Final

Syllabus

- → Probability Theory ✓ (video)
- → Naive Bayes / Bayes Theorem (Video).
- → Decision Tree ✓ (video)
- → Linear Regression. (Maliha)

Video 12

Probability Theory

$\boxed{\text{Picture}}$ (থেকে output (দেয় → Cat <.75> → Cat Final Decision

→ Dog <.25>

Agent → Store knowledge

System → detects the general weather of march.

weather → sunny, rainy, cold.

Need data of previous years march weather.

50 years → 45 march (sunny/hot)

In general, March → P(sunny/hot) → $\frac{45}{50}$

$= 90\%$

Machine learning classifier → Naive Bayes classifier

↳ model

↳ based on probability theory.

Terminology of probability

Events / variables → A B C } has domain value

Event A → (A , ¬A → not A)

Event B → (B , ¬B → not B)

একটা variable এর under এ multiple domain

ও থাকতে পারে।

Variable weather → ⟨ sunny , rainy , cloudy , snowy ⟩

$\underbrace{\text{sunny, rainy, cloudy, snowy}}_{\text{4 Domain Values}}$

প্রত্যেকটা Domain এর সাথে কিছু probabilistic value associated থাকবে, and summation of all probabilistic value will be 1.

Variable weather → ⟨'sunny', 'rainy', 'cloudy', 'snowy'⟩

⟨.72, .1, 0.08, 0.1⟩

$\Rightarrow .72 + .1 + .08 + .1 = 1$

$P(\text{sunny}) = .72$ $P(\text{rainy}) = 0.1$

$P(\text{cloudy}) = .08$ $P(\text{snowy}) = .1$

এভাবে represent করি Domain এর probabilistic value.

$P(A)$ → হওয়ার probability

$P(\neg A)$ → না হওয়ার probability

$P(A)$ $P(\neg A)$
$P(B)$ $P(\neg B)$

representation.

Probability এর value always $\boxed{0 \leq P(A) \leq 1}$

also can't be negative

$P(\neg A) = 1 - P(A)$

$A \to A, \neg A$

$P(A) + P(\neg A) = 1$

$P(\neg A) = 1 - P(A)$

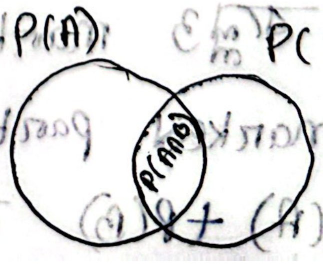
we can write.

Probability multiple variable combination এও হতে পারে।

$P(A \cap B) \leftarrow$ Joint Probability.

$P(A \cap B) \rightarrow$ আমাদের Duita event A and B একসাথে হওয়ার Probability.

Venn Diagram দিয়ে represent করতে পারি।



$P(A \cup B) \rightleftarrows P(A \text{ or } B)$

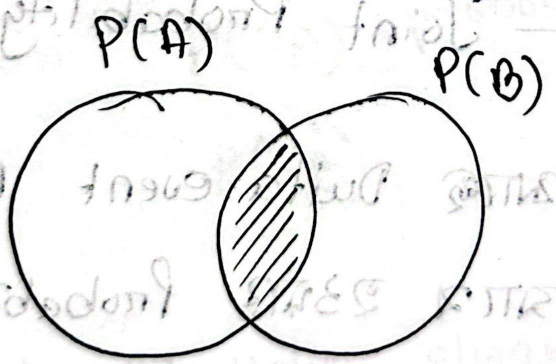
এটির মানে হলো হয় দুটি একসাথে হবে আর না হলে একটি হবে।

$P(A \cap B) \rightleftarrows P(A, B) \rightleftarrows P(A \text{ and } B)$ } Same meaning

formula

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

} relation between or, and.



P(A) বল্লেও marked part আসতিছে

P(B) বল্লেও marked part আসতিছে।

and marked part হচ্ছে $(A \cap B)$.

As, $P(A) + P(B)$ করলে marked part 2 বার

আমাদের তাই একবার $P(A \cap B)$ substract করেছি,

Another formula

$P(A \backslash B)$, A $\rightarrow$ covid, B $\rightarrow$ fever.

$\downarrow$

A given B মানে হলো B situation already occured. Given that situation, what is the outcome of A.

Situation: আমার already জ্বর আছে, Now what is my probability of getting affected by covid.

$$P(\text{covid} \backslash \text{fever}) = \frac{P(\text{covid} \cap \text{fever})}{P(\text{fever})}$$

covid $\rightarrow$ A, fever $\rightarrow$ B

$$= \frac{P(A \cap B)}{P(B)}$$

Joint Probability

$\hookrightarrow P(A \cap B) \rightarrow$ Probability of A and B happening at the same time.

$A \rightarrow A, \neg A$
$B \rightarrow B, \neg B$

A ∩ B এর outcome এ 4টি outcome আসতে পারে।

$P(A \cap B)$ $P(\neg A \cap \neg B)$
$P(A \cap \neg B)$ $P(\neg A \cap B)$

add করে 1 পাবো

$P(A \cap B)$: A and B একসাথে হওয়া।

$P(A \cap \neg B)$: A হওয়া and B না হওয়া।

Joint Probability distribution.

$P(\neg A \cap B)$: A না হওয়া and B হওয়া

$P(\neg A \cap \neg B)$: A and B কোনটিই একসাথে না হওয়া

we can represent it in tabular form.

	A	¬A
B	$P(A \cap B)$	$P(\neg A \cap B)$
¬B	$P(A \cap \neg B)$	$P(\neg A \cap \neg B)$

Joint distribution table.

	A	¬A
B	.75	.15
¬B	.05	.05

Sum = 1.

$.75 + .15 + .05 + .05 = 1$

$P(A \cap B) = .75$

$P(\neg A \cap B) = .15$

$P(A \cap \neg B) = .05$

$P(\neg A \cap \neg B) = .05$

alarm $\longrightarrow$ alarm, $\neg$alarm

↓ ↑

buglary $\longrightarrow$ buglary, $\neg$buglary.

	alarm	$\neg$ alarm
buglary	.09	.01
$\neg$buglary	.1	.8

Exam.

