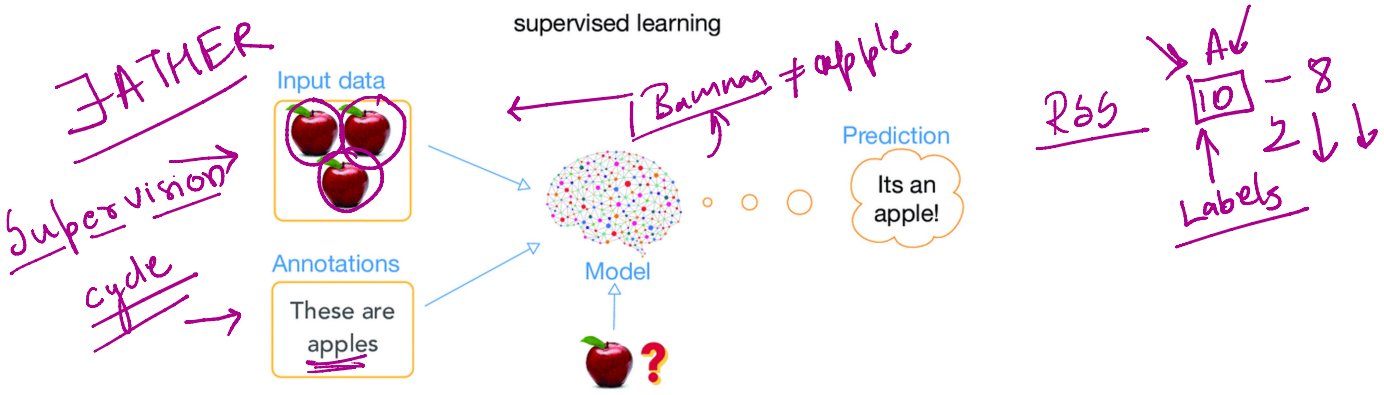
Unsupervised Learning

Sunday, January 8, 2023 8:29 PM

Difference → Supervised and Unsupervised

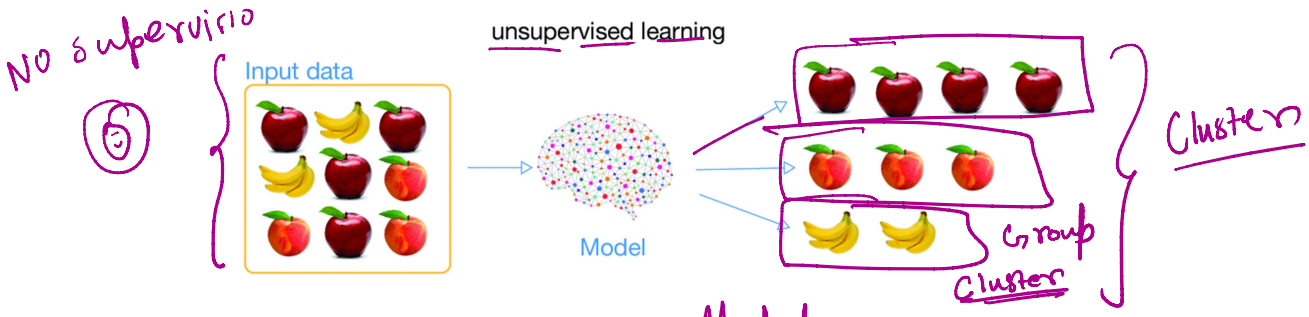
Labels — Without Labels



FATHER

Supervision cycle →

RSS → A → 10 − 8 → 2 → Labels



No supervision

Model is going to learn

[redacted]

