UNIT-III :

Supervised Learning :

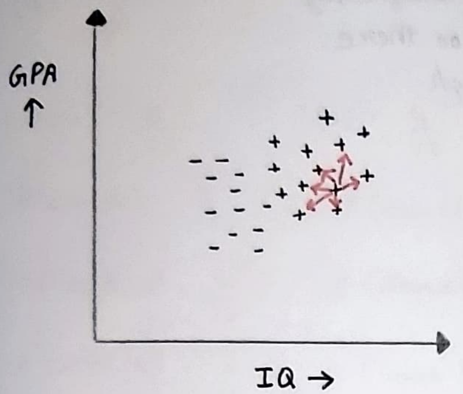
- K-nearest neighbours
- Naive Baye's
- Decision Tree

$\underbrace{\text{K-nearest neighbours, Naive Baye's, Decision Tree}}_{\text{CLASSIFICATION}}$

- Linear Regression
- Logistic regression

$\underbrace{\text{Linear Regression, Logistic regression}}_{\text{PREDICTION}}$

• K-nearest neighbour (Knn):



+ → placed (Y)

− → not placed (N)

— Distance metrics :

1) Euclidean Distance

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

2) Manhattan Distance

$$d(x,y) = \sum_{i=1}^{n} |x_i - y_i|$$

3) Minkowski Distance

$$d(x,y) = \left[\sum_{i=1}^{n} (x_i - y_i)^p\right]^{1/p}$$

Eg:

IQ	CGPA	Placed (Y/N)	distance		
73	7.2	Y	$\sqrt{3^2 + (.6)^2}$	=> 3.06	(1)
86	8.4	Y	$\sqrt{10^2 + (-.6)^2}$	=> 10.01	(3)
45	4.3	N	$\sqrt{31^2 + 3.5^2}$	=> 31.19	(6)
56	5.8	Y	$\sqrt{20^2 + (-2)^2}$	=> 20.09	(5)
32	3.0	N	$\sqrt{44^2 + 4.8^2}$	=> 44.26	(8)
95	9.1	Y	$\sqrt{(-19)^2 + (1.3)^2}$	=> 19.04	(4)
68	6.5	Y	$\sqrt{8^2 + 1.3^2}$	=> 8.104	(2)
35	3.2	N		=> 41.25	(7)

Q) which class does point (76, 7.8) belongs to? (Y/N)

i) K = 3 : Y = 3 ; N = 0 : (Y)

ii) K = 5 : Y = 5 ; N = 0 : (Y)

Ignore taking (K) to be even, it have chances of causing a TIE.

— How to choose value of (K):

(i) Heuristic method : $K = \sqrt{n}$ → # of data points

(ii) Experimental method : cross-validation : data → Train, Test ($K=3$)

→ Put various values of K and then check their accuracy.

K=1 (very low)
↳ case of overfitting
↳ case of variance

K=n (very high)
↳ case of underfitting
↳ bias can be there

→ K should not be very low or very high.

— Application of Knn :

(1) Spam detection.

(2) Health status detection

(3) Speech detection / classification

— Advantages of Knn :

(1) Easy to implement.

(2) No training is required, therefore known as Lazy Learning Technique

(3) Very few parameters required.

— Limitations of Knn :

(1) Large dataset.

(2) Not reliable in high dimensions as calculating distance is hard in multiple dimension and Knn depends on it.

(3) Outliers are problem as Knn is very sensitive to outliers.

(4) Imbalanced data.

(5) Scale of features / Non-homogeneous Scale of dataset can cause problem.

Naive Bayes :

18-03-25

	Dear	Friend	Lunch	Money	Total
N (8)	7	5	2	1	15
S (4)	2	1	0	4	7

$P(N) = \frac{8}{12}$ $\quad$ $P(S) = \frac{4}{12}$

$P(\text{Dear}/N) = \frac{7}{15}$ $\quad$ $P(\text{Dear}/S) = \frac{2}{7}$

$P(\text{Friend}/N) = \frac{5}{15}$ $\quad$ $P(\text{Friend}/S) = \frac{1}{7}$

$P(\text{Lunch}/N) = \frac{2}{15}$ $\quad$ $P(\text{Lunch}/S) = 0$

$P(\text{Money}/N) = \frac{1}{15}$ $\quad$ $P(\text{Money}/S) = \frac{4}{7}$

- $P(\underbrace{\text{Dear Friend}}_{x})$ is normal ; $P(N/x)$

$$\underbrace{\boxed{P(N/x)}}_{\text{Conditional probability}} \propto \underbrace{\boxed{P(N)}}_{\text{Prior}} \times \underbrace{\boxed{P(\text{Dear}/N) \times P(\text{Friend}/N)}}_{\text{Likelihood}}$$

- P(Lunch Money) is normal or spam
  - This will give 0 probability for spam that is BIASNESS, so we increment all the frequencies by (1).

