Machine learning.

Supervised ML algo

Regression.

1. Ln reg — l1reg, l2reg
2. ~~Log reg~~
3. SVR.
4. DTR.
5. RFR
6. GBR
7. XBR
8. KNNR.

Classification

1. Log reg
2. SVC
3. DTC
4. RFC
5. XGB.
6. GBC / ABC
7. KNNC

Unsupervised ml algo

1. k-means → k-meant+
2. hirarchical.
3. DBScan.

target / dependent / supervisor

Supervised → algo

Supervisor

Height	Weight	BMI. (target)
170	60	21
180	65	22
160	70	20
165	75	18
140	55	19

Unsupervised ML algo → Clustering → grouping

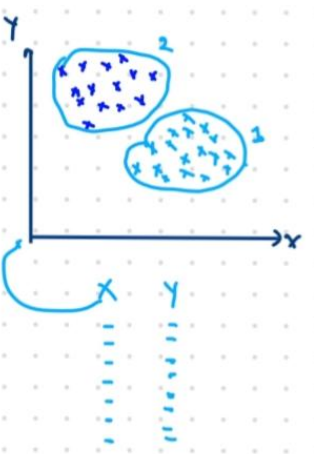
Data set

Height	Weight	BMI.	Country
170	60	21	IND
180	65	22	UK
160	70	20	USA
165	75	18	IND
140	55	19	USA

3-cluster — 3-group

mathmatically

1. k-means.
2. hirarchical.
3. DBSCan



① K-means → Data → Similarity → Distance → Euclidean dist

✓ ① centroid.
✓ ② Distance.
✓ ③ mean.

= K-means.

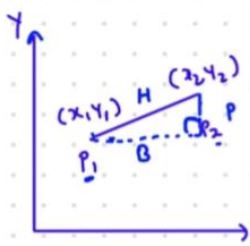
no. of centroid.
2, 3, 4, 5 ...

→ ELBOW method.

→ WCPS.

= [Inter cluster / Intra cluster]

{ → Dunn' index. / → Silhoutee coff. }



Pythagoras

$H^2 = P^2 + B^2$

$(P_1P_2)\,H = \sqrt{P^2 + B^2}$

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

K-means

		Height	Weight	→ Clustering
①	c_1	185	72	
②	c_2	170	56	
③		168	60	
④		179	68	
⑤		182	72	
⑥		188	77	
⑦		180	71	
⑧		160	59	
⑨		183	84	
⑩		180	88	
⑪		180	C_3	
⑫		167	76	

① Centroid. [randomly]

= K=2

NO of centroid = Number of cluster

c_1 (185,72) c_3 (168,60)

E.D. → $d(C_1,3)$, $d(C_2,3)$

$C_1 \to 3 = 5$ $C_2 \to 3 = 8$

$= \sqrt{(168-185)^2+(60-72)^2}$

$D(C_1,3) = = \sqrt{20.80}$

$= (185,72)$ $C_1, 3$

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

$\sqrt{(2)^2+(4)^2} = \sqrt{4+16}$

$= \sqrt{20} = 4.4$

C_2 (170,56) (168,60)

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

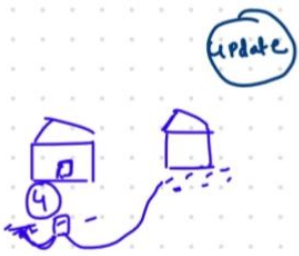
$= (169, 58)$

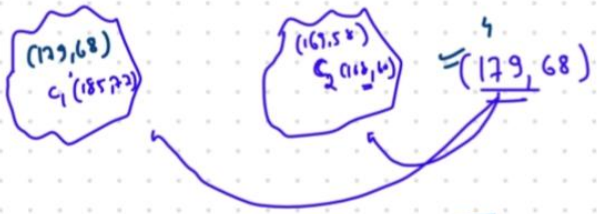
K-means. → K=2 → 2 centroid. → dis → similarity → min

c_1 (185,72)

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

$\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$





$$\frac{179+185}{2} = 182$$

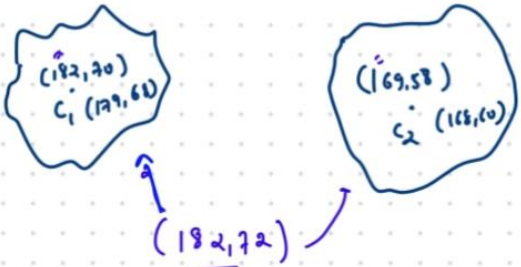
$$\frac{68+72}{2} = 70$$

$$D(c_1, 4) = \sqrt{(185-179)^2 + (72-68)^2} = 7.21$$

$$D(c_2, 4) = \sqrt{(169-179)^2 + (58-68)^2} = 14.14$$

Why = ? = low distance, min dist.

