HITESH

FUZZY C MEANS CLUSTERING

HITESH
A023119820027

HITESH

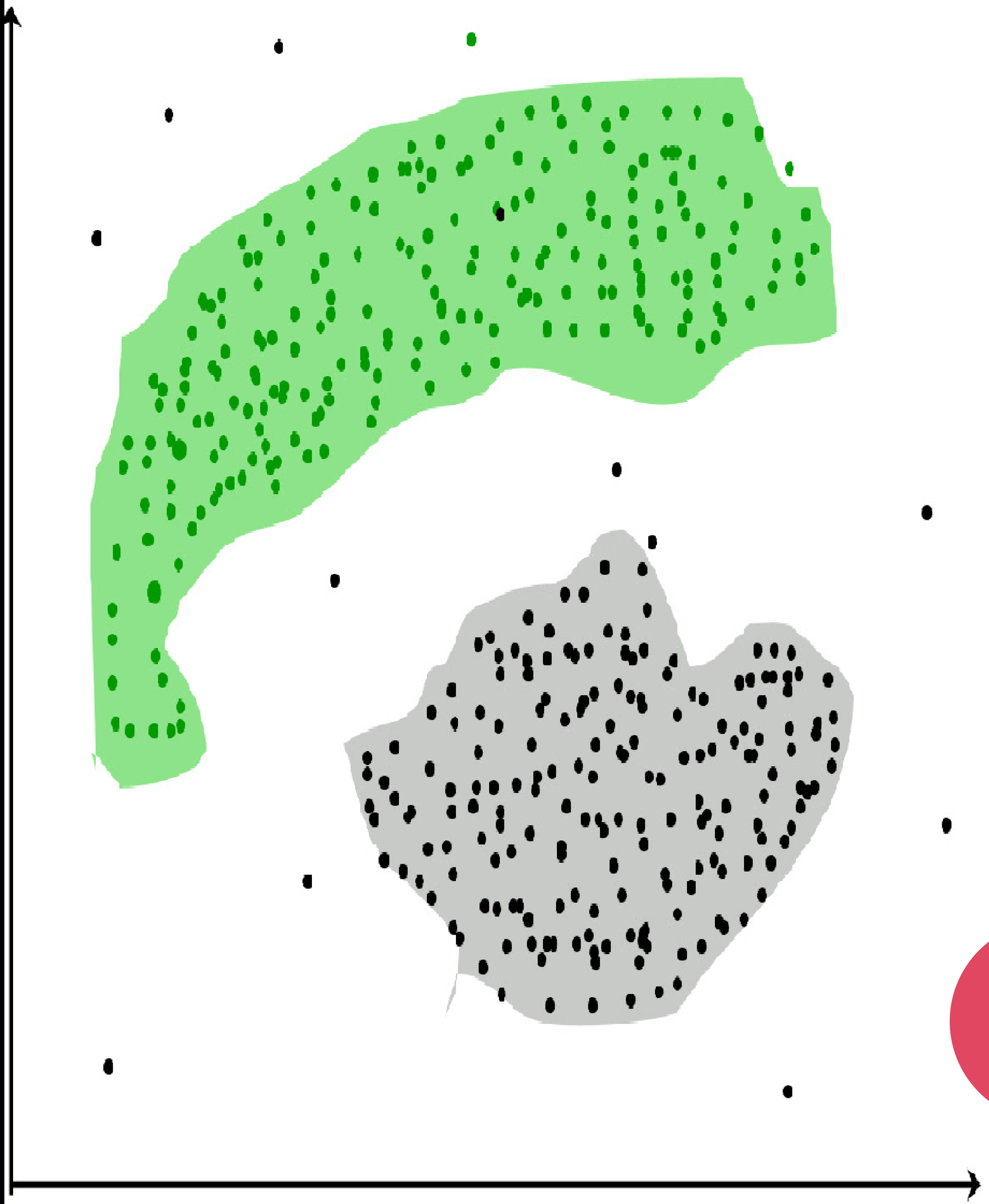
CLUSTERING

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups.