Amazon : Labels

100M → Unsupervised Labels → 100M

1.	♀	10,000	Tshirt	Branded
2.	♀	299	Tshirt	Non Branded

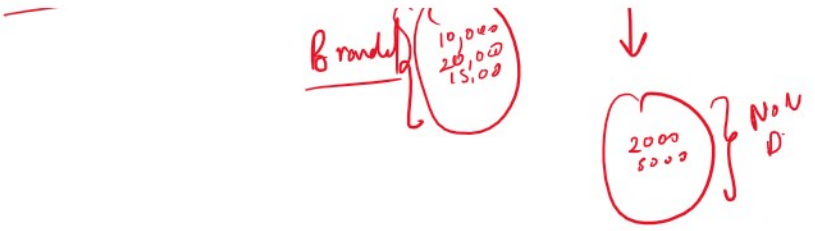
High
Low
Medium

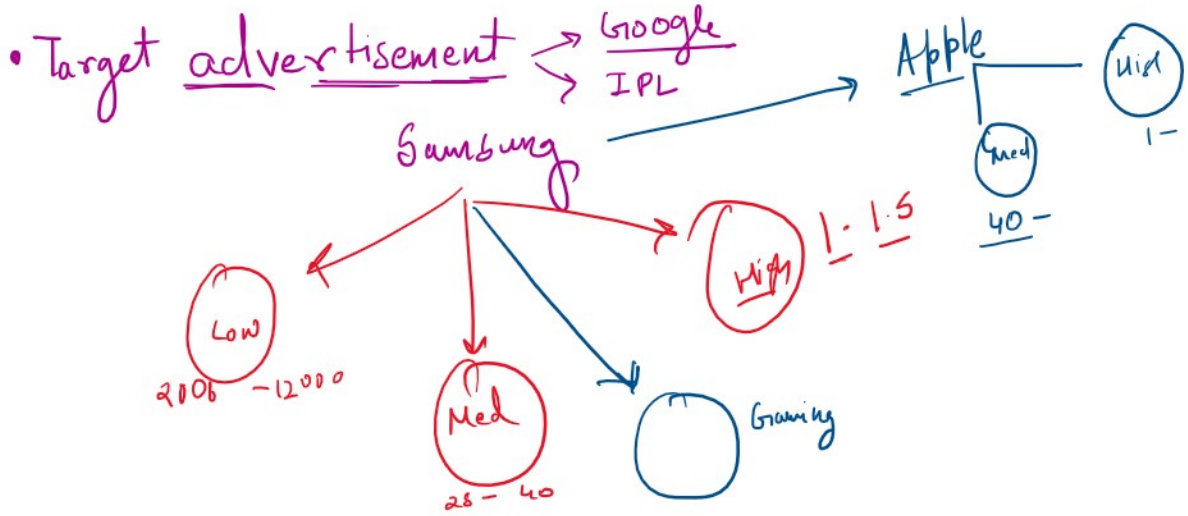
20 { Team Data Labelling

1 Month → T, M, Delayed

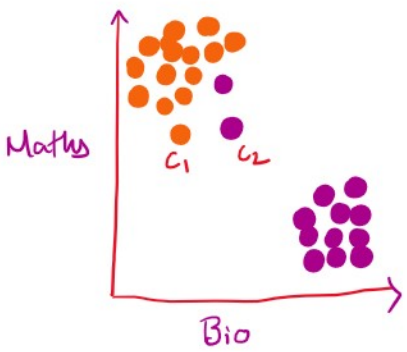
200 500 } 1 copy

Branded 10,000 20,000 15,000





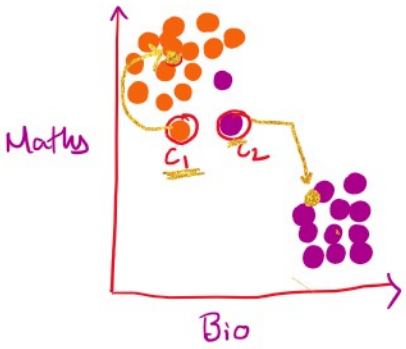
K Means



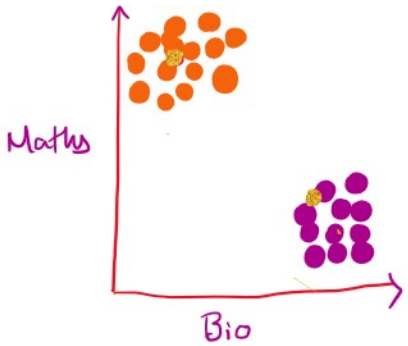
2 Step

→ 1 step
: Randomly assign Clusters Centroid

→ 2 step
Assignment Step



2 Step
→ 1 step centroid for cluster
: ~~Randomly assign clusters centroid~~
→ 2 step Assignment Step



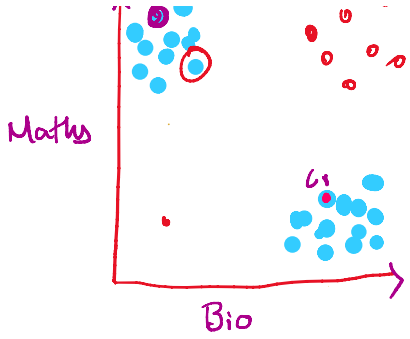
2 Step
→ 1 step centroid for cluster
: ~~Randomly assign clusters centroid~~
→ 2 step Assignment Step

- Pick the first centroid point (C_1) randomly.
- Compute distance of all points in the dataset from the selected centroid. The distance of x_i point from the farthest centroid can be computed by

