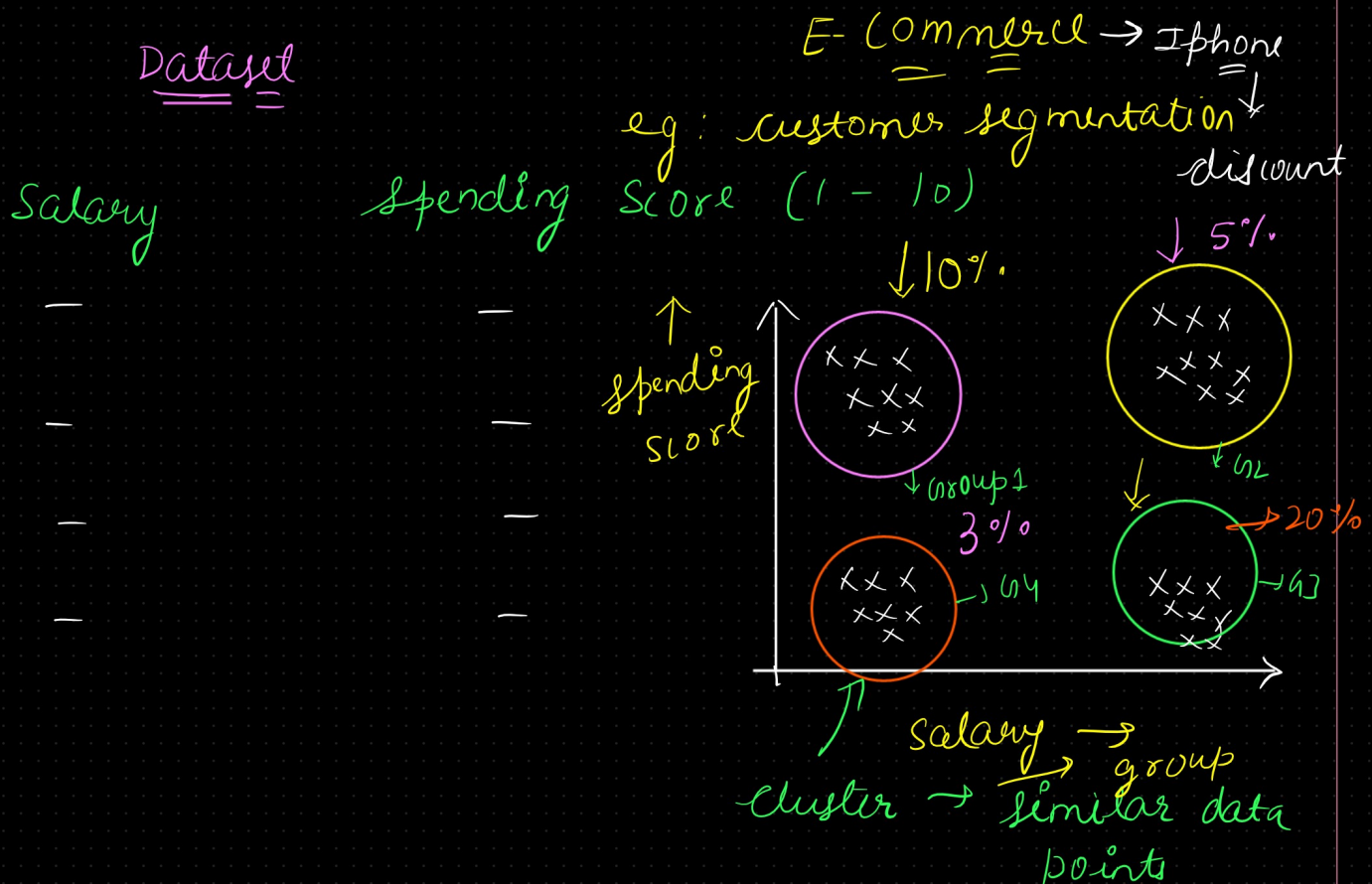
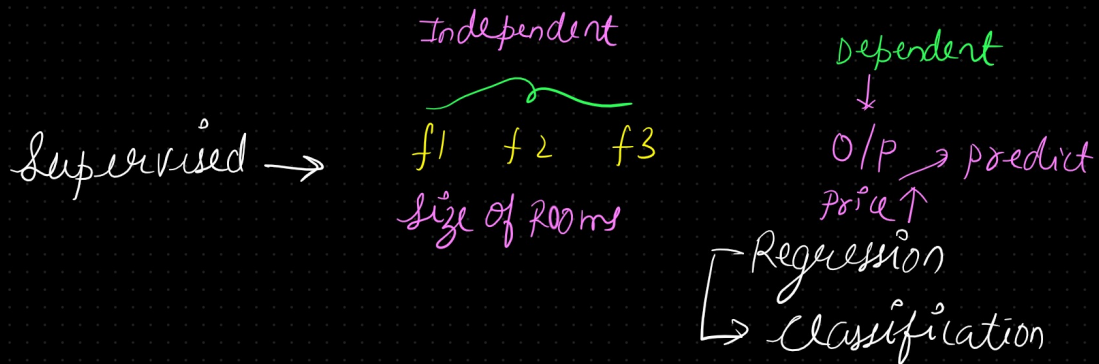


