K-Means Clustering

Tuesday, 29 April 2025 07:01

Unsupervised Learning

↓

Data with input cols only

Amazon Customer's Data

↓

	Name	Age	City	Gender	Purchases	Amount spent
0						
1						
2						
3						
4						
5						

↓

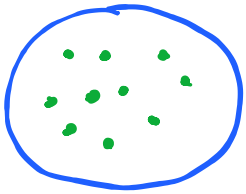
Unsupervised Algorithm

↓

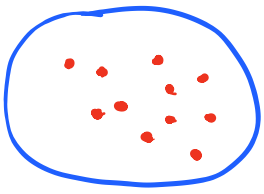
They will learn two things :

→ How many groups / clusters exist in our data ?

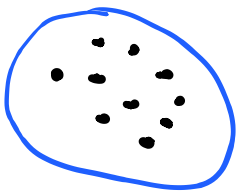
→ Which customer / datapoint belongs to which cluster.



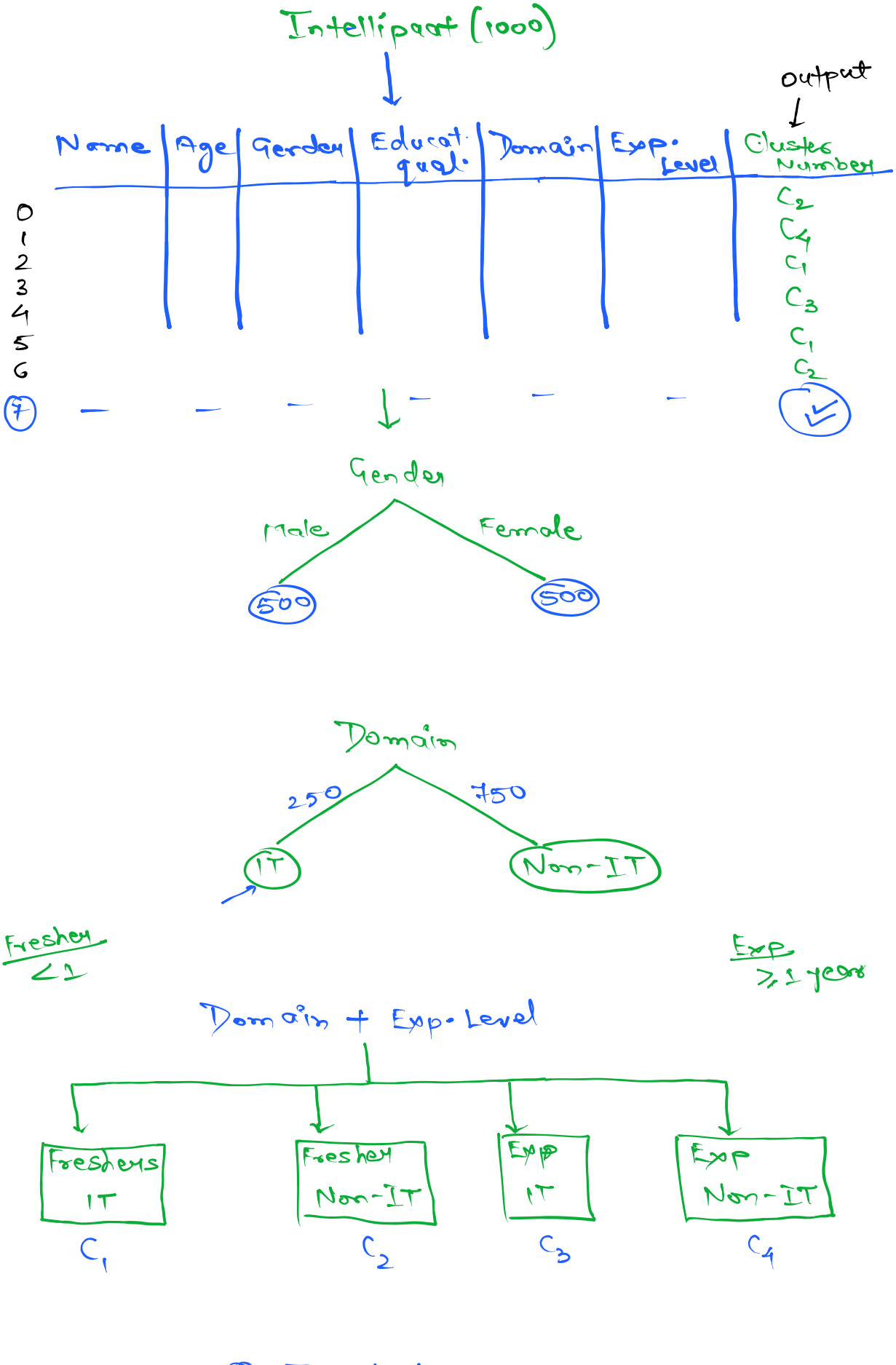
Low spenders



Medium spenders



High spenders



① Targeted marketing

② Recommendations

③ offers, coupons, discount