$P(\text{alarm} \cap \text{buglary}) = .09$

$P(\neg\text{alarm} \cap \text{buglary}) = .01$

$P(\neg \text{buglary} \cap \text{alarm}) = .1$

$P(\neg\text{alarm} \cap \neg \text{buglary}) = .8$

$P(\text{alarm}) = .09 + .1 = .19$

$P(\neg\text{alarm}) = .01 + .8 = .81$

$P(\text{alanm})$

$= p(\text{alanm} \cap \text{buglany}) + P(\text{alanm} \cap \neg \text{buglany})$

$= .09 + .1$

$= .19$

$P(\text{buglany}) = P(\text{buglany} \cap \text{alanm}) + P(\text{buglany} \cap \neg \text{alanm})$

$= .09 + .01$

$= .1$

$P(\neg \text{buglany}) = P(\neg \text{buglany} \cap \text{alanm}) + P(\neg \text{buglany} \cap \neg \text{alanm})$

$= .1 + .8$

$= .9$

$P(\neg \text{alanm}) = P(\neg \text{alanm} \cap \text{buglony}) + P(\neg \text{alanm} \cap \neg \text{buglony})$

$= .01 + .8$

$= .81$

P(alarm \ buglary) → conditional probability

$$\Rightarrow \frac{P(\text{alarm} \cap \text{buglary})}{P(\text{buglary})} = \frac{.09}{.1} = .9$$

এটা বুঝাচ্ছে যে buglary হলে, alarm বাজার probability .9 অথবা 90%.

$$P(\neg \text{alarm} \setminus \neg \text{buglary}) = \frac{P(\neg \text{alarm} \cap \neg \text{buglary})}{P(\neg \text{buglary})}.$$

$$= \frac{.8}{.9}$$

$$= .88$$

↳ buglary না হলে alarm না বাজার probability.

Cheated on College Exam?

	Yes	No
Male	.32	.22
Female	.28	.18

P (Male ∩ Yes) = .32

P (Male ∩ No) = .22

P (Female ∩ Yes) = .28

P (Female ∩ No) = .18

P (Yes) = .32 + .28 = .6

P (No) = .22 + .18 = .4

P (Male) = .32 + .22 = .54

P (Female) = .28 + .18 = .46

$$P(\text{Male} \setminus \text{cheated}) = \frac{P(\text{Male} \cap \text{cheated})}{P(\text{cheated})}$$

$$= \frac{.32}{.6} = .533$$

P (Female ∩ No) = .18

$$P(\text{Male} \setminus \text{Yes}) = \frac{P(\text{Male} \cap \text{Yes})}{P(\text{Yes})} = \frac{.32}{.6} = .53$$

if any already did cheat then, ~~propa~~ probability of the person being Male is .53 %.

	Right handed	Left handed	
Male	.41	.08	.49 = Male
female	.45	.06	.51 = female
	.86	.14	
	"	"	
	Right handed	Left handed.	

(a)

$$P(M \mid R) = \frac{P(M \cap R)}{P(R)} = \frac{.41}{.86} = .477$$

(b)

$$P(R \mid M) = \frac{P(R \cap M)}{P(M)} = \frac{.41}{.49} = .837.$$

(c)

$$P(F \mid L) = \frac{P(F \cap L)}{P(L)} = \frac{.06}{.14} = .429$$

(d)

$$\left.\begin{array}{l} P(F \mid L) = .429 \\ P(F) = .51 \end{array}\right\} \neq$$

So, F and L are not independent.

Joint Distribution table ২টির বেশি variable নিয়েও হতে পারে।

	tootache		¬tootache	
	Show on XRay	¬Show on XRay	Show on XRay	¬Show on XRay
Cavity	.108	.012	.072	.008
¬cavity	.016	.064	.144	.576

t = toodache
xr = xRay
c = cavity.

3 variable each 2 outcomes.

$2^3 = 8$ values.

if 4 variable with 2 outcomes

then, $2^4 = 16$ outcome values.

if 5 variable with 2 outcomes.

then, $2^5 = 32$ outcome values.

$P(t \cap a \cap c) = .108$

$P(\neg t \cap \neg x \cap \neg c) = .008$

$P(x) = .108 + .016 + .072 + .144$

= marginal probability

Question

↳ $P(t \cap c)$

$= .108 + .012$

$= .12$

↳ $P(\neg c \cap x)$

$= .016 + .144$

$= .16$

↳ $P(\neg x)$

$= .012 + .064 + .008 + .576$

$= .66$

↳ $P(t \mid \neg x) = \dfrac{P(t \cap \neg x)}{P(\neg x)} = \dfrac{.012 + .064}{.66} = \dfrac{.076}{.66}$

$= .115\%$

Video 13

	Toothache		¬ Toothache	
	Xray	¬ Xray	Xray	¬ Xray
Cavity	.108	.012	.072	.008
¬ Cavity	.016	.064	.144	.576

Question

↳ $P(\text{toothache} \mid \text{cavity} \cap \text{Xray})$

$$= \frac{P(\text{toothache} \cap \text{cavity} \cap \text{Xray})}{P(\text{cavity} \cap \text{Xray})}$$

$$= \frac{.108}{.108 + .072}$$

$$= \frac{.108}{.18} = 0.6$$

Independence

Two branch

- → full independence
- → Conditional independence

independence usually 2 টি event এর perspective এ discuss করা হয়।

A B

A এর outcome doesn't effect the outcome of B → A and B are independent.

Scenario

Fair Coin toss, $P(h) = .5$ $P(t) = .5$

30 বার টানা head পড়লেও 31 বার এ head and tail পড়ার probability same. Doesn't affect. Independent

So, fair coin tosses are independent.

Formula

$P(A \cap B) = P(A) \times P(B)$

$L.H.S = \cancel{P(H) *} \; P(H \cap T)$

$= .25$

$R.H.S = P(A) * P(B)$

$= P(H) * P(T)$

$= .5 * .5$

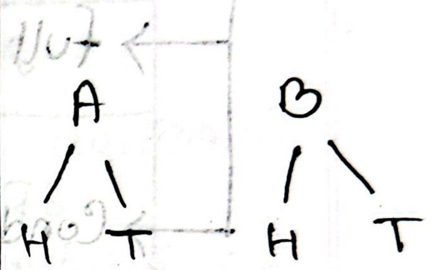
$= .25$

$\therefore L.H.S = R.H.S$

$P(A \cap B) = P(A) * P(B)$

So, coin tosses are independent.