Unsupervised Machine Learning

Clustering Algorithm [K-Means, Hierarchical, DBSCAN]

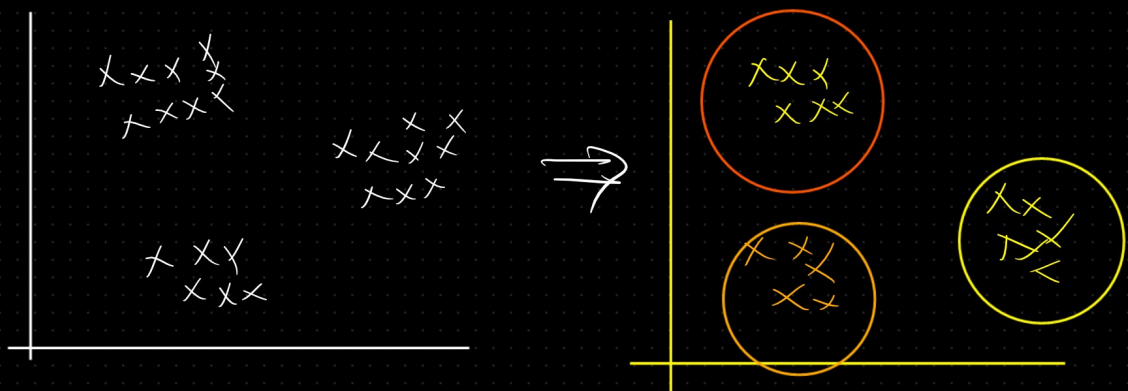
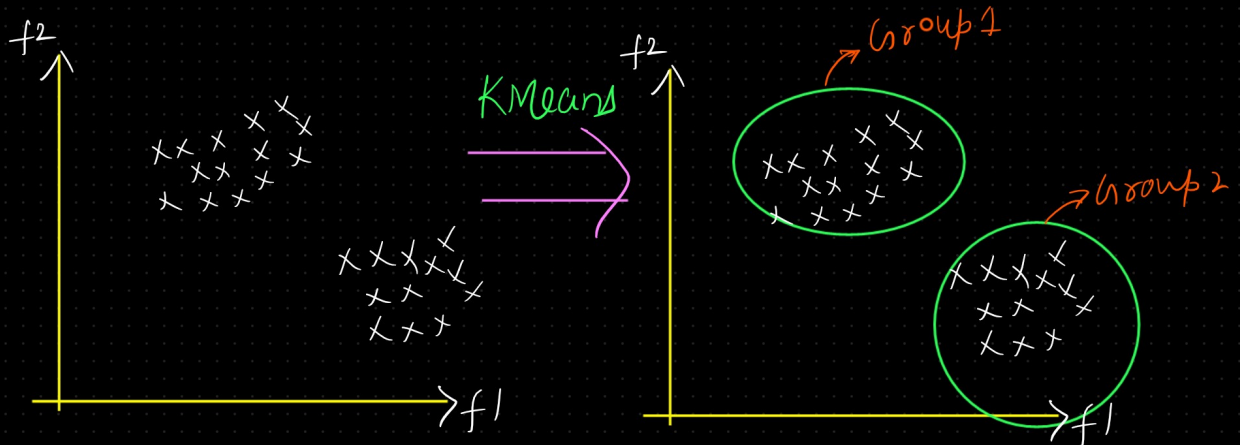
- Unsupervised machine learning involves finding patterns or structures in data without explicit guidance or labeled outcomes
- One common technique is clustering, where the algorithm groups similar data points together
- For example, in customer segmentation, you might use clustering to group customers based on their purchasing behavior without telling the algorithm what those groups should be



