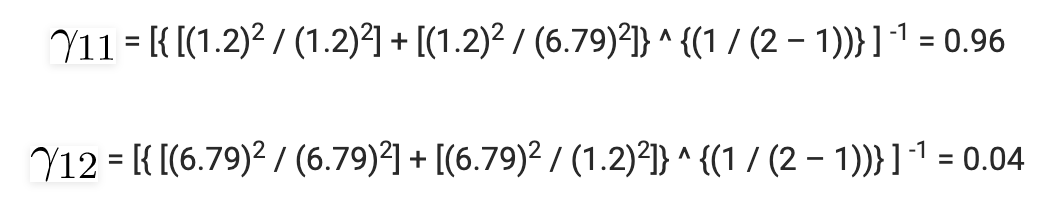
D12 = ((1 - 5.35)2 + (3 - 8.215)2)0.5 = 6.79

Similarly, the distance of all other points is computed from both the centroids.

**Step 4**: Updating membership values.



For point 1 new membership values are:



Alternatively,



Similarly, compute all other membership values, and update the matrix.

**Step 5**: Repeat the steps 2-4 until the constant values are obtained for the membership values or the difference is less than the tolerance value (a small value up to which the difference in values of two consequent updations is accepted).

**Step 6**: Defuzzify the obtained membership values.