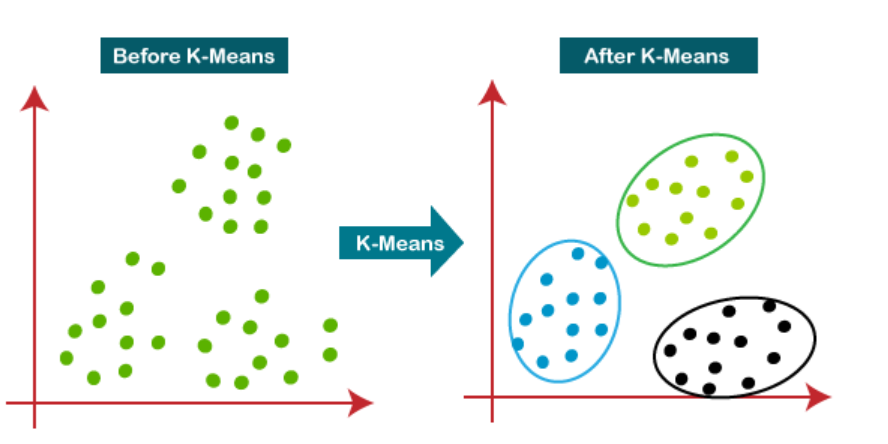
**Unsupervised learning:**

**kmeans**



les etapes de l’algorithme kmeans:

The working of the K-Means algorithm is explained in the below steps:

**Step-1:** Select the number K to decide the number of clusters.

**Step-2:** Select random K points or centroids. (It can be other from the input dataset).

**Step-3:** Assign each data point to their closest centroid, which will form the predefined K clusters.

**Step-4:** Calculate the variance and place a new centroid of each cluster.

**Step-5:** Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.

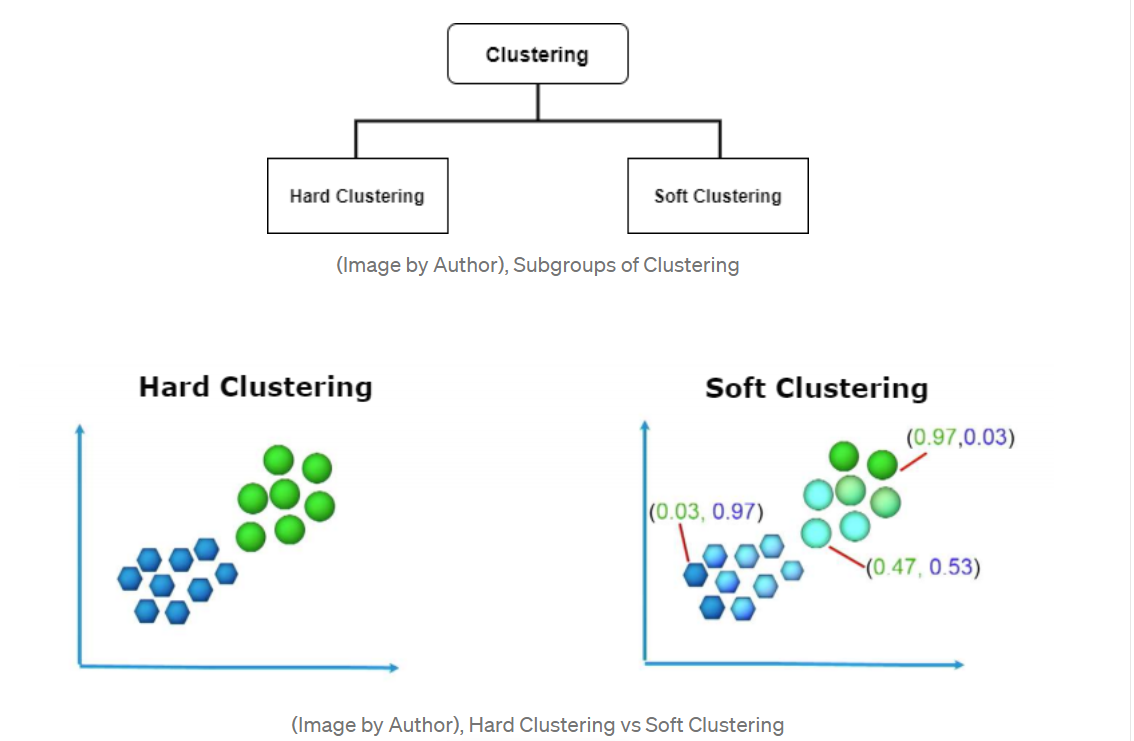
**Step-6:** If any reassignment occurs, then go to step-4 else go to FINISH.

**Step-7**: The model is ready.

**Elbow method:**

WCSS(Within-Cluster Sum of Square):

WCSS is defined as **the sum of the squared distance between each member of the cluster and its centroid**.

**----------------------------------------------------------------**

**hard clustering :**K-Means.

**Soft Clustering: Fuzzy C-means** clustering (FCM) Algorithm.

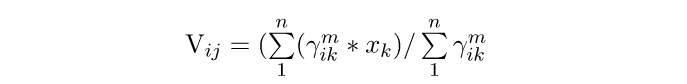
# Fuzzy C-Means Clustering:

Fuzzy C-Means clustering is a soft clustering approach, where each data point is assigned a likelihood or probability score to belong to that cluster. The step-wise approach of the Fuzzy c-means clustering algorithm is:

* Fix the value of **c** (number of clusters), and select a value of **m**(generally 1.25<m<2), and initialize partition matrix U.

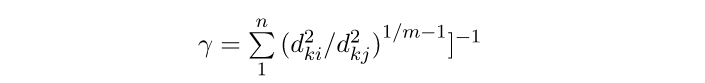
https://miro.medium.com/max/555/1*5pBPTQmdRWZPdVDmh5arGQ.png

* Calculate cluster centers (centroid).



Here,  
**µ: Fuzzy membership value  
m: fuzziness parameter**

* Update Partition Matrix



* Repeat the above steps until convergence.

-----------------------------------

Fuzzy clustering algorithms seeks to minimize cluster memberships and distances, but we will focus on Fuzzy C-Means Clustering algorithm. Fuzzy c-means developed in 1973 and improved in 1981. It’s very similar to k-means algorithm in a structure way:

* Choose number of clusters.
* Assign coefficients randomly to each data point for being in the clusters.
* Repeat until algorithm converged (Objective Function C minimizes cluster memberships and distances):
  + Compute the centroid for each cluster
  + Compute each data points’ coefficients of being in the clusters.

The main difference with k-means cluster is that objective function for fuzzy c-means algorithm allows different cluster membership with probability values, where k-means cluster has strict objective function allows only one cluster membership.