	Dear	Friend	Lunch	Money	Total
N (8)	8	6	3	2	19
S (4)	3	2	1	5	11

$x = \{\text{Lunch Money}\}$

$P(N/x) \propto P(N) \times P(\text{Lunch}/N) \times P(\text{Money}/N)$

$\propto \frac{8}{12} \times \frac{3}{19} \times \frac{2}{19} = 0.011$

$P(S/x) \propto P(S) \times P(\text{Lunch}/S) \times P(\text{Money}/S)$

$\propto \frac{4}{12} \times \frac{1}{11} \times \frac{5}{11} = 0.013$

$\therefore$ hence x is more likely to be a SPAM.

Q)

Outlook	weather	wind	Cricket (Y/N)
Rainy	cool	True	N
Overcast	mild	False	Y
sunny	hot	T	Y
R	hot	T	N
R	mild	F	Y
S	cool	F	Y
S	hot	F	Y
O	mild	T	N
S	cool	T	N
O	hot	T	Y

→ we will find likability

Outlook	Y	N
Rainy	1/6	1/2
Sunny	1/2	1/4
overcast	1/3	1/4

weather	Y	N
cold	1/6	1/2
mild	1/3	1/4
hot	1/2	1/4

wind	Y	N
T	1/3	2/2 = 1
F	2/3	0

- $x = \{ \text{Sunny}, \text{hot}, \text{True} \}$

$$P(y|x) = \frac{P(y) \cdot P(x/y)}{P(x)} \Rightarrow \frac{P(y) \cdot P(\text{Sunny}/y) \cdot P(\text{hot}/y) \cdot P(\text{True}/y)}{P(\text{Sunny}) \cdot P(\text{hot}) \cdot P(\text{True})}$$

$$\Rightarrow \frac{\frac{6}{\cancel{10}} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3}}{\frac{4}{\cancel{10}} \cdot \frac{4}{10} \cdot \frac{6}{10}} \Rightarrow \frac{10 \cdot 10 \cdot}{4 \cdot 4 \cdot 2 \cdot 2 \cdot 3} \Rightarrow 0.520$$

$$P(N|x) = \frac{\frac{4}{\cancel{10}} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot 1}{\frac{\cancel{4}}{\cancel{10}} \cdot \frac{4}{10} \cdot \frac{6}{10}} \Rightarrow \frac{10 \cdot 10}{4 \cdot 4 \cdot 4 \cdot 6} \Rightarrow 0.260$$

[WE WILL PLAY]

- $x = \{ \text{Rainy}, \text{cool}, \text{True} \}$

$$P(y/x) = \frac{\frac{6}{\cancel{10}} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{3}}{\frac{3}{\cancel{10}} \cdot \frac{3}{10} \cdot \frac{6}{10}} \Rightarrow 0.102$$

$$P(N/x) = \frac{\frac{4}{\cancel{10}} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot 1}{\frac{3}{\cancel{10}} \cdot \frac{3}{10} \cdot \frac{6}{10}} \Rightarrow 1.85$$

incorrect dataset

— new likelihood: input { outlook, weather } O/P { wind }

outlook	T	F
Rainy	1/3	1/4
Sunny	1/3	2/4
overcast	1/3	1/4
weather	**T**	**F**
cold	1/3	1/4
mild	1/6	1/2
hot	1/2	1/4

- $x = \{ \text{overcast}, \text{hot} \}$

$$P(T/x) = \frac{\frac{6}{\cancel{10}} \cdot \frac{1}{3} \cdot \frac{1}{2}}{\frac{3}{\cancel{10}} \cdot \frac{4}{10}} \Rightarrow 0.833$$

$$P(F/x) = \frac{\frac{4}{\cancel{10}} \cdot \frac{1}{4} \cdot \frac{1}{4}}{\frac{3}{\cancel{10}} \cdot \frac{4}{10}} \Rightarrow 0.208$$

[WIND WILL BLOW]

2013

8)

Chills	headache	Runny nose	fever	flue
Y	mild	N	Y	N
Y	no	Y	N	Y
Y	Strong	N	Y	Y
N	mild	Y	Y	Y
N	no	N	N	N
N	Strong	Y	Y	Y
N	Strong	Y	N	N
Y	mild	Y	Y	Y

