Today's Agenda

Supervised Learning

Label, target → Supervisor

1. Regression ⇒ target, label ⇒ {Numeric} (Cont., Discr.)
2. Classification ⇒ target or label ⇒ categorical
   - nominal
   - ordinal

Binary → 2 class

Multiclass → more than 2 class

Unsupervised Learning

✗ No label, No target ✗

↓

No Supervisor

Clustering or grouping

1. K-Means
2. Hierarchical
3. DBScan
4. Mean shift

Data → Weight | BMI | Height

Best fit line ?

Linear reg ⇒ Y = mx + c

M, C

→ Geometric intuition

W

B

H

(Act - Pred)

1. Calculation ⇒ $y = mx + c$
2. Loss ⇒ MSE $\frac{1}{N}\sum_{i=1}^{N}(Y - \hat{Y})^2$
3. optimization → GD → $W_{new} = W_{old} - \eta \frac{\partial L}{\partial W}$

Minimize ⇒ Model ⇒ $Y = mx + c$

Mathematical or algebraic

Unsupervised example

Weight	Weight
50	160
55	175
85	178
90	185
80	165

K-Means ⇒ mathematical

Hierarchical

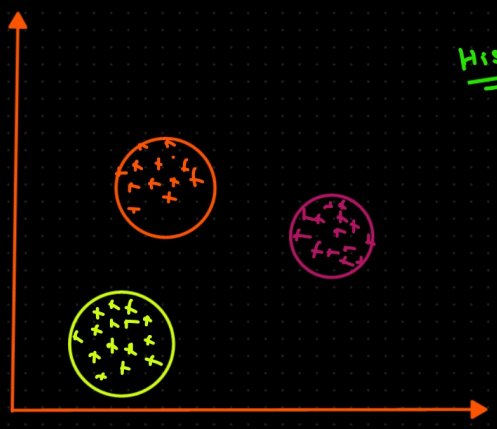
DBScan

→ Cluster

↓

collection of the elements

Group



History →

1. What is K-Means ?
2. KMeans v/s KMeans++
3. Hirarchical clustering
4. DBScan.
5. Hirarchical v/s K-Means
6. Why we use DBScan
7. evaluation Matrix
   1) Theory
   2) DUNN
   3) shiloute
8. inter & intra clustering
9. WCSS

K-Means

Data

1. Decide the K-value (K is number of cluster or K is Number of Centroid)
   K= 2,3,4,5 ...
2. Randomly initlize the Centroid.
3. Find out the Distance to all the Point and make cluster
4. Update the Centroid
5. Again you need to perform Step No. 3

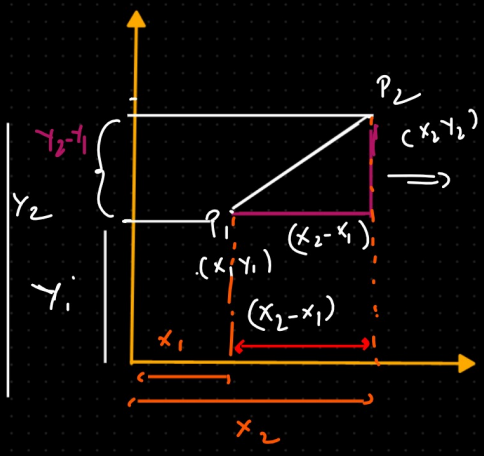
{until we don't get the Optimize cluster}

Distance

1. Euclidian Dist-
2. Manhattan dist-
3. Minkowski { Categorical }

Unsupervised → Clustering = Grouping — 3rd

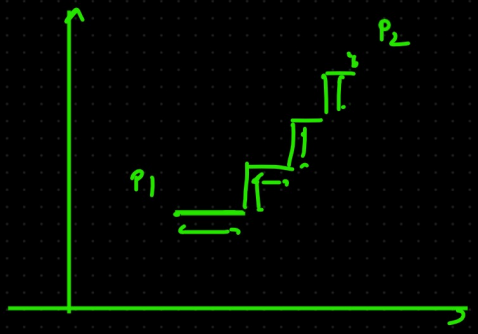
2nd < Distance >
- Similarity
- Data — 1st



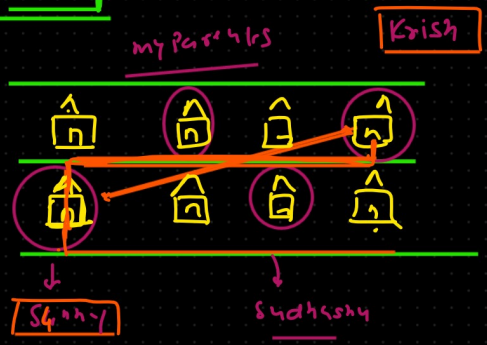
$$(y_2-y_1)\quad H^2 = 2^2 + 8^2$$

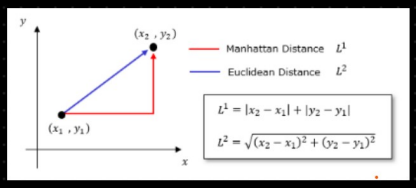
$$P_1P_2 = \sqrt{(y_2-y_1)^2+(x_2-x_1)^2}$$