Two coin toss.



Outcomes:

HH → .25

HT → .25

TH → .25

TT → .25

Conditional Independence

Bias Coin Toss {But don't know which side bias}.

Scenario : Coin toss করলাম, 3 বার, 3 বারই head পরলো and আমি জানি coin টি bias. So, 4th time head and tail পরার probability same থাকবে না।

$\boxed{\begin{aligned} P(H) &= .5 \\ P(T) &= .5 \end{aligned}}$ Prothom 2 বার ও র পর probability change হয়ে গেলো।

$P(H) = .75$

$P(T) = .25$

Bias টা condition হিসেবে set করে দিলে তখন এটা conditional independence

We are introducing a particular condition and after this conditional independence is achieved.

Example: 2 টা person একটি place এ খেতে গেছে and আমি জানি Storm হলে তাদের বাসায় আসতে late হবে।
initially আমাকে বলা হয় নাই যে, Storm হচ্ছে নাকি না।

Person 1 late করে আসলো বাসায়।
person 1 কে late করতে দেখে আমি assume করলাম, যে good chance যে Storm হচ্ছে।
So, person 2 ও বাসায় late করে আসতে পারে।

Person 1 এর late এ আসা বা না আসা Person 2 এর আসা বা না আসার probability effect করতিছে।

Now, Storm হচ্ছে এটা একটা given condition.
already আমি জানি যে ~~late 2~~ Storm হচ্ছে,

So, তখন both person 1 and 2 এর বাসায় late করে আসার probability বেশি থাকবে।

And, তখন, person 1 late করে আসতিসে নাকি না এটার উপর base করে person 2 এর late করে আসার probability change হবে না।

Condition → Snow storm is happening. Based on the condition person 1 and person 2 independent.

Formula

$$A \underset{\text{independence}}{\rightleftarrows} B \xrightarrow[\text{condition}]{\text{Given}} C$$

L.H.S　　　　R.H.S

$$P(A \cap B \mid C) = P(A \mid C) * P(B \mid C)$$

* কোন event independent মানে যে তারা conditional independent হবে তেমন না।

L.H.S = R.H.S

equal হলে conditional independent otherwise not.

$$P(\underbrace{A \cap B}_{A} \mid \underbrace{C}_{B}) = P(A \mid C) * P(B \mid C)$$

same

$$\frac{P(A \cap B \cap C)}{P(C)} = P(A \mid C) * P(B \mid C)$$

same

$$P(A \cap B \cap C) = P(A \mid C) * P(B \mid C) * P(C)$$

Maths of Normal Independence

$P(A \cap B) = P(A) * P(B)$

if equal then independent.

	Smart		¬Smart	
	Study	¬Study	Study	¬Study
Prepared	.432	.16	.084	.008
¬prepared	.048	.16	.036	.072

↳ is smart independent of study.

$P(\text{Smart} \cap \text{Study}) = P(\text{smart}) * P(\text{study})$

L.H.S

$= P(\text{smart} \cap \text{study})$

$= P(.432 + .048)$

$= .48$

R.H.S

$P(\text{smart}) * P(\text{study})$

$= (.432 + .16 + .048 + .16) *$

$(.432 + .048 + .084 + .036)$

$= .8 * .6$

$= .48$

L.H.S = R.H.S

Smart and study are independent.

↳ Is prepared independent of study.

P(prepared ∩ study) = P(prepared) * P(study).

L.H.S	R.H.S
P(prepared ∩ study)	= p(prepared) * P(Study)
= P(.432 + .084)	= P(.432 + .16 + .084 + .008) * (.432 + .048 + .084 + .036)
= P(.516)	= (.684) * (0.6)
= .516	= .4104.

So, they are not independent.

Math of Conditional Independence

	Smart		¬ smart	
	Study	¬ Study	Study	¬ Study
Prepared	.432	.216	.084	.008
¬ prepared	.048	.16	.036	.072

↳ is smart conditionally independent of prepared, given study.

P(Smart ∩ Prepared \ Study) = P(Smart \ Study) * P(Prepared \ Study)

L.H.S

$P(\text{smart} \cap \text{Prepared} \setminus \text{Study})$

$= \dfrac{P(\text{Smart} \cap \text{Prepared} \cap \text{Study})}{P(\text{study})}$

$= \dfrac{.432}{.432 + .048 + .084 + .036} = \dfrac{.432}{.6}$

$= .72$

R.H.S

$= P(\text{smart} \setminus \text{Study}) * P(\text{Prepared} \setminus \text{Study})$

$= \dfrac{P(\text{Smart} \cap \text{Study})}{P(\text{Study})} * \dfrac{P(\text{Prepared} \cap \text{Study})}{P(\text{Study})}$

$= \dfrac{.432 + .048}{.6} * \dfrac{.516}{.6}$

$= .8 \times .86$

$= .688$

$L.H.S = .72$

$R.H.S = .688$

$\therefore L.H.S \neq R.H.S$

$\therefore$ Smart and prepared are not conditionally independent given study.

↳ Is study conditionally independent of prepared, given smart?

$$P(\text{study} \cap \text{prepared} \setminus \text{smart}) = P(\text{study} \setminus \text{smart}) * P(\text{prepared} \setminus \text{smart})$$

L.H.S

$= P(\text{study} \cap \text{prepared} \setminus \text{smart})$

$= \dfrac{P(\text{study} \cap \text{prepared} \cap \text{smart})}{P(\text{smart})}$

$= .432 / .8$

$= .54$

R.H.S

$P(\text{study} \setminus \text{smart}) * P(\text{prepared} \setminus \text{smart})$

$= \dfrac{P(\text{study} \cap \text{smart})}{P(\text{smart})} * \dfrac{P(\text{prepared} \cap \text{smart})}{P(\text{smart})}$

$= \dfrac{.432 + .048}{.8} * \dfrac{.432 + .16}{.8}$

$= .6 * .74$

$= .44$

L.H.S = .54

R.H.S = .44

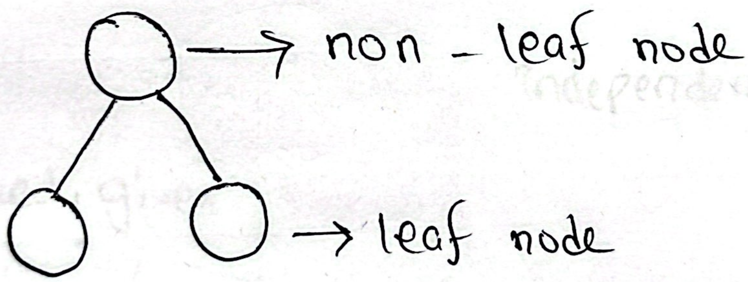
$\therefore$ L.H.S $\neq$ R.H.S

So, study and prepared are not conditionally independence given smart.

