Unsupervised Learning

What is Unsupervised Learning?

- **Unsupervised learning is a type of machine learning where the algorithm learns patterns and structures from unlabeled data. In contrast to supervised learning, there are no explicit labels (targets) provided to the model. The goal is to infer the underlying structure of the data.**

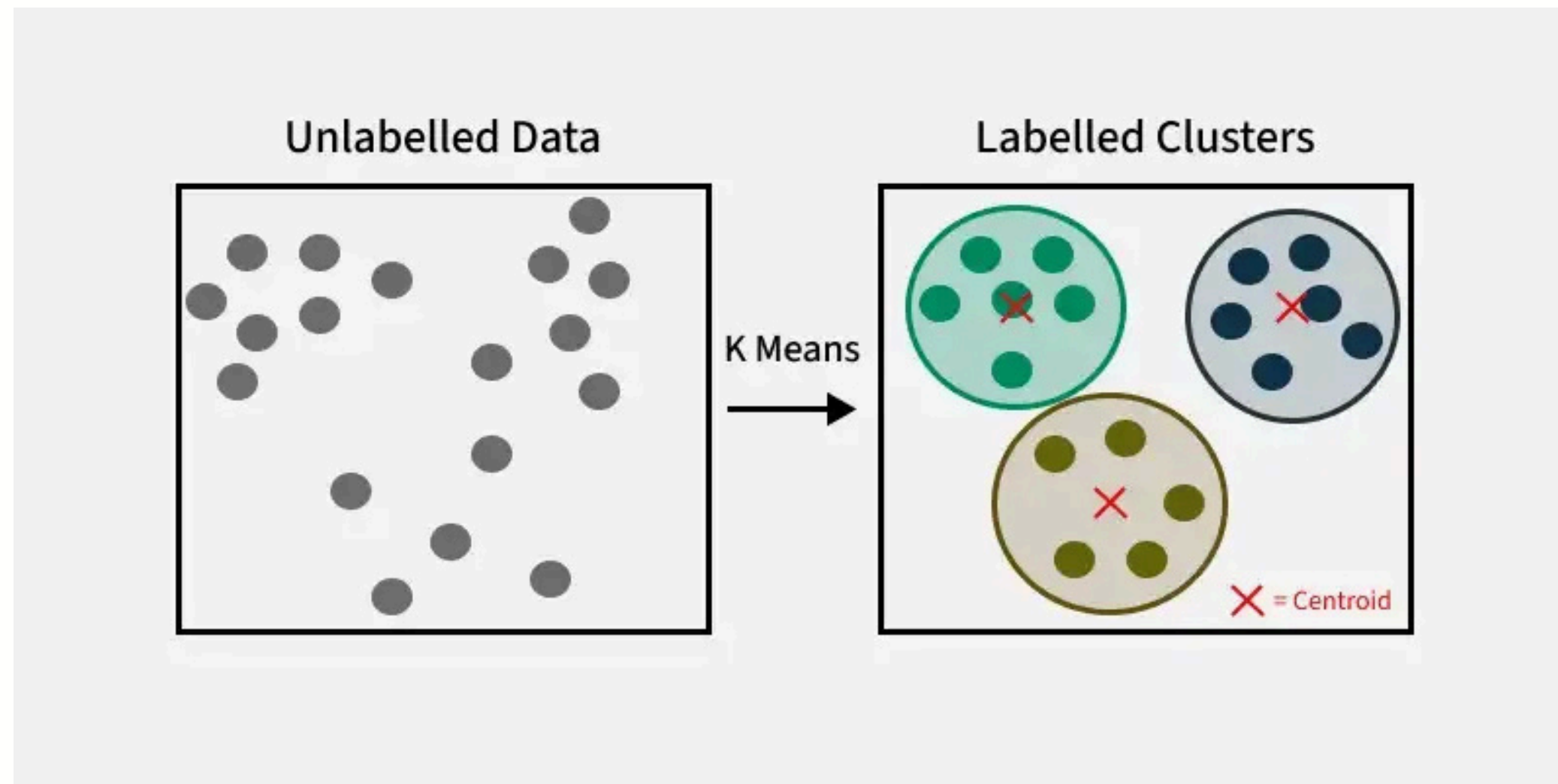# Unsupervised Learning Algorithms

Clustering

Association Rule Learning
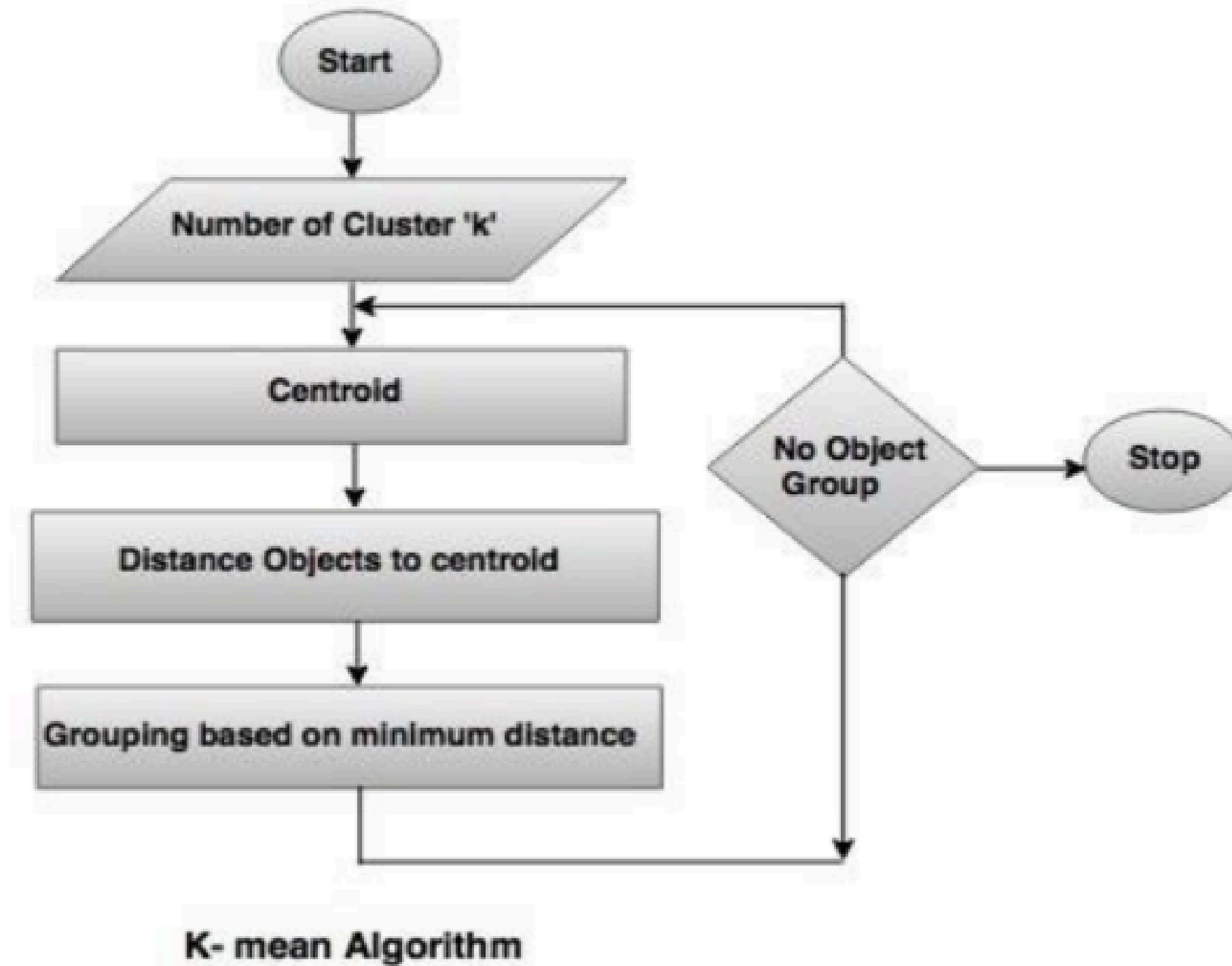
Dimensionality Reduction

Understanding K-means Clustering

- **We are given a data set of items with certain features and values for these features like a vector. The task is to categorize those items into groups.**
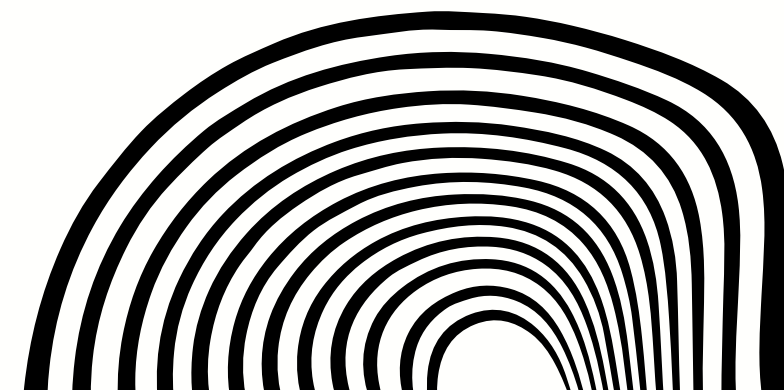
How k-means clustering works?

K- mean Algorithm

Euclidean Distance:

- For two points in 2D space:
- If
- Point 1 = (x₁, y₁)
- Point 2 = (x₂, y₂)
- Then the Euclidean distance between them is:

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

Elbow Method in K-Means Clustering

- **The Elbow Method helps us find this optimal k value. Here's how it works:**
- **We iterate over a range of k values, typically from 1 to n (where n is a hyper-parameter you choose).**
- **For each k, we calculate the Within-Cluster Sum of Squares (WCSS).**

WCSS measures **how well the data points are clustered around their respective centroids**. It is defined as the **sum of the squared distances between each point and its cluster centroid**:

$$\text{WCSS} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \text{distance}(x_j^{(i)}, c_i)^2$$

where,

$\text{distance}(x_j^{(i)}, c_i)$ represents the distance between the j-th data point $x_j^{(i)}$ in cluster i and the centroid c_i of that cluster.

Elbow Method in K-Means Clustering