Manhattan Path



Colony

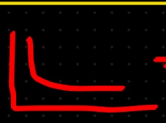




BLUE ⇒ Eclidian

Red ⇒ Manhattan

K-Means ⇒ No. of centroid ⇒ **Elbow Method** ⇒ WCSS

2, 3, 4, 5 ... ⟶  ⇒ With in cluster some of square

(1) Intra cluster dist
(2) Inter cluster dist

Data

	Height	Weight
1	185	72
2	170	56
3	168	60
4	179	68
5	182	72
6	188	77
7	180	71
8	180	70
9	183	84
10	180	90
11	180	67
12	167	76

(1) K ⇒ value

K=2 ⇒ 2 cluster
2 Centroid

(2) Randomly initial centroid

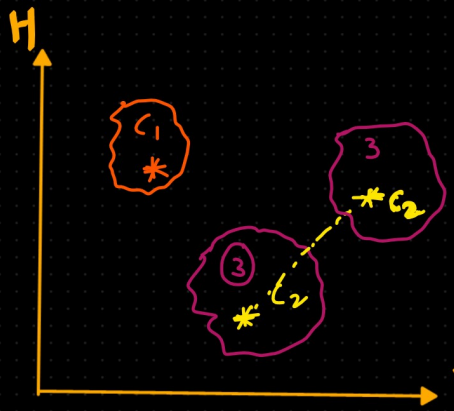
$C_1 \rightarrow (185, 72)$
$C_2 \rightarrow (170, 56)$

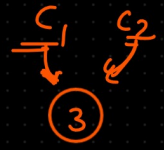
No. of centroid = No. of cluster

(3) Dist ⇒ $C_1 \rightarrow 3 = (185, 72)(168, 60)$
$C_2 \rightarrow 3 = (170, 56)(168, 60)$

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

$D(C_1, 3) = \sqrt{(185-168)^2 + (72-60)^2} = 20.80$
$D(C_2, 3) = \sqrt{(170-168)^2 + (56-60)^2} = 4.47$





(4) Update the centroid

$C_1 \rightarrow X$
$C_2 \rightarrow \checkmark$

$C_2 (170, 56)$
$3 (168, 60)$

$\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}$

$\left(\frac{170+168}{2}, \frac{56+60}{2}\right)$

New Updated centroid ⟶ (169, 59) ✓
(x, y)

C_1 (185, 72) ⟶ ④ $D(C_1, ④)$

C_2 (169, 59) ⟶ ④ $D(C_2, ④)$

$C_2 \Rightarrow$

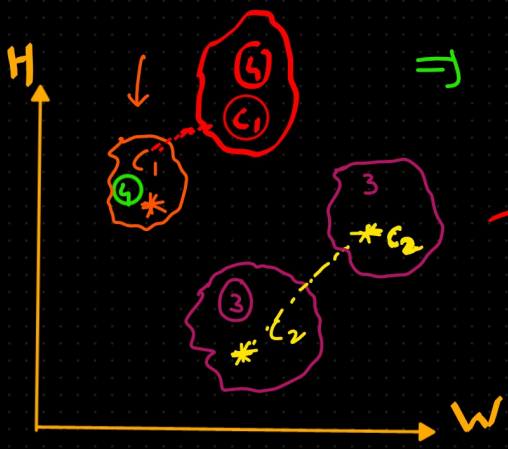
$D([185,72), (179,68)) = \sqrt{(185-179)^2 + (72-68)^2} = \sqrt{} = 7.21$

$D((169,59), (179,68)) = \sqrt{(169-179)^2 + (59-68)^2} = \sqrt{} = 13.45$

④

C_1 C_2

In whatever cluster I am going to append the Data Point I will update that centroid only.



⇒ Update Centroid

C_1 & C_2

(185, 72) C_1

(179, 68) ④

$\left[\frac{185+179}{2}, \frac{72+68}{2}\right]$

$\Rightarrow (182, 70)$ → C_1

Can You Do it?

All the Data Point ⇒ H.W.