Likability

Chills	Y flu	N
Y	3/5	1/3
N	2/5	2/3

headache	Y flu	N
mild	2/5	1/3
no	1/5	1/3
strong	2/5	1/3

fever	Y flu	N
Y	4/5	1/3
N	1/5	2/3

• $x = \{y, mild, y\}$

$$P(Y|x) = \frac{\frac{5}{8} \times \frac{3}{5} \times \frac{2}{5} \times \frac{4}{5}}{\frac{4}{8} \times \frac{3}{8} \times \frac{5}{8}} = \frac{8}{5 \times 5 \times 4} = 1.024$$

$$P(N|x) = \frac{\frac{3}{8} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3}}{\frac{4}{8} \times \frac{3}{8} \times \frac{5}{8}} = \frac{8 \times 8}{4 \times 5 \times 3 \times 3 \times 3} \Rightarrow 0.118$$

(There will be Flu)

runny nose	Y	N
Y	4/5	1/3
N	1/5	2/3

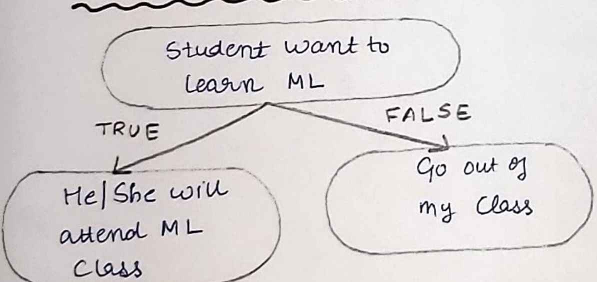
• $x = \{y, n, mild, y\}$

$$P(Y|x) = \frac{\frac{5}{8} \times \frac{3}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{4}{5}}{\frac{4}{8} \times \frac{5}{8} \times \frac{3}{8} \times \frac{5}{8}} = 0.546$$

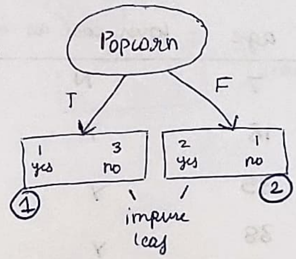
$$P(N|x) = \frac{\frac{3}{8} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3}}{\frac{4}{8} \times \frac{5}{8} \times \frac{3}{8} \times \frac{3}{8}} = 0.210$$

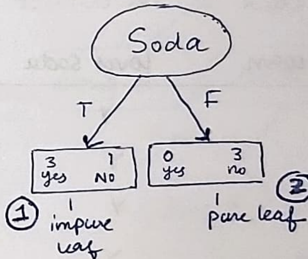
[Flu will be there]

Decision Tree:



Eg:





Eg: Loves popcorn	Loves soda	Age	Loves cool as ice
Y	Y	7	N
Y	N	12	N
N	Y	18	Y
N	Y	35	Y
Y	Y	38	Y
Y	N	50	N
N	N	83	N
0.4	0.21 min m	0.34	

① $GI(\text{Popcorn}) = 1-(P(y))^2-(P(n))^2 = 1-\left(\frac{1}{4}\right)^2-\left(\frac{3}{4}\right)^2 = 0.37$

② $GI(\text{Popcorn}) = 1-(P(y))^2-(P(n))^2 = 1-\left(\frac{2}{3}\right)^2-\left(\frac{1}{3}\right)^2 = 0.44$

– because no. of people in sets aren't same we will calculate the weighted impurity and then add it.

$\rightarrow \frac{4}{7} \times 0.37 + \frac{3}{7} \times 0.44 = 0.4$

① $GI(\text{soda}) = 1-\left(\frac{3}{4}\right)^2-\left(\frac{1}{4}\right)^2 = 0.37$

② $GI(\text{soda}) = 1-0-\left(\frac{3}{3}\right)^2 = 0$

weighted imp $= 0.37 \times \frac{4}{7} + 0 \times \frac{3}{7}$

$= 0.21$

Age : 7 12 18 35 38 50 83

avg : 9.5 15 26.5 36.5 44 66.5

Age < 9.5

T: 0 Y, 1 N | F: 3 Y, 3 N

GI (LL) = 0 GI(RR) = 0.5

weighted impurity = $0.5 \times \frac{6}{7}$ = 0.428

Age < 15

T (2): 0 Y, 2 N | F (5): 3 Y, 2 N

GI(LL) = 0 GI(RR) = $1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2$ = 0.485