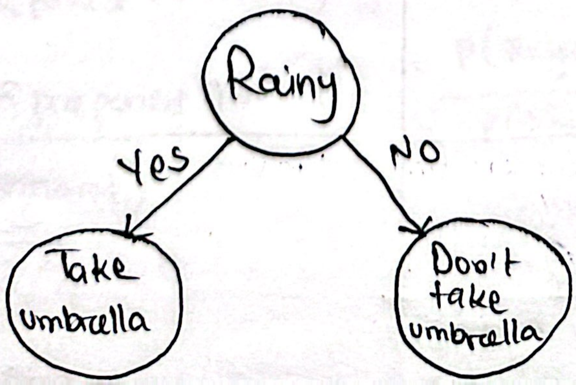
Video 16

Learning (Decision Tree)

in real life, if we are planning to get out of home, we take a look outside if we have to carry our umbrella or not. So, watching/observing the weather we are taking a decision.



যেই criteria অনুযায়ী আমাদের Decision নিচ্ছি সেগুলো leaf node এ রাখছি।



Basic tree

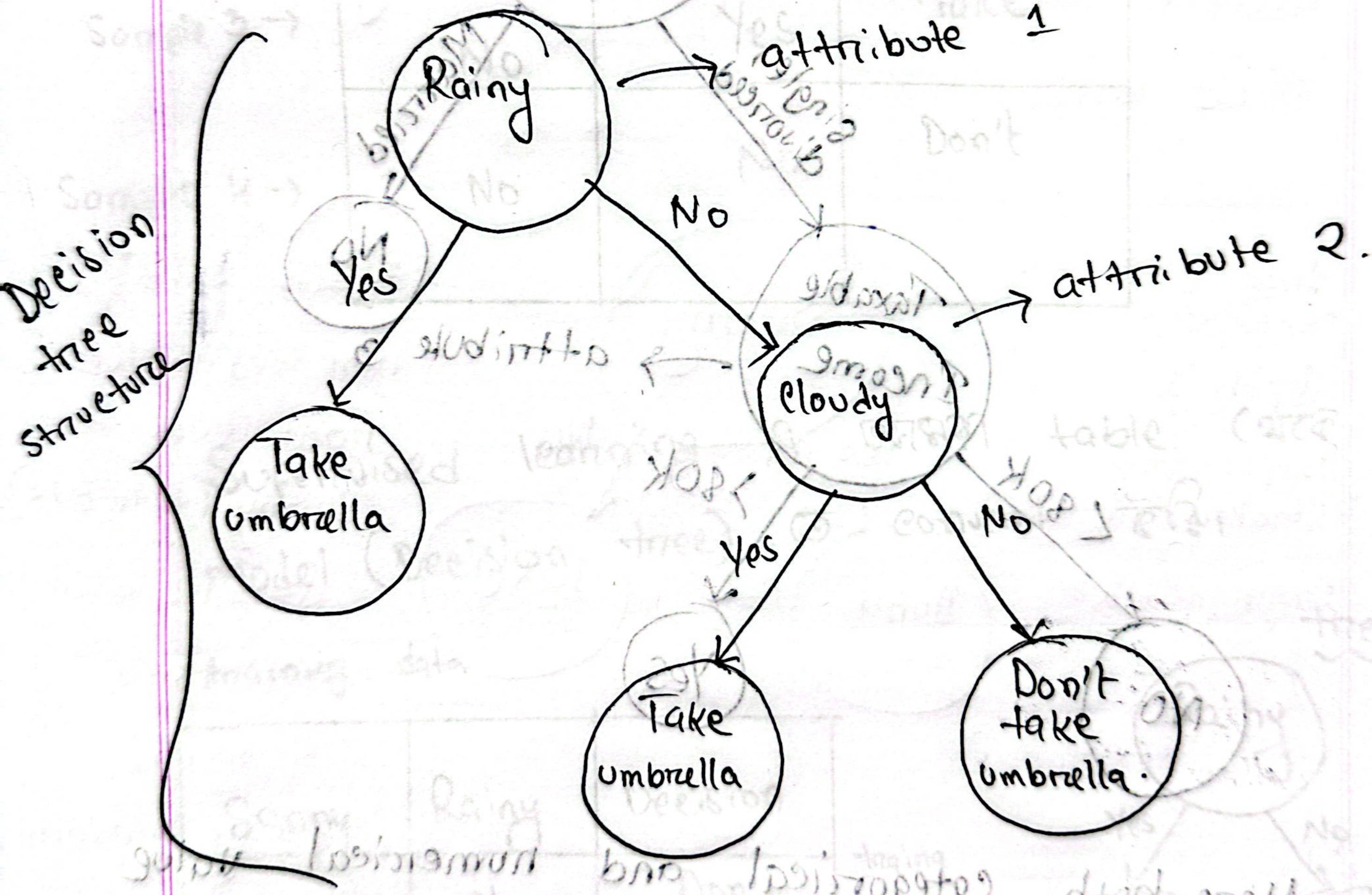
1 attribute

2 value

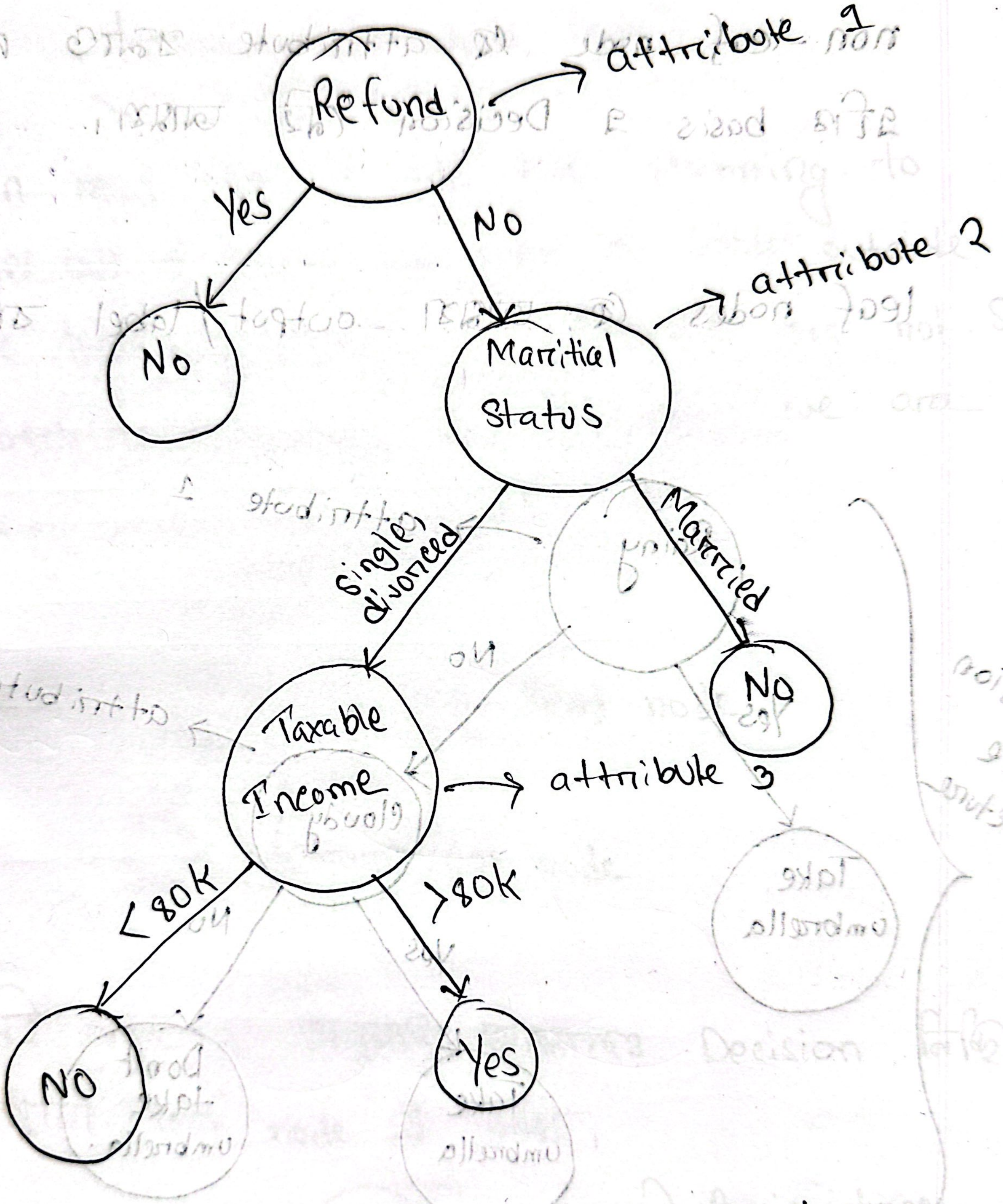
2 level on basis of value.

non-leaf node কে attribute বলতে পারি as এটির basis এ Decision নেই আমরা।

leaf nodes কে আমরা output / label বলি।



Yes, No → categorical value.



Here, both categorical and numerical value are present. Categorical data is converted to numerical.

1st tree into tabular form