K-Means Centroid **VS** K-Means Specific Approach

With Random Data Point from the data { for initializing the centroid }

KMeans

1. Centroid
2. Distance
3. include the DP in cluster and update the centroid.

K = 2, 3, 4, 5, 6.. n

WCSS => With in Cluster Sum of square

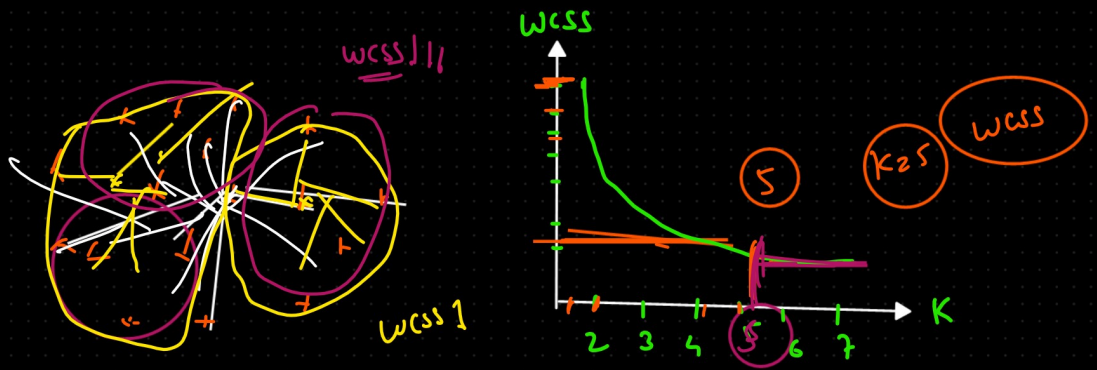
K = 1 — 2, 3, 4, 5 - - - - - - - - K = N

↓ All the Data Point will be in the same cluster

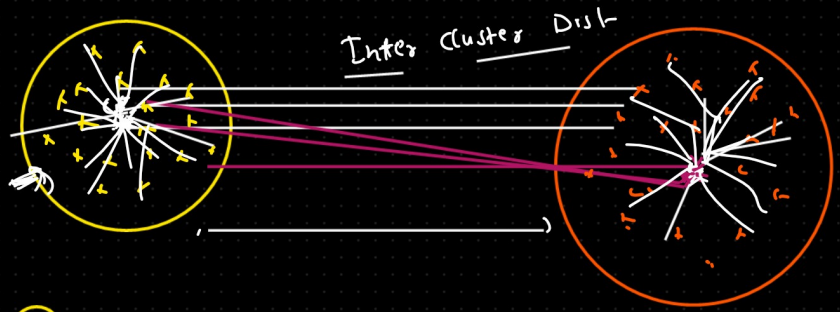
(every Data Point will be Consider as a individual Point-)

↑ ↑ ↑ ↑ WCSS

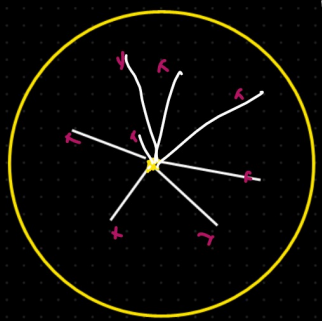
WCSS WCSS WCSS



WCSS => With in cluster Sum of Square



1. Inter Cluster Distance (B/w)
2. Intra Cluster Distance
   - (itself)



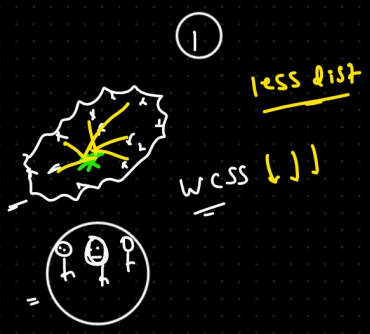
$W_i C_j \leq S$

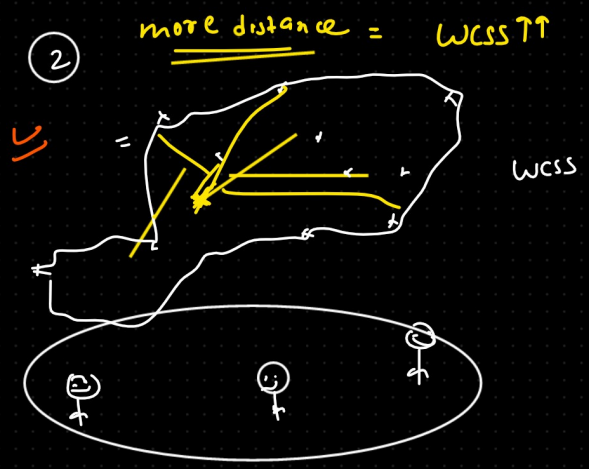
Within cluster

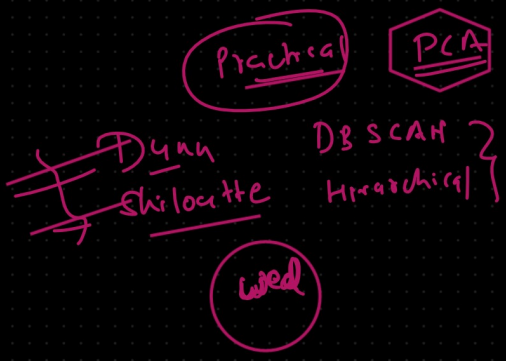
inside cluster

Dist => Square => Sum

$$= \sum_{i=1}^{n} \left(\sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2} \right)^2$$


1
less dist
WCSS ↓↓


2
more distance = WCSS ↑↑
WCSS


Practical
PCA
Dunn
Silhouette
DBSCAN
Hierarchical
used