In simple words, the aim is to segregate groups with similar traits and assign them into clusters

Types of Clustering -

1. Soft Clustering
2. Hard Clustering



HITESH

K-MEANS AND FUZZY C-MEANS CLUSTERING

01

Fuzzy based Unsupervised Clustering Technique

02

An extension of K-means

03

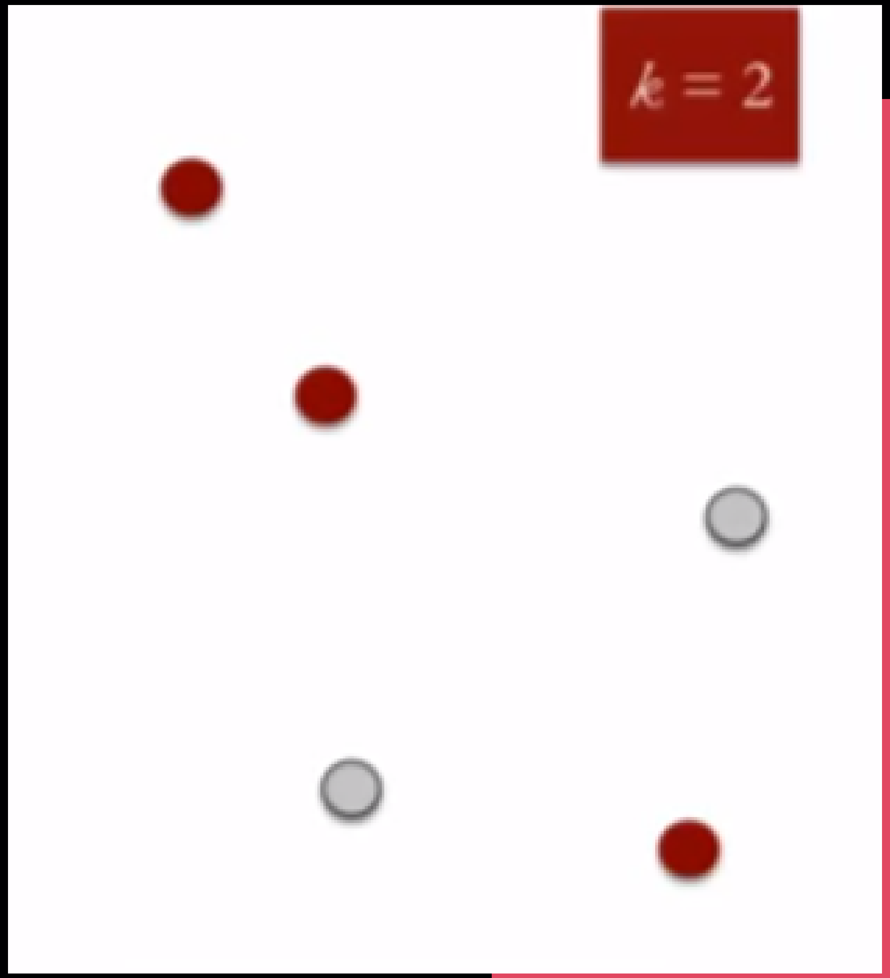
