life was EZ (until) Machine Learning.

Name: Ahtisham. Roll No 2127 (A.I)

DataSet

Train → [Model training]

input → features, Label.

Test → [Mode Evaluation]

features → input

lablels → output.

Baye's Theorem.

formula: $$P(H/X) = \frac{P(X|H)\,P(H)}{P(X)}$$

prior Probability.

$P(H) = P(C)$

↓

$P(C_1)$ & $P(C_2)$

↓ ↓

$\left(\frac{50}{100}\right)$ & $\left(\frac{50}{100}\right)$

X = features

C, H, $C_{1,2,3}$ > Classes.

Bayesian

Examples:-

if a customer bought computer or Not.

S.NO	Age	Income	Student	Credit Rating	Buy/No Buy.
1	<30	High	No	Fair	No
2	>40	Medium	No	fair	Yes
3	30-40	Low	Yes	Excellent	No.

features = 4 (Age, Income, Student, Credit) → (All Discrete. Because Ranges)

Label = 1 → 2 Classes (No / Yes)

→ New Customer X comes:-

age < 30
income = medium
student = yes
Credit = fair

Solving:-

$\Rightarrow P(C_i|X) = P(C_i)\, P(X|C_i)$

$P(C_i) = \begin{cases} P(C_1) = 9/14 \\ P(C_2) = 5/14 \end{cases}$

$P(X/C_1)$ & $P(X/C_2)$

$x_1, x_2, x_3, x_4 / C_1$

$x_1, x_2, x_3, x_4 / C_2$

$P(age^{<30} / Yes) = 2/9$

$P(age<30/No) = 3/5$

$P(medium/Yes) = 4/9$

$P(medium/No) = 2/5$

$P(Student/Yes) = 6/9$

$P(Student/No) = 1/5$

$P(fair/Yes) = 6/9$

$P(fair/No) = 2/5$

x (Mul)

x (Mul)

13/12/21

(Continous Variables)

Name: Roll No Page:

Same previous example, but income is continous.

PC_1 = same

PC_2 = same.

Discrete features = same = (age, credit, student).

Continous = income = ? — Dataset

Dataset → Separate (Yes) → Take Mean.

Dataset → (NO) separate. (same) → Take Mean.

Take Mean.

$$\mu = \bar{X} = \sum_{i=1}^{N} \frac{x_i}{n}.$$

↓

(1) $\frac{\text{Sum of all values}}{\text{total no. of value}}$.

↓

Take variance

$$\sigma^2 = \frac{\Sum (x-\mu)^2}{N}$$

↓

(2) every value '−' from Mean. / total no.

$$\frac{(x_1-\mu)^2 + (x_2-\mu)^2 \ldots}{N}$$

↓

(3) Gaussian formula.

$$\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right] \Rightarrow \left[\frac{1}{\sqrt{2(3.14)(14630)^2}} \exp\left(\frac{-(405-334)^2}{2(14630)^2}\right)\right]$$

Last Step

{ at the end Multiply Class Probability with $P(x/c)$.

(K-NN Classifier.)

nearest neighbour — NN.

We check distance between points of dataset and then classify furthur.

When new coordinate comes, we check its distance with dataset.

imp.

k=(odd value)

{because we don't want a tie

Example.

9,13 Good

9,16 Bad

9 - Good

13,16 Bad.

Bad 6

Goo 8,7,10,12

Read Api

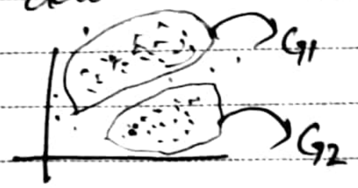
December.

24/12/21.

Name: Roll No Page

K-mean Clustring.

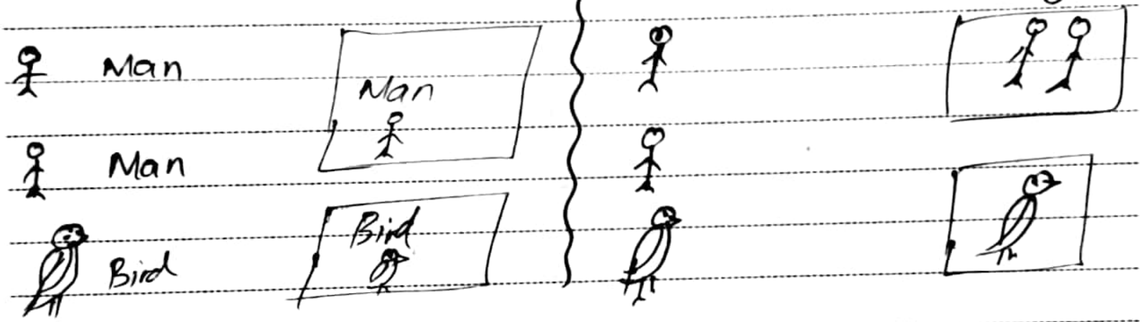
- Dividing data into groups based on common features.



Classification. VS Clustering.

Supervised
Classes

Un-Supervised
No-Classes (only Groups).



Centerpoint : Centroid is must for grouping.

$K == C$

Calculate distance of all points to all Centroids.

Value of $K \to$ Random but important.

Centroid $= C \Rightarrow$ based on value of K.

Customer DataSet.

x	y
2	10
2	5
8	4
5	8
7	5
6	4
1	2
4	9

Suppose $k=3$
so, $C=3$.

$C_1 = 2,5$
$C_2 \quad 5,8$
$C_3 \quad 1,2$

→ Gerated Randomly By algo.
No Algorithm. (Can be out of DataSet)

Dist. from C_1	Dist C_2	Dist C_3
$\sqrt{(2-2)^2+(10-10)^2}$	$\sqrt{(2-5)^2+(10-8)^2}$	$\sqrt{(2-1)^2+(10-2)^2}$
$\sqrt{(2-2)^2+(5-10)^2}$	$\sqrt{(2-5)^2+(5-8)^2}$	$\sqrt{(2-1)^2+(5-2)^2}$
⋮	⋮	⋮

	C_1	C_2	C_3	Class
P_1	0	5	9	C_1
P_2	5	6	4	C_3

{Pick lowest values. and assign to relative C.}

when made Clusters; check new/~~[illegible]~~ dataset.
update Cluster and perform Distance check again.
When value of C stops update then stop.

Calculate Centroid. Avg. $\frac{(n_1+n_2+n_3)}{n}$

Scanned with CamScanner