Unsupervised Learning Algorithms →

① K-Means clustering.

② Hierarchical Clustering.

③ DBScan clustering.

④ Gaussian mixture models (GMMs).

K-means Clustering

Two requirements of k-means :

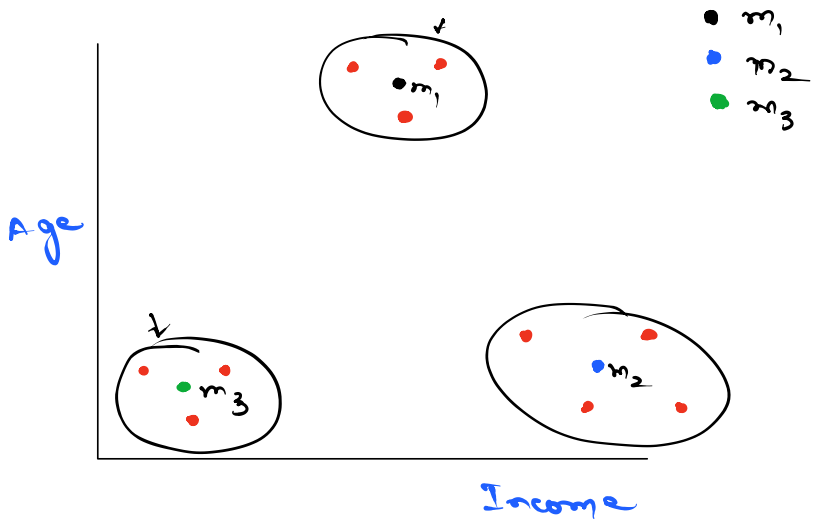
① Data (input cols) → 'X'

② k-value : No of clusters to be identified.

X = 10 rows

K = 3

k-means



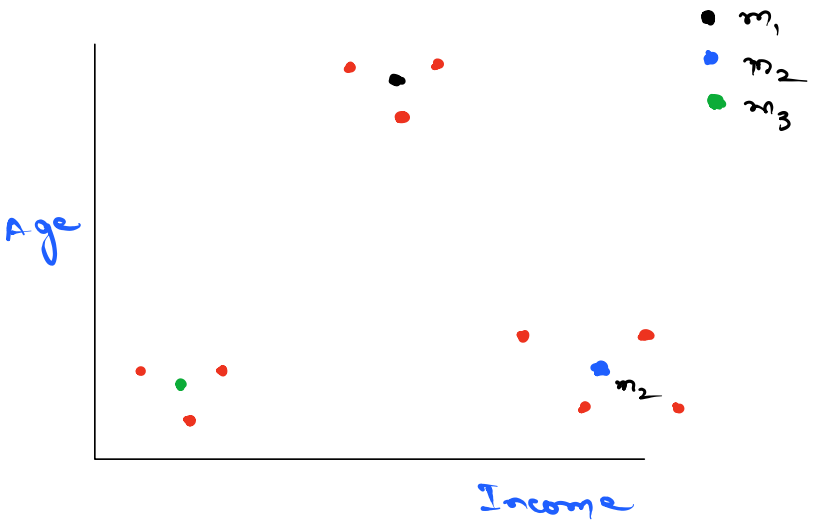
1. Create centroids / cluster means on the data.

2. Calculate the distance b/w each data point to each centroid / cluster mean.

3. Put together the datapoints with their closest centroids in clusters.

4. Update the centroid values.

5. Repeat steps 2-4 until no further changes observed.



Initialization Trap :

$K=3$