

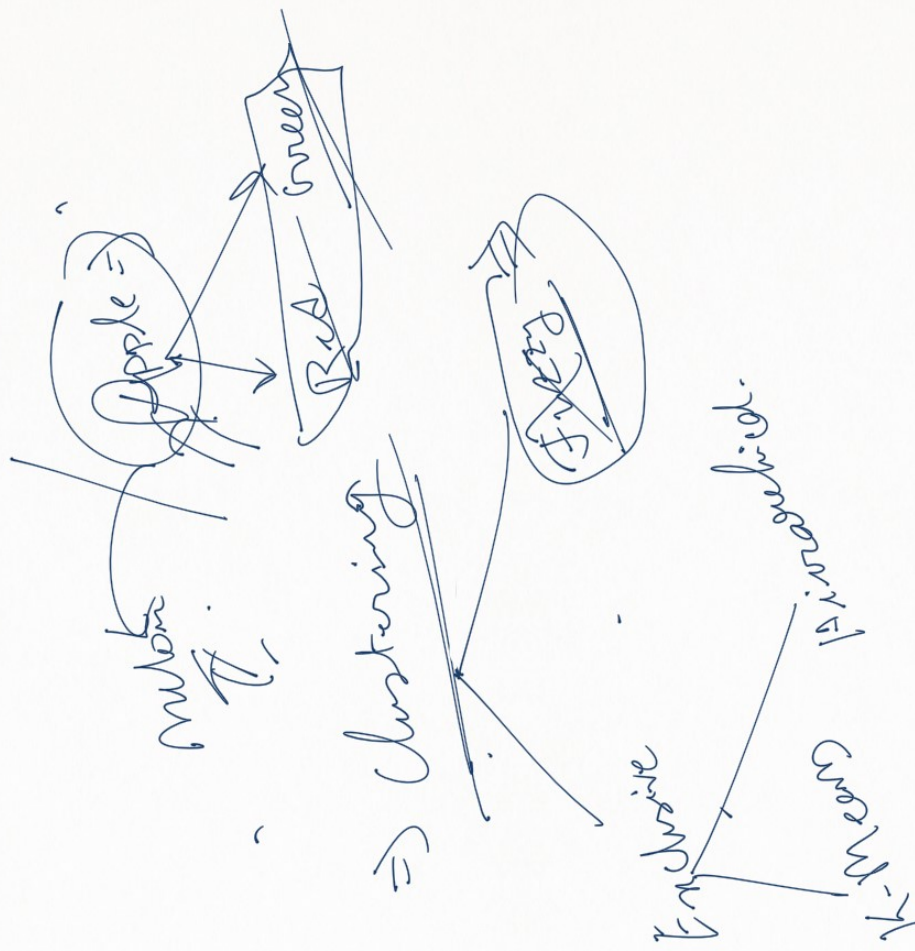
Unsupervised Learning

Label \Rightarrow Target

x_1, x_2, \dots, x_n

But then

Find similarity



K-Means Clustering

$K \Rightarrow$ No. of clusters $\Rightarrow \{c_1, c_2, \dots, c_k\}$

Extra Cluster Variance \Rightarrow ~~Total~~ \Rightarrow Low \Rightarrow

Distance betn clusters

$$\sum_{k=1}^K \sum_{(i,j) \in C_k} (x_{ij} - x_{i2})$$

$$\sum_{(i,j) \in C_1} (x_{i1} - x_{i1})^2 + \sum_{(i,j) \in C_2} (x_{i2} - x_{i2})^2$$

K-Means Clustering

- Random
K-Centroids \Rightarrow Avg. Point in a Cluster.
- i) Define K Points and centroids.
 - ii) Calculate Diff b/w Point Δ Centroid to be selected.
 - iii) min Diff of Point Δ Centroid
 - iv) Assign Point to min Centroid
 - v) Re-Calculate Centroids (Step-ii) to Step-iv) till
 - vi) Repeat from Step-ii no movements.
- Here are no

$$\begin{aligned} & \underline{C_1, C_2} \\ & x_1, x_2 \\ & \Rightarrow (x_1 - C_1), (x_1 - C_2) \\ & \Rightarrow (x_2 - C_1), (x_2 - C_2) \end{aligned}$$