K-means generates partitions

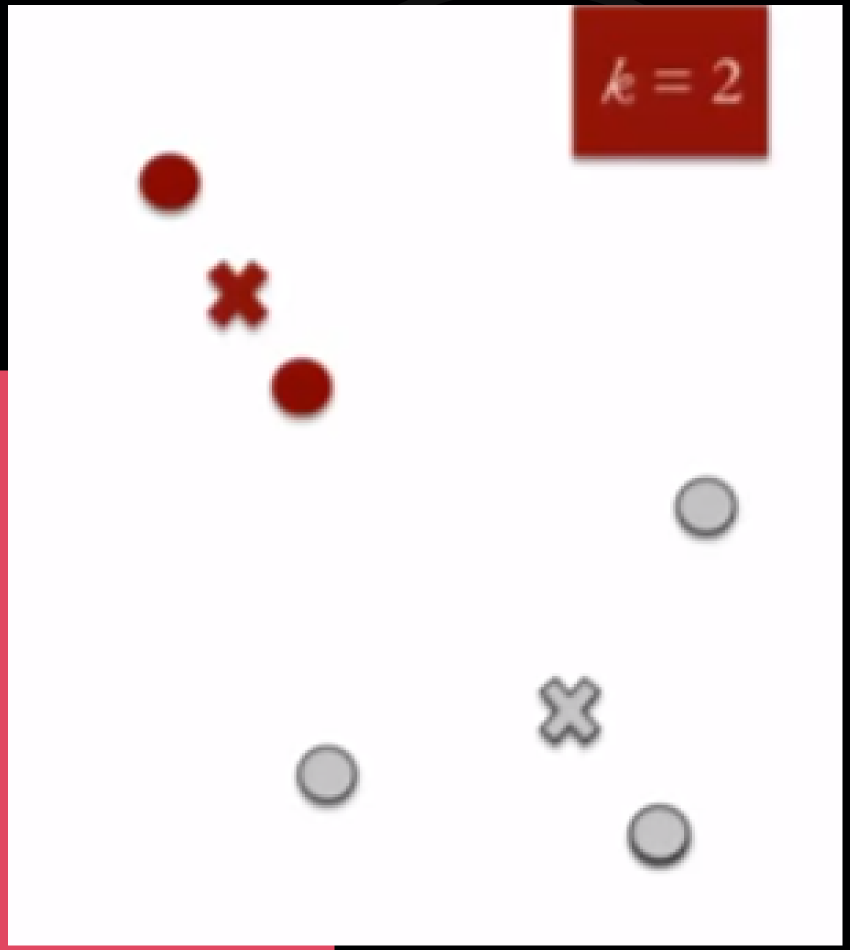
Each data point can only be assigned in one cluster

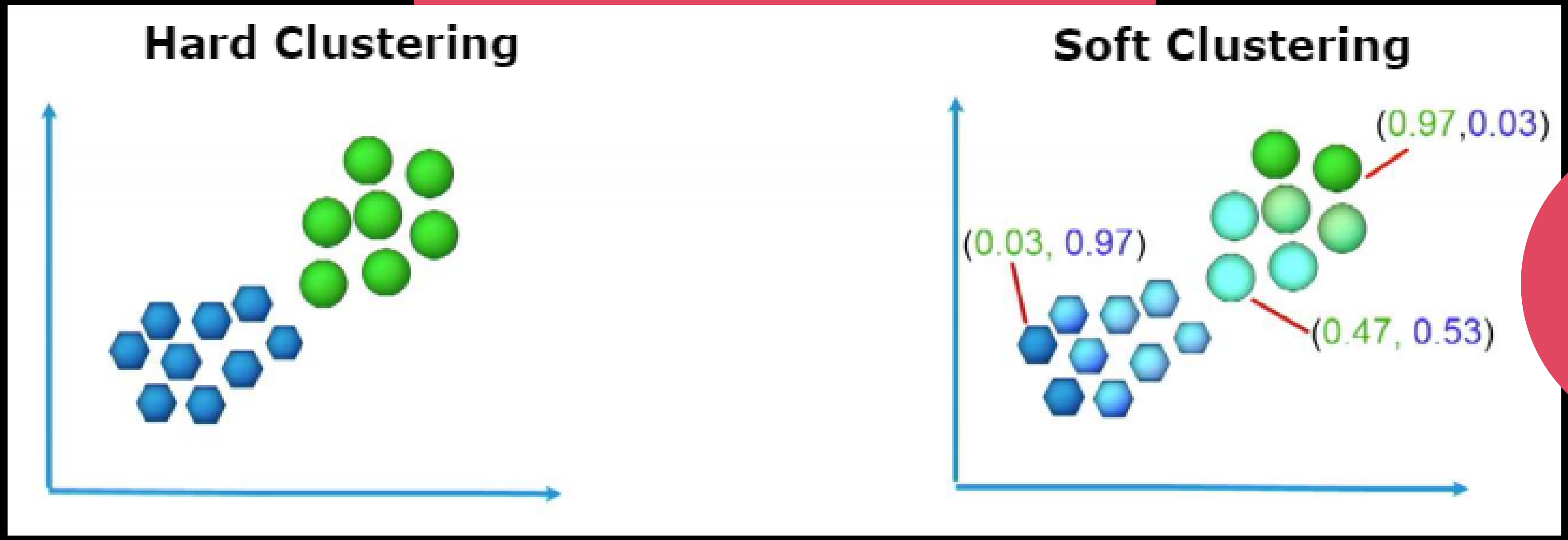
04

Fuzzy C means allows data points to be assigned into more than one cluster

Each data point has a degree of membership (or probability) of belonging to each cluster