weighted impurity = $0.195 \times \frac{5}{7}$ = 0.~~1~~34

Age < 26.5

T (3): 1 Y, 2 N | F (4): 2 Y, 2 N

GI(L) = 0.44 GI(R) = 0.5

weighted imp. = 0.47

A < 36.5

T: 2 Y, 2 N | F: 1 Y, 2 N

GI(L) = 0.5 GI(R) = 0.44

Avg weighted imp = 0.46

A < 44

T: 3 Y, 2 N | F: 0 Y, 2 N

GI(L) = 0.48 GI(R) = 0

Avg weighted imp = 0.342

A < 66.5

T: 3 Y, 3 N | F: 0 Y, 1 N

GI(L) = 0.5 GI(R) = 0

Avg weighted imp = 0.428

after chosing minm avg weighted impurity, fix that node

Loves Soda

(4): 3 Y, 1 N — impure leaf | (3): 0 Y, 3 N — pure leaf

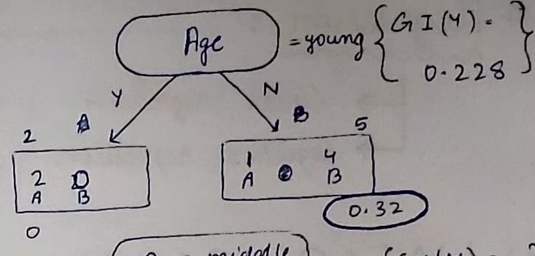
So we will create a new dataset with only (Y) values :

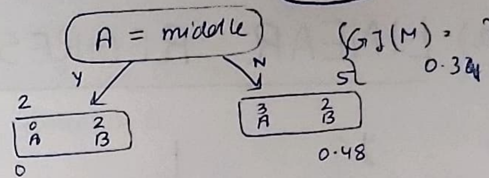
Loves pop corn	loves soda	age	loves cool as ice
Y	Y	7	N
N	Y	18	Y
N	Y	35	Y
Y	Y	38	Y

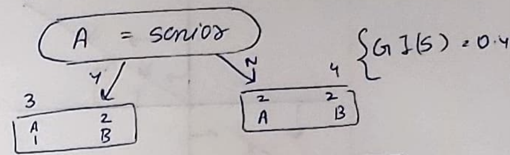
Q) Make decision tree for following data:

Age	Gender	BP	Cholestrol	Drug
Young	F	High	normal	A
Young	F	High	hight	A
middle	F	High	normal	B
Senior	F	normal	normal	B
Senior	M	low	normal	B
Senior	M	low	high	A
middle	M	low	high	B

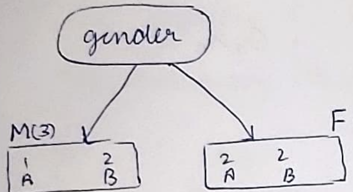
new







gender



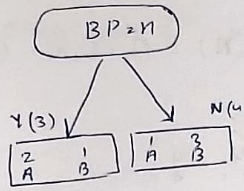
$GI(L) = 1-\left(\frac{1}{3}\right)^2-\left(\frac{2}{3}\right)^2$ $= 0.44$

$GI(R) = 1-\left(\frac{1}{2}\right)^2-\left(\frac{1}{2}\right)^2$ $= 0.5$

Avg GI $= \frac{3}{7} \times 0.44 + \frac{4}{7} \times 0.5$

$\Rightarrow 0.47$

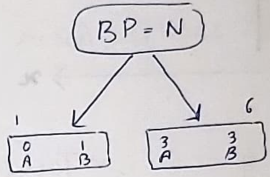
BP



GI (H) = 0.44 GI(R) = 0.375

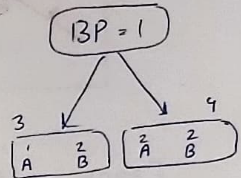
Avg $= \frac{3}{7} \times 0.44 + \frac{4}{7} \times 0.375$