C_1, x_2

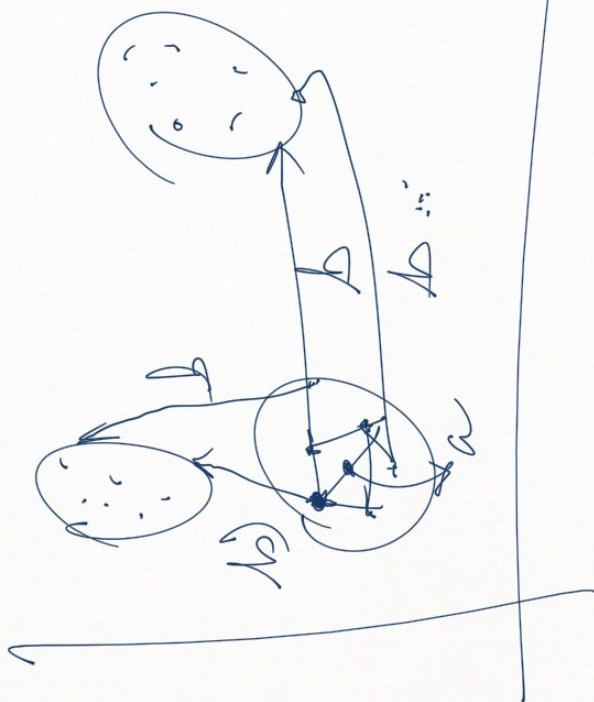
(x_1, x_2)

Silhouette Formula \Rightarrow

$$\frac{b - a}{\max(a, b)}$$

$b \Rightarrow$ Distance betw. a point to its
nearest cluster which it is not part of

$a \Rightarrow$ intra cluster difference.



$$\frac{b-a}{\max(a,b)} \approx 1 \Rightarrow$$

$$\underline{b > a}$$

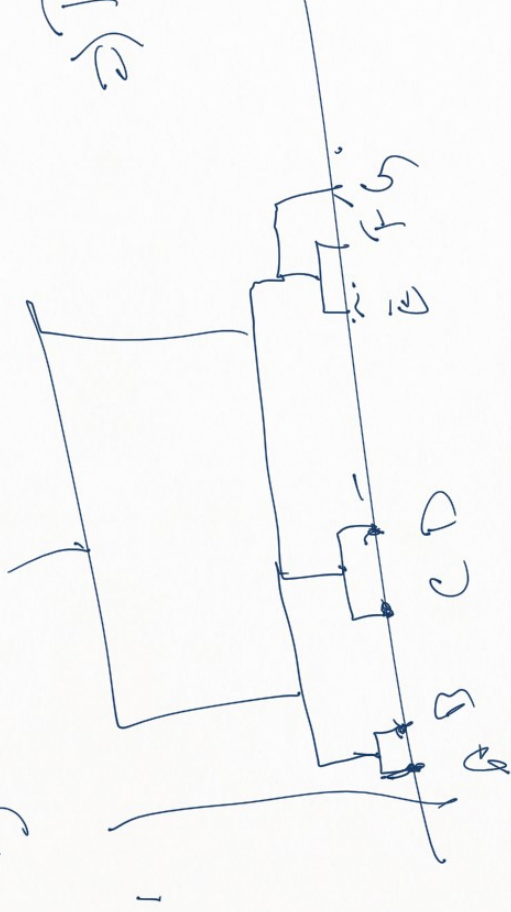
$$, a > b = -ve$$

Hierarchical Clustering

i) Every individual data points are a cluster. \Rightarrow

ii) Entire Dataset is a cluster.

Bottom Up \Rightarrow Agglomerative



Top Down \Rightarrow Divisive Clustering

- i) Which points to consider in a cluster \Rightarrow Euclidean distance.
- ii) Which clusters to fuse.

