1. Algorythm description

Fuzzy C Means algorythm is a clustering method, an extension of K Means, that allows one data point to belong to two or more clusters based on a membership function. This clustering method is frequently used in pattern recognition.

In a hard clusterig algorythm, the membership function the membership function takes the shape

and in a fuzzy clustering method we might have a membership function similar to this

The goal of the FCM algorythm is to minimize the objective function

where wij is the degree of membership for the datapoint xi to the centroid cj and m>=1 is a hyperparameter for the clustering which determines the fuzzyness of the clusters.

The centroids are computed using :

and the membership degrees :

In order to start we need to choose a number of clusters and randomly assign data points to this clusters.

For each iteration we compute the centroids for each cluster, for each data point we compute the membership degree for each cluster based on these new centroids.

We repeat the iterations until the maximum change in membership degrees between two iterations is less then the sensitivity threshold ε

2. Importance and practical applications

Clustering is an important classification method. FCM, being an unsupervised machine learning algorythm with an acceptable order of complexity can be easily used for clustering large datasets with good results. Due to the nature of fuzzy logic that deals with approximate and non-precise logic, FCM is more suited for coping with the nature of reality then hard clustering methods.

Main implemetations include

Clustering in medical diagnostic systems

market segmentation

image segmentation

divide chemical compounds into toxicity classes

patient classification

Web log analisys

3. Pros and Cons

Some of the main advantages of FCM are

Always converges

Unsupervised

A more natural approach to clustering

Relative simple implementation

Cons

Complexity is O(ndc²i) vs K-Means O(ndci)

Sensitivity to initial guess

Sensitivity to noise / outliers

One issue encountered in mult-dimensional datasets is the need to normalize the scales, since unbalanced scales might lead the shorter scaled dimension to be ignored in the clustering process.

4. Datasets used

a) Weightlifting

[] Dunn 1973

[]Bezdek 1981

https://www.win.tue.nl/~mbottoto/files/papers/conference\_papers/fuzzy\_citi2017.pdf