$= 0.4$



GI(L) = 0 GI(R) = 0.5

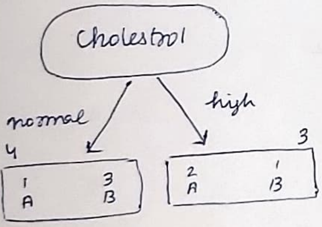
Avg = 0.42



GI(L) = 0.44 GI(R) = 0.5

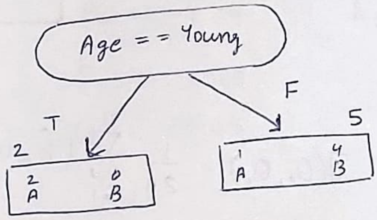
Avg = 0.188 + 0.255

= 0.47

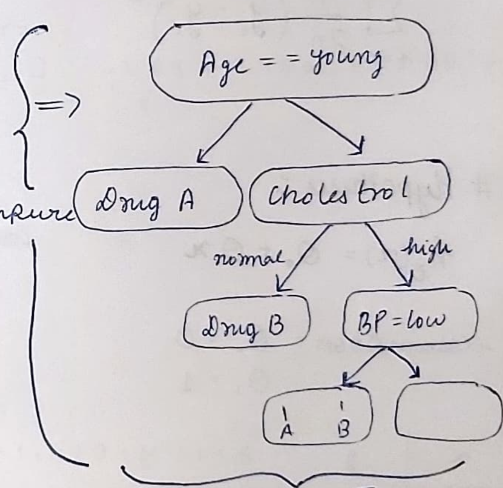


GI(L) = 0.375 GI(R) = 0.44

Avg = 0.40

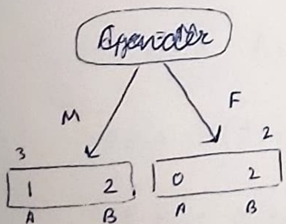


} impure



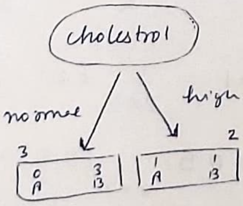
Decision Tree

2nd table:



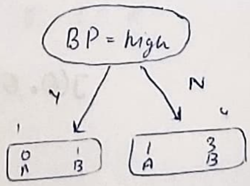
GI(L) = 0.44 GI(R) = 0

Avg = 0.264

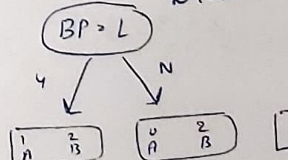


GI(L) = 0 GI(R) = 0.5

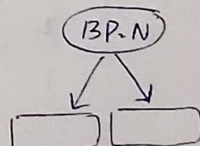
Avg = 0.2



Avg wt = 0.296

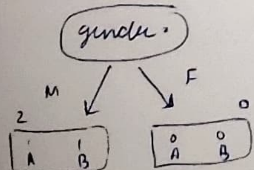


Avg wt = 0.264



Avg wt : 0.296

BP · low



→ 0.5

→ avg = 0.5

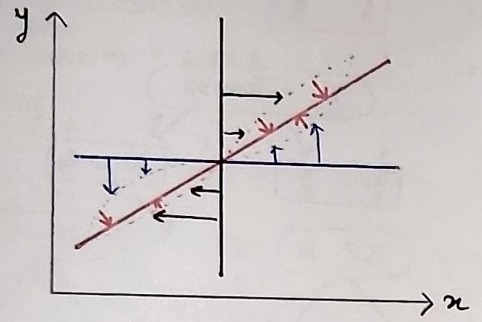
03/04/25

Regression :

- Linear regression : prediction
- logistics regression : classification using prediction

a) LINEAR REGRESSION :

AIM: To find the best fit line with minimum error.



$y = mx + c$

↑ slope ; intercept

$h_\theta(x) = \theta_0 + \theta_1 x$ ← for 1 input

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$

Error $= \hat{y}_i - y_i$; Total $= \sum_{i=1}^{n} (\hat{y}_i - y)$

Cost function :

$$\sum_{i=1}^{n} \frac{1}{2n} (\hat{y}_i - y_i)^2 \Rightarrow \frac{1}{2n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \Rightarrow \boxed{\frac{1}{2n} \sum_{i=1}^{n} (h_\theta(x)^{(i)} - y^{(i)})^2}$$

mean squared error

Hypothesis :

$h_\theta(x) = \theta_0 + \theta_1 x$

cost f^n : $J(\theta_0, \theta_1) = \frac{1}{2n} \sum_{i=1}^{n} (h_\theta(x)^i - y(i))^2$

assumption : $\theta_0 = 0$
$\theta_1 = 1$

x	y
1	1
2	2
3	3

$x = 1 ; y : 0 + 1.1 = 1$
$x = 2 ; y = 2$
$x = 3 ; y = 3$

$\theta_1 = 0.5$
$x = 1 ; y = 0.5$
$x = 2 ; y = 1$
$x = 3 ; y = 1.5$

$\theta_1 = 0$
$x = 1 ; y = 0$
$x = 2 ; y = 0$
$x = 3 ; y = 0$

$J(\theta_0, \theta_1) ; \theta_1 = 1 = \frac{1}{2 \times 3} (0^2 + 0^2 + 0^2) = 0$

$; \theta_1 = 0.5 = \frac{1}{2 \times 3} (0.5^2 + 1^2 + 1.5^2) = 0.58$

$; \theta_1 = 1.5 = \frac{1}{2 \times 3} (1^2 + 2^2 + 3^2) = 2.3$