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

ii) Four clusters \Rightarrow

Linkage of

Single \Rightarrow

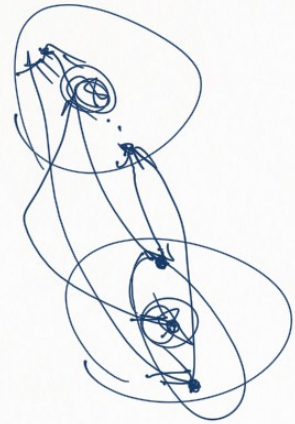
Complete \Rightarrow

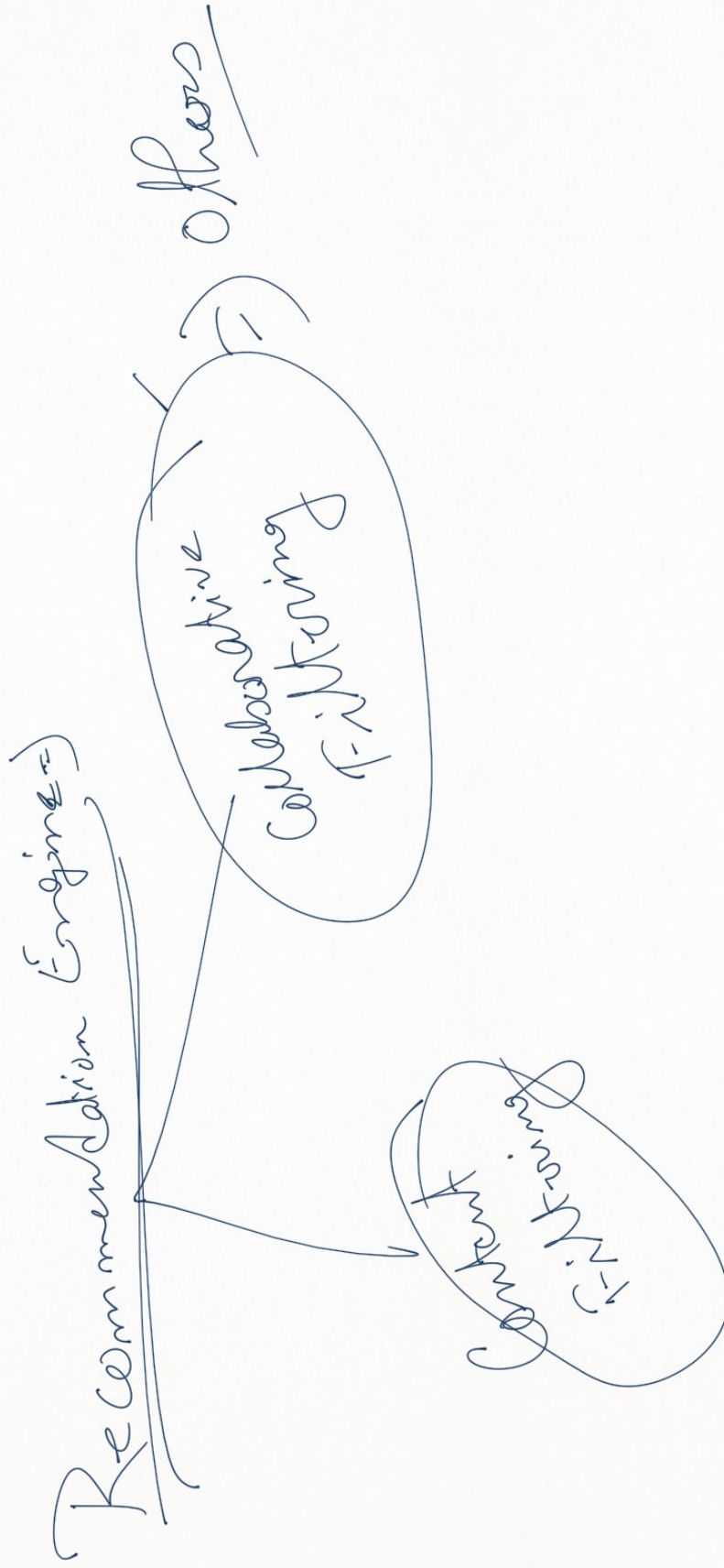
Group Avg \Rightarrow

Centroid \Rightarrow

Ward \Rightarrow

Distance \sim between clusters
 minimum \sim between clusters
 maximum \sim between clusters
 of all \sim between clusters
 avg \sim between clusters
 take diff \sim between clusters
 take \sim between clusters
 = \sim between clusters
 = \sim between clusters





Associate Rule Mining

Antecedent $A \Rightarrow B$ Consequent

Support Confidence

$$C(A \Rightarrow B) = \frac{S(A, B)}{S(A)}$$

$S(A) =$ Total no. of item appearance

Total transaction $S(Bread) = 20$

$$S(Milk) = 17 \quad S(Bread) = 20$$

$$S(Milk) = 17 \quad S(Bread) = 20$$

$$\begin{aligned} Milk &\Rightarrow 17 \Rightarrow 20 \\ Bread &\Rightarrow 10 \\ Egg &\Rightarrow 7 \end{aligned}$$

$$\{Milk, Bread\} = 8$$

$$\{Milk, Egg\} = 1$$

$$\{Bread, Egg\} = 4$$

$$\begin{aligned} L(A \Rightarrow B) &= \frac{S(A, B)}{S(A) * S(B)} \\ &= \frac{8}{17 * 20} = \frac{16}{170} \end{aligned}$$

$$\begin{aligned} L(MB) &= \frac{8}{17} \times \frac{20}{10} = \frac{16}{17} \\ L(B, E) &= \frac{4}{10} \times \frac{7}{7} = \frac{4}{7} \\ L(ME) &= \frac{1}{17} \times \frac{10}{7} = \frac{10}{119} \end{aligned}$$

$$\begin{aligned} C(M, B) &= \frac{8}{17} \\ C(B, E) &= \frac{4}{10} \\ C(M, E) &= \frac{1}{17} \end{aligned}$$

Apriori \Rightarrow Finds a threshold of support, confidence, lift.

Stores a threshold of single items.

i) Calculate support of threshold.

ii) Reject all with support $<$ threshold. \swarrow - starts from 2.

iii) Calculate support confidence, lift for k -item.

kill no item remains

iv) Calculate all $<$ any threshold.

v) Reject all \swarrow by k to $k+1$

vi) Repeat from

vii) Repeat