=



$$d(c_1, 5) = \sqrt{(182-182)^2 + (72-70)^2} = 2$$

$$d(c_2, 5) = \sqrt{(182-169)^2 + (72-58)^2} = 19.1$$

(182, 70)
(182, 72)

$$\frac{182+182}{2}, \frac{70+72}{2}$$

= (182, 71)

K-means.

1. Centroid
2. Distance. [compare, min]
3. Include Point in cluster, update the centroid.

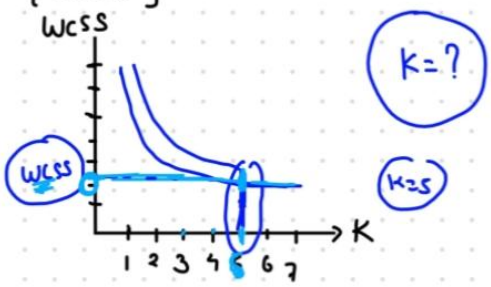
Iterate until all the data Point

K-means.

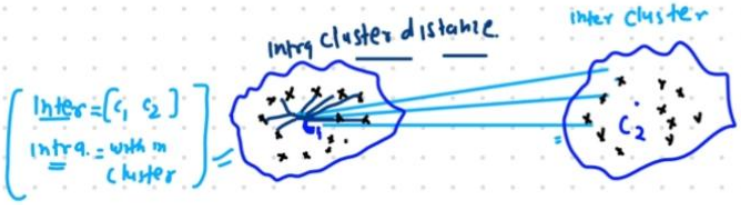
K = 2, 3, 4, 5, 6

[2 cluster].

[ELBOW.]



WCSS → with in cluster some of sqare



K=1

within cluster sum of sqre

= $WCSS_1$

K=2

$WCSS_2$ =

= $WCSS_1 > WCSS_2$

WCSS = within cluster sum of sqare

$$= \sum_{i=1}^{n} d(C, x_i)^2$$

K=3

$WCSS_1 > WCSS_2 > WCSS_3$

WCSS

K-means.

K = 1, 2, 3, 4, 5 ....

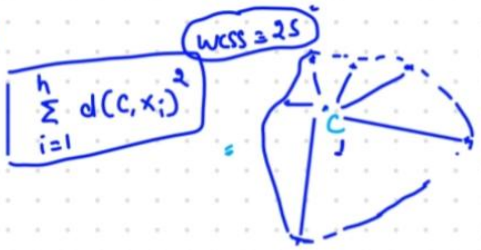
12 data Point

minimum. = 1

maximum = 12

WCSS.

K=1



K=2

WCSS = 20

K=3

WCSS = 15

= avg

diff b/w K-means | K-means++ =

how to validate cluster → K=5

Distance

- (1) dunn index.
- (2) Silhouette Score

Linreg

Regression.

→ R^2 / adj-R^2 R^2 = [0-1]

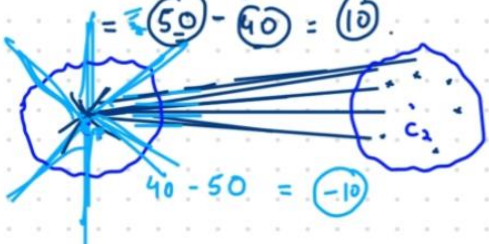
Classification worst [0] best [1]

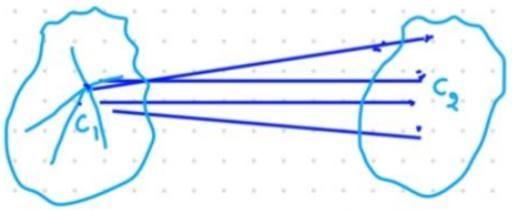
→ Roc AUC / Confusion mat-

(1) dunn index = $\frac{\max dist(x_i, x_j)}{\max dist(y_i, y_j)}$

(2) Silhouette score = $\frac{b_i - a_i}{\max(b_i - a_i)}$ same cluster [-1 to 1] = worst -1, best +1

= 50 - 40 = 10.

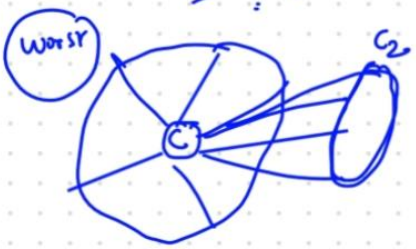




Same cluster

a_i b_i

$a_i > b_i$



K≥2

+ve

-1

(1) unsupervised.

(2) K-means. → (K=?)

(3) how to chose optimal K = ?

ELBOW | WCSS

(4) K≥5 → how to validate.

Silhoutte. → [-1, +1]

worst best

(1) random centroid.

(2) Distance, min.

(3) update the centroid

(2) hirarchical.

(3) DBScan

(4) Partical.

(1) You have to revise

(2) k means v/s k means++

(3) Dunn index.

How to make a best model or optimize sol^{n}.

Custom learning or Custom model = Supervised + Unsupervised.
(Semi Supervised)