feature

	Rainy	Cloudy	Decision
Sample 1 →	Yes	Yes	Take
Sample 2 →	Yes	No	Take
Sample 3 →	No	Yes	Take
Sample 4 →	No	No	Don't

~~feature~~ label

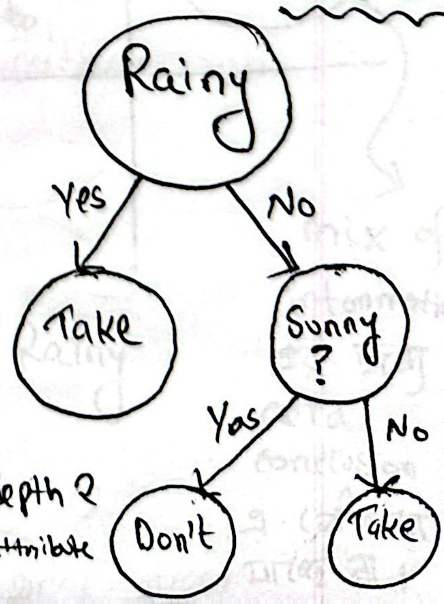
Supervised learning এ আমরা table থেকে model (Decision tree) এ convert করি।

training data

Sunny	Rainy	Decision
Yes	No	Don't
No	Yes	Take
No	No	Take

traing → phase (learning phase)

tree 1



depth 2
2 attribute

test data

Yes	Yes	? take

take → Decision Tree থেকে দেখে।

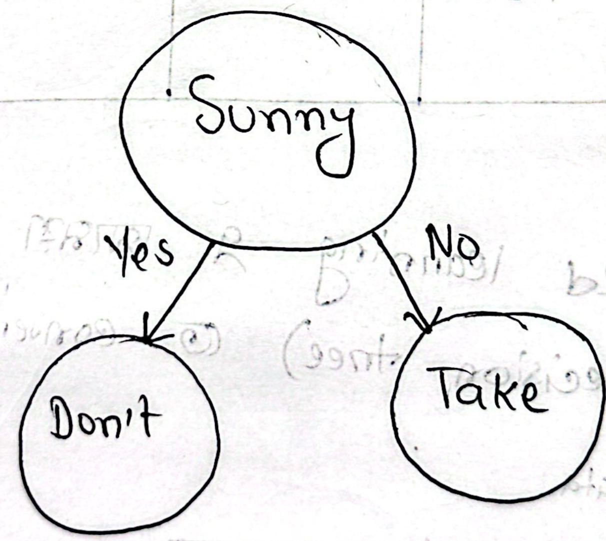
Decision tree আমরা supervised learning এ কাজে করতে পারি।

tree 2

depth 1

comparatively smaller tree

single attribute



যখন আমরা, non leaf node/ root node এ Sunny (attribute) নিয়েছি তখন পরে গিয়ে আমাদের raining consider ই করা লাগতেছে না।