HITESH

STEPS IN ALGORITHM

01 **Randomly initialize the membership matrix using this equation -**

$$\sum_{j=1}^{C} \mu_j(x_i) = 1 \qquad i = 1,2 \ldots k$$

02 **Calculate the centroid using equation**

$$C_j = \frac{\sum_i [\mu_j(x_i)]^m x_i}{\sum_i [\mu_j(x_i)]^m}$$

03 **Calculate dissimilarity between the data points and centroid using Euclidean distance**

$$D_i = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

04 **Update the new membershiop matrix using the equation**

$$\mu_j(x_i) = \frac{\left[\frac{1}{d_{ji}^2}\right]^{1/m-1}}{\sum_{k=1}^{C} \left[\frac{1}{d_{ki}^2}\right]^{1/m-1}}$$

here m is a fuzzification parameter m ϵ [1.25, 2]

05 **Repeat the steps(2-4) until the constant values are obtained for the membership values or the difference is less than the tolerance value ϵ**

$$\left|\mu_j(x_i)^{(k+1)} - \mu_j(x_i)^{(k)}\right| < \epsilon$$

HITESH

EXAMPLE

STEP 1

Suppose the given data points are {(1, 3), (2, 5), (6, 8), (7, 9)}

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

HITESH

EXAMPLE

STEP 2

Find out the centroid

$$C_1 = \frac{(0.82*1 + 0.72*2 + 0.22*4 + 0.12*7)}{(0.82 + 0.72 + 0.22 + 0.12)} = 1.568$$

$$C_2 = \frac{(0.82*3 + 0.72*5 + 0.22*8 + 0.12*9)}{(0.82 + 0.72 + 0.22 + 0.12)} = 4.051$$

$$C_3 = \frac{(0.22*1 + 0.32*2 + 0.82*4 + 0.92*7)}{(0.22 + 0.32 + 0.82 + 0.92)} = 5.35$$

$$C_4 = \frac{(0.22*3 + 0.32*5 + 0.82*8 + 0.92*9)}{((0.22 + 0.32 + 0.82 + 0.92)} = 8.215$$

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

HITESH

EXAMPLE

STEP 3

Find out the distance of each point from centroid

$$D11 = ((1 - 1.568)^2 + (3 - 4.051)^2)^{0.5} = 1.2$$

$$D12 = ((1 - 5.35)^2 + (3 - 8.215)^2)^{0.5} = 6.79$$

HITESH

EXAMPLE

$$\mu_{11} = \frac{\left(\frac{1}{(1.2)^2}\right)^{\frac{1}{2-1}}}{\left(\frac{1}{(1.2)^2} + \frac{1}{(6.79)^2}\right)^{\frac{1}{2-1}}} = 9.6$$

$$\mu_{12} = \frac{\left(\frac{1}{(6.79)^2}\right)^{\frac{1}{2-1}}}{\left(\frac{1}{(1.2)^2} + \frac{1}{(6.79)^2}\right)^{\frac{1}{2-1}}} = 0.04$$

STEP 4

Updating membership values

STEP 5

Repeat the steps(2-4) until the constant values are obtained for the membership values