K-Means Clustering

→ K-means clustering is a popular unsupervised machine learning algorithm used for partitioning a dataset into K distinct, non-overlapping clusters

→ The goal of K-means is to find cluster centers that minimize the sum of squared distances from each data point to its nearest cluster center.



How it works

① Initialization

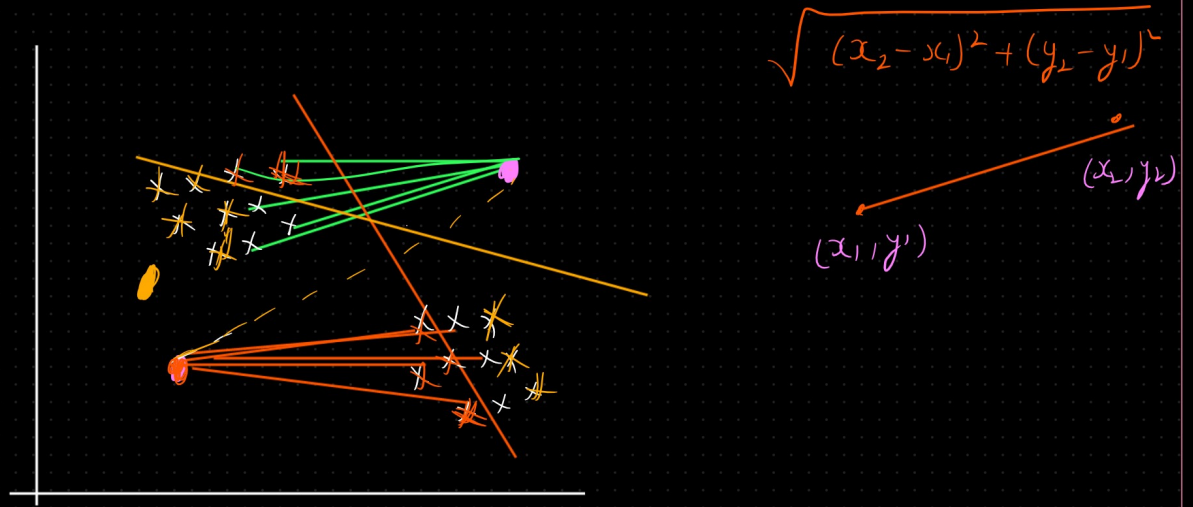
Randomly select three data points as initial centroids

② Assignment

$K = 2$ → centroid
↓
center of cluster

For each data point, calculate its distance to each centroid using a distance metric like Euclidean distance. Assign each data point to the nearest centroid.





③ Update:

Recalculate the centroids of the clusters by taking the mean of all the data points assigned to each centroid.

(Move the centroid) \rightarrow Mean

④ Repeat:

Repeat steps 2 and 3 until the centroids no longer change significantly or until a maximum number of iterations is reached.

Mathematical Formula

$$J = \sum_{i=1}^N \sum_{k=1}^K \mu_{ik} \|x^i - c_k\|^2$$

K = NO. of cluster

N = NO of data points

x_i = data points

c_k = centroid of k cluster

x_{ik} = Indicator variable

$$\text{otherwise } \begin{cases} x_{ik} = 1 & \text{if } x_i \rightarrow \text{cluster} \\ x_{ik} = 0 & \text{if cluster is not} \\ & \text{assign} \end{cases}$$

The objective function of K-means is to minimize the sum of squared distances

/

J from each data point to its assigned cluster centroid: