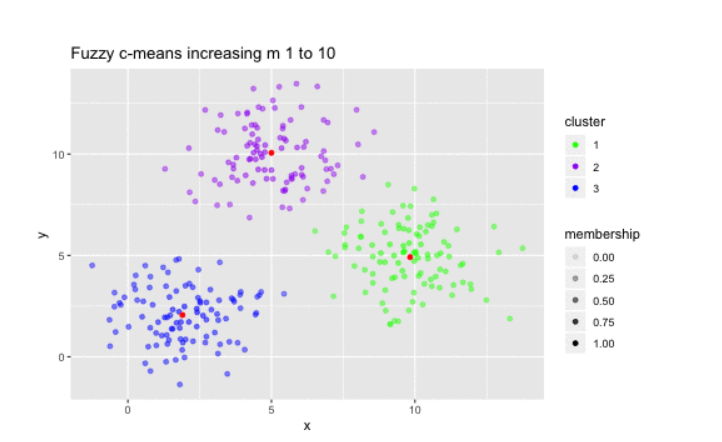
**Fuzzy C-Means Clustering**

Fuzzy C-Means (FCM) is a clustering algorithm used to group similar data points based on their similarity with each other. It is an extension of the K-means clustering algorithm, which assigns a data point to only one cluster. FCM, on the other hand, allows a data point to belong to multiple clusters with different degrees of membership.

Fuzzy logic principles can be used to cluster multidimensional data, assigning each point a membership in each cluster center from 0 to 100 percent. This can be very powerful compared to traditional hard-threshold clustering where every point is assigned a crisp, exact label. This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one.

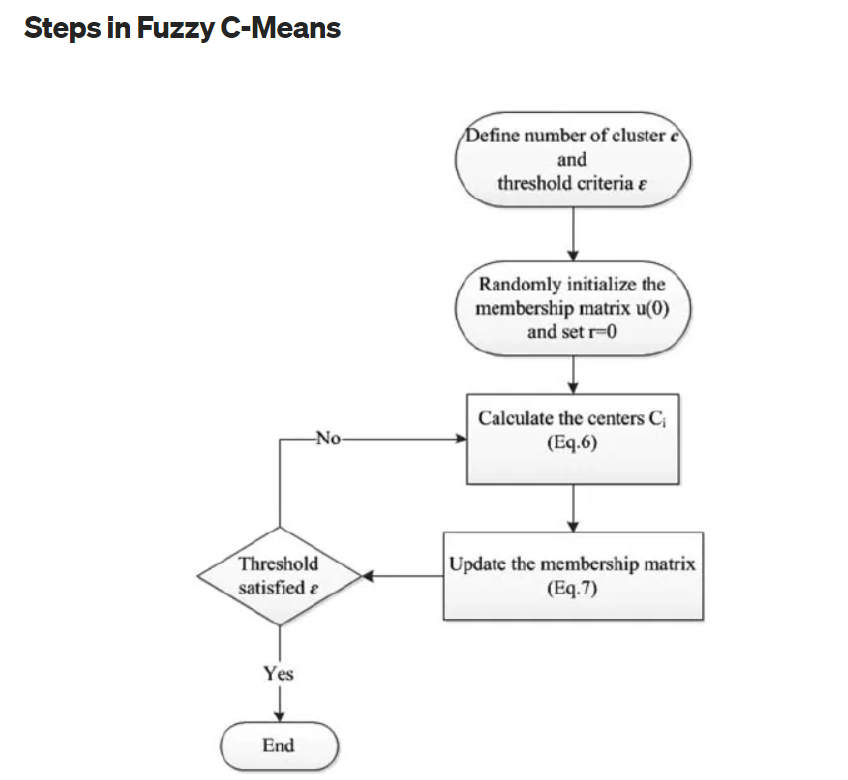
It is an unsupervised clustering algorithm that permits us to build a fuzzy partition from data. The algorithm depends on a parameter m which corresponds to the degree of fuzziness of the solution. Large values of m will blur the classes and all elements tend to belong to all clusters. The solutions of the optimization problem depend on the parameter m. That is, different selections of m will typically lead to different partitions. Given below is a gif that shows the effect of the selection of m obtained from the fuzzy c-means.



et us compare these two powerful algorithms to get a clear idea of where the fuzzy c-means algorithm fits in.

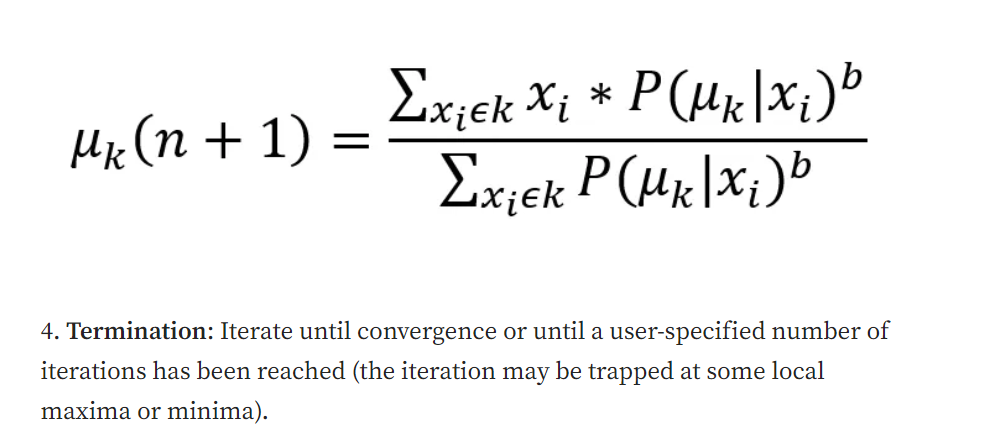
1. **Attribution to a cluster**: In fuzzy clustering, each point has a probability of belonging to each cluster, rather than completely belonging to just one cluster as it is the case in the traditional k-means. In Fuzzy-C Means clustering, each point has a weighting associated with a particular cluster, so a point doesn’t sit “in a cluster” as much as has a weak or strong association to the cluster, which is determined by the inverse distance to the center of the cluster.
2. **Speed**: Fuzzy-C means will tend to run slower than K means, since it’s actually doing more work. Each point is evaluated with each cluster, and more operations are involved in each evaluation. K-Means just needs to do a distance calculation, whereas fuzzy c means needs to do a full inverse-distance weighting.
3. **Personal Opinion**: FCM/Soft-K-Means is “less stupid” than Hard-K-Means when it comes to elongated clusters (when points otherwise consistent in other dimensions tend to scatter along a particular dimension or two).

We should realize that fuzzy c-means is a special case of K-means when the probability function used is simply 1 if the data point is closest to a centroid and 0 otherwise.



The process flow of fuzzy c-means is enumerated below:

1. **Assume** a fixed number of clusters k.
2. **Initialization**: Randomly initialize the k-means μk associated with the clusters and compute the probability that each data point xi is a member of a given cluster k, P(point xi has label k|xi, k).
3. **Iteration**: Recalculate the centroid of the cluster as the weighted centroid given the probabilities of membership of all data points xi:



**Unlike k-means where data point must exclusively belong to one cluster center here data point is assigned membership to each cluster center as a result of which data point may belong to more than one cluster center.**

**Fuzzy C-Means clustering is a powerful unsupervised machine learning technique that can be used to group data points with similar characteristics. It is particularly useful when the boundaries between clusters are not well-defined, or when a data point could belong to multiple clusters with different degrees of membership.**