HITESH

PROS AND CONS

FCM Advantages

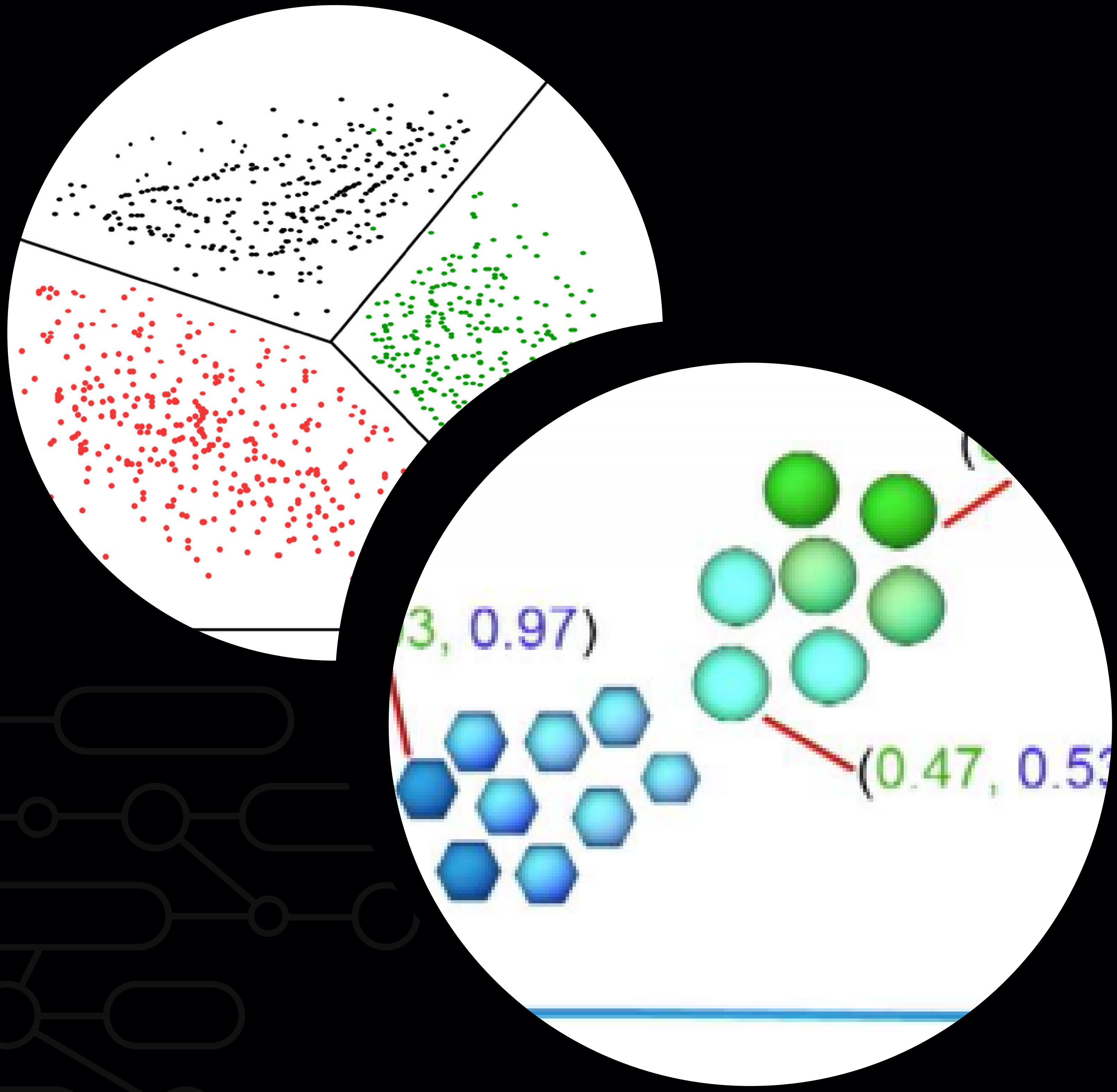
- Gives best result for overlapped dataset and comparatively better then k-means algorithm
- Unlike k-means where data point must exclusively belong to one cluster center here data point is assigned membership to each cluster center as a result of which data point may belong to more than one cluster center

FCM Disadvantages

- A prior specification of the number of clusters
- With lower value of the ϵ, we get better result but at the expense of more iterations
- Euclidean distance measures can unequally weight underlying factors

HITESH

APPLICATIONS OF CLUSTERING



- Marketing: It can be used to characterize & discover customer segments for marketing purposes.
- Biology: It can be used for classification among different species of plants and animals.
- Libraries: It is used in clustering different books on the basis of topics and information.
- Insurance: It is used to acknowledge the customers, their policies and identifying the frauds.
- City Planning: It is used to make groups of houses and to study their values based on their geographical locations and other factors present.

HITESH

THANK YOU