আমরা concise decision tree form করার try করবো। { কিছু criteria }।

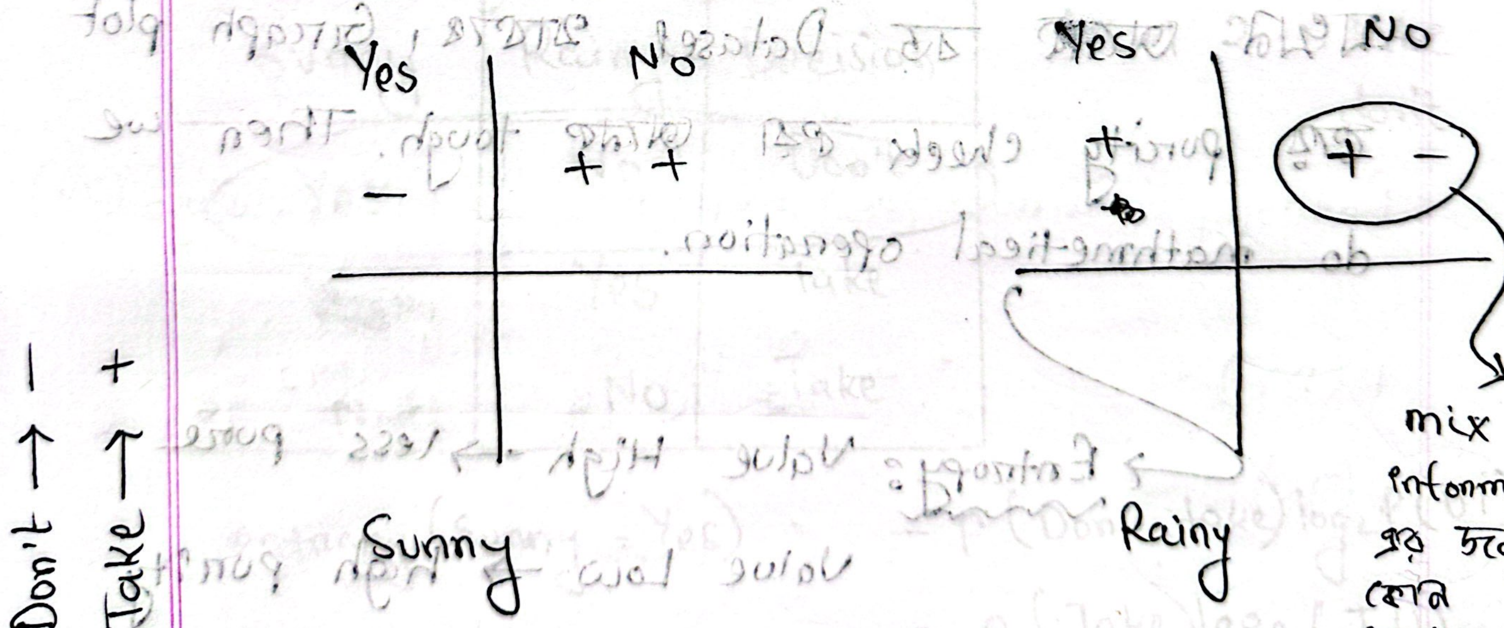
কথা হচ্ছে কোন criteria decision tree কে concise করে।

Root node Sunny হলে Decision tree concise.

Root node Rainy হলে Decision tree long.

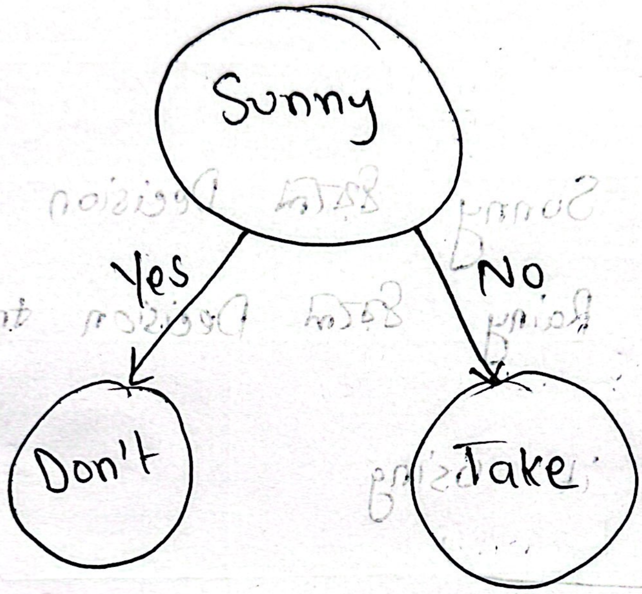
We describe it using

Information Purity :



mix of information এর জন্য কোন conclusion এ পৌঁছানো যাচ্ছে না।

∴ Sunny is more pure than Rainy. No mix of information. And it has more information. And, Sunny should have more priority.



As, purity Sunny এর বেশি, so এটি দিয়ে graph (decision tree) বানালে concise.

যখন অনেক বড় Dataset থাকবে, Graph plot করে purity check করা অনেক tough. Then we do mathmetical operation.

→ Entropy: Value High → less pure
Value Low → high purity

0 ←→ 10
more pure — entropy — less pure

Entropy = $\sum_{i=n}^{n} -p_i \log p_i$

i = number of labels.

entropy attribute এর each unique value নিয়ে বের করি।

↳ entropy (Sunny = Yes)

↳ entropy (Sunny = No)

↳ entropy (Rainy = Yes)

↳ entropy (Rainy = No)

Sunny	Rainy	Decision
Yes	No	Don't
No	Yes	Take
No	No	Take

label 2 → take, don't

$$\text{entropy}(\text{Sunny} = \text{Yes}) = -p(\text{Don't take}) \log_2 p(\text{DT}) - p(\text{Take}) \log_2 (\text{TU})$$

$$= -1/1 \log_2 1/1 - 0/1 \log_2 0/1$$

$= -\log_2 1$

$= 0$. (very pure).