From <https://towardsdatascience.com/understanding-k-means-k-means-and-k-medoids-clustering-algorithms-ad9c9fbf47ca>

K Means ++ } Smart Cluster centroid

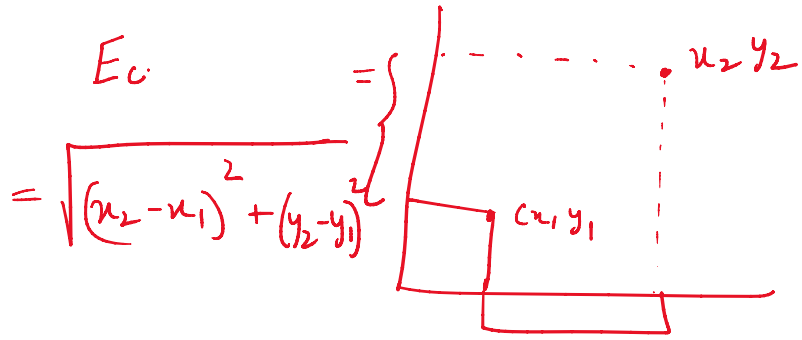
1. Randomly Select 1 Point



1. Randomly select 1 Point from the given Point

② dist is higher highest would be your next centroid

③ Calc the dist to its nearest centroid



K Means

Centroid

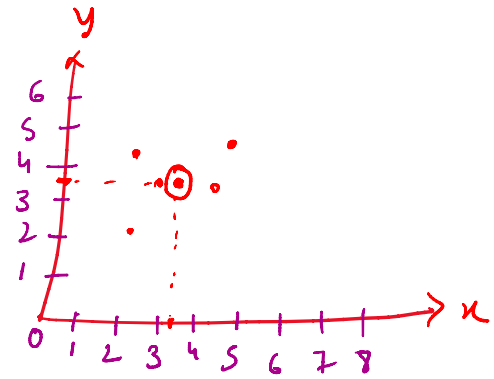
↓

No of cluster

Centroid

$$x = \frac{2+2+3+4.5+5}{5} = 3.3$$

$$y = \frac{2+3.5+4.5+3.5+4.7}{5} = 3.6$$



$d_1^2 + d_2^2$

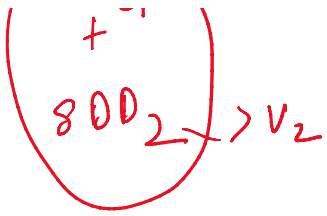
$(10)^2$

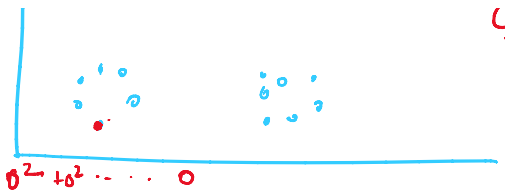
$(5)^2$

Elbow → sum of square distance

Cluster K=1 → SSD = V_1

K=2 SSD = V_2





Cluster K=1 SSD = V1

K=2 SSD = V2

K=3

⋮

K=n SSD = 0

Sum $d_1^2 + d_2^2 \ldots\ldots d_n^2$ n

{ Silhouttee, Hierarchical } → { PCA, LDA } TS