Entropy (Sunny = NO) $= -P(DT)\log P(DT) - P(TU)\log_2(TU)$

$= -\frac{0}{2}\log\frac{0}{2} - \frac{2}{2}\log_2\frac{2}{2}$

$P(DT) = 0/2$

$P(TU) = 2/2$

$= -\log_2 1$

$= 0$ (very pure).

Entropy (Rainy = Yes) $= -P(DT)\log P(DT) - P(TU)\log_2(TU)$

$P(DT) =$

$= -\frac{0}{1}\log\frac{0}{1} - \frac{1}{1}\log_2\frac{1}{1}$

$P(TU)$

$= 0$ (very pure).

$$\text{Entropy}(\text{Rainy} = \text{No}) = -p(DT)\log_2 P(DT) - p(TU)\log_2 (TU)$$

$$p(DT) = \quad = -\frac{1}{2}\log_2 \frac{1}{2} - \frac{1}{2}\log_2 \frac{1}{2}$$

$$= -0.5\log_2 0.5 - 0.5\log_2 0.5$$

$$= 1$$ (High entropy = less pure).

$$\text{Entropy}(\text{Sunny}) < \text{Entropy}(\text{Rainy})$$

Sunny is more pure.

Video-17

Entropy → Information Gain → Select feature

Decision tree algorithm → ID3

Day	Outlook	Temp	Humidity	Wind	Decision
1	Sunny	Hot	High	weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	weak	Yes
4	Rain	Mild	High	weak	Yes
5	Rain	cool	Normal	weak	Yes
6	Rain	cool	Normal	Strong	No
7	Overcast	cool	Normal	Strong	Yes
8	Sunny	Mild	High	weak	No
9	Sunny	cool	Normal	weak	Yes
10	Rain	Mild	Normal	weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	weak	Yes
14	Rain	Mild	High	Strong	No

↳ Decision tree → root node (anything).

↳ → but কোনটা আসলে বসাবো,

Entropy → Information Gain → feature selection.

Information Gain (Outlook)

Entropy of Decision $= -p(yes)\log_2(yes) - p(no)\log_2(no)$

$$= -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14}$$

$$= .940$$

Entropy of (Sunny) $= -p(yes \backslash sunny)\log_2 (yes \backslash sunny) -$

$- p(No \backslash sunny)\log_2(No \backslash sunny)$

$$= -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}$$

$$= .971.$$

$$Entropy(Rain) = -p(yes \backslash rain)\log_2(Yes\backslash Rain) - p(No\backslash rain)\log_2(No\backslash Rain)$$

$$= -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5}$$

$$= .971.$$

$$Entropy(Overcast) = -p(Yes/overcast)\log_2(Yes/overcast) - p(No/overcast)\log_2(No/overcast)$$

$$= -\frac{4}{4}\log_2\frac{4}{4} - \frac{0}{4}\log_2\frac{0}{4}$$

$$= 0.$$

Information Gain(Outlook) = E(Decision) - E(Sunny) - E(Rain) - E(overcast).

এখানে Sunny and Rain 5 বার appear করেছে overcast appear করেছে 4 বার।

So, information gain calculate করার সময়, যেই particular category একটু কম appear করবে তাকে একটু কম weight দেয়া and বেশি যারা appear করবে তাদেরকে বেশি weight দেওয়া।

So,

$$IG(\text{outlook}) = E(\text{Decision}) - P(\text{sunny}) * E(\text{sunny}) - P(\text{Rain}) * E(\text{Rain}) - P(\text{overcast}) * E(\text{overcast}).$$

$$= .940 - \left(\frac{5}{14} \times .971\right) - \left(\frac{5}{14} \times .971\right) - \left(\frac{4}{14} \times 0\right)$$

Information gain = .246 ⟶ so outlook থেকে আমরা এই amount f information পাবো।

So, Temp, Humidity and wind এর জন্যও same ভাবে information gain বের করতে হবে and then যার gain বেশি তাকে root node বানিয়ে decision tree বানাতে হবে।

Entropy (Homidity = High):

$$= -\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7}$$

$= .985$

Entropy (Homidity = Normal)

$$= -\frac{6}{7}\log_2\frac{6}{7} - \frac{1}{7}\log_2\frac{1}{7}$$

$= .592$.

Information gain (Homidity) = E(Decision) − P(High) * E(High) − P(Normal) * E(Normal)

$$= .940 - \left(\frac{7}{14}\times .985\right) - \left(\frac{7}{14} * .592\right)$$

Information Gain (Humidity) = .151

$\Rightarrow IG(\text{Outlook}) = .246$

$\Rightarrow IG(\text{Humidity}) = .151$

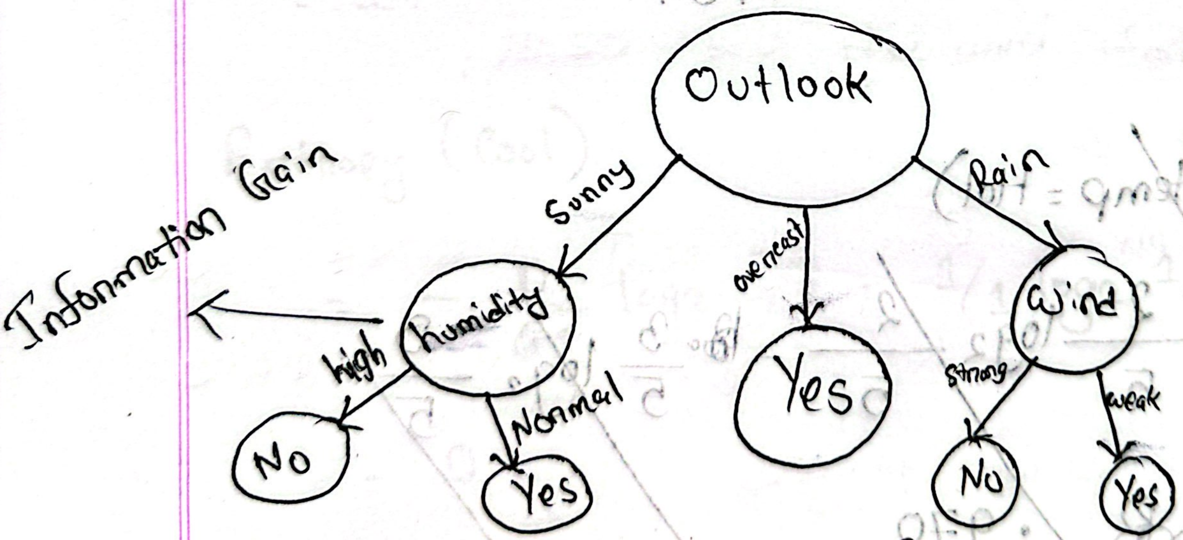
আমাদের যদি Outlook and Humidity থেকে select করতে হয় যে কোনটি root node নিবো তাহলে Outlook নিবো as, এখান থেকে বেশি information পাই।

$\Rightarrow IG(\text{Wind}) = .048$

$\Rightarrow IG(\text{Temp}) = .029$

More information Gain = more priority

Outlook has the most information gain



আমরা outlook থেকে overcast পেলে Yes দিছি, Rain & Sunny পেলে কিছু দিতে পারতিছি না as mix info. So,

আমরা table থেকে sunny Row আলাদা করলাম then,

Day	Outlook	Temp	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	mild	High	Weak	Yes
9	Sunny	cool	Normal	Weak	Yes
11	Sunny	mild	Normal	Strong	Yes.

$$\text{Entropy (Decision)} = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5}$$

$$= .970.$$

Entropy (temp = Hot)

$$= -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}$$

$$= .970$$

Entropy (Hot)

$= -p(no/hot)\log_2 p(no/hot) - p(yes/hot)\log_2 p(yes/hot)$

$= -2/2 \log 2/2 - 0/2 \log_2 0/2$

$= 0.$

Entropy (Mild)

$= -\frac{1}{2}\log_2 \frac{1}{2} - 1/2 \log_2 1/2$

$= 1.$

Entropy (Cool)

$= -\frac{0}{1}\log_2 0/1 - 1/1 \log_2 1/1$

$= 0.$

$IG(Temp) = E(D) - P(Hot) * E(Hot) - P(Mild) * E(Mild)$
$- P(Cool) * E(Cool)$

$= .970 - (2/5 * 0) - (2/5 * 1) - (1/5 * 0)$

= .570.

∴ Information gain (Temp) = .570

Same ভাবে Humidity and wind এর জন্য করতে হবে।

IG (Humidity) = .970

IG (Wind) = .019

So, আমরা humidity use করবো।

* last Graph a Rain এর জন্য আমাদের table থেকে Rain এর Row গুলো isolate করতে হবে, then Same process.

$$\text{Entropy (Decision)} = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}$$

$$= .970.$$

Entropy (~~Rain~~ ~~Temp~~ Wind) =

Video 14

Naive Bayes

it is a classifier. Basic but effective.

it mainly came from Bayes. Theorem.

Bayes Theorem

from probability theory we saw,

$$P(A \backslash B) = \frac{P(A \cap B)}{P(B)} \qquad P(B \backslash A) = \frac{P(B \cap A)}{P(A)}$$

$$\Rightarrow P(A \backslash B) P(B) = P(A \cap B) \qquad \Rightarrow P(B \backslash A) P(A) = P(B \cap A)$$

Now,

$P(A \cap B) = P(B \cap A)$ [are equal]

So, $P(A \backslash B) P(B) = P(B \backslash A) P(A)$

formula

$$\boxed{\therefore P(A \backslash B) = \frac{P(B \backslash A) P(A)}{P(B)}}$$

→ Bayes Theorem.

bayes theorem অনুযায়ী কোন particular problem space এর A and B এর marginal probability জানলে আমরা P(A|B) বের করতে পারবো.

The Problems that we can address using Bayes theorem

1) HIV global prevalence = .008

Test with 95% Specificity and Sensitivity

$p(T \mid HIV) = 95\%$ → থাকলে and True (sensitivity)

$p(\neg T \mid \neg HIV) = 95\%$ → না থাকলে and true (specificity)

Perform a ~~fi/st~~ test and the result is positive.

~~[illegible]~~

~~[illegible]~~ Person is likely HIV effected? or not?

$p(HIV) = .008$ $\quad\therefore P(\neg HIV) = 1 - .008$

$\qquad\qquad\qquad\qquad\qquad\qquad = .992.$

$p(HIV) + p(\neg HIV) = 1$

$P(T|HIV) = \cancel{\dfrac{P(\cdot 95 + \cdot 008)}{P(0.008)}}$

$\cdot 95$

$P(\neg T / HIV) = 1 - \cdot 95$

$= \cdot 05$

$P(\neg T / \neg HIV) = \cdot 95$

$P(T / \neg HIV) = 1 - \cdot 95$

$= \cdot 05$

Given true,

$P(HIV/T)$

$P(\neg HIV / T)$

compare.
যার value বেশি তাকে
more likely হবে ধরবো।

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

$$P(HN/True) = \frac{P(True|HN)\,P(HN)}{P(True)}$$

compare

$$P(\neg HN/True) = \frac{P(True/\neg HN)\,P(\neg HN)}{P(True)}$$

=

As, denominator same and both are positive, it won't effect our comparision. So, we don't consider them.

$$\underbrace{P(T|HN)\,P(HN)}_{\text{numeric}} \rightleftharpoons \underbrace{P(T/\neg HN)\,P(\neg HN)}_{\text{numeric}}$$

$L.H.S = P(T/HIV)\,P(HIV)$

$= .95 * .008$

$= .0076$

$R.H.S = P(T/\neg HIV)\,P(\neg HIV)$

$= .95 * .992$

$= .0496$

$R.H.S > L.H.S$

$P(\neg HIV/T) > P(HIV/T)$

Person is less likely to have HIV.

Done Comparison.

(c

$$P(HIV/T) = ?$$

$$P(HIV/T) = \frac{P(T/HIV)\,P(HIV)}{P(T)}$$

$$= \frac{P(T/HIV)\,P(HIV)}{P(A \cap HIV) + P(T \cap \neg HIV)}$$

$$= \frac{P(F/HIV)\,P(HIV)}{P(T/HIV)\cdot P(HIV) + P(T/\neg HIV)\,P(\neg HIV)}$$

$$= \frac{.95 \times .008}{.95 \times .008 + .05 \times .992}$$

$$= \frac{.0076}{.0572}$$

$$= .1329$$

(Ans).