F1 score = 2 X

only Accuracy can be misleading.

Covid 19

Actual	predicted
0	0
0	0.
1	0
0	0
0	0
0	0
1	0
0	0
0	0
0	0

MT2 syllabus
A* search
Gameplaying
Regression
Evaluation metrics
upto this syllabus

TP 0	FN 2
FP 0	8 TN

Accuracy = $\frac{0+8}{10} = 0.8$ 80%

But here it is not valid.

Recall = $\frac{0}{0+2}$ undefined Mathematically not correct = 0.

precision = $\frac{0}{0}$ = 0

Classification Algorithm.

1 Decision Tree.

Data:

Input, output

Day	outlook (sunny, overcast, Rainy)	Temperature (Hot, mild, cold)	Humidity (High, Normal)	wind	play Golf?
1	S	H	H	W	N
2	S	H	H	S	N
3	O	H	H	W	Y
4	R	M	H	W	Y
5	R	C	N	W	Y
6	R	C	N	S	N
7	O	C	N	S	Y
8	S	M	H	W	N
9	S	C	N	W	Y
10	R	M	N	W	Y
11	S	M	N	S	Y
12	O	M	H	S	Y
13	O	H	N	W	Y
14	R	M	H	S	N

After MT2
17/11/2023
Friday

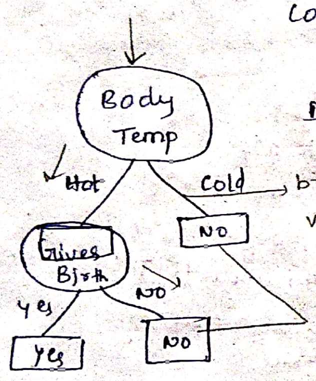
Classification:

Decision Tree:

Outlook	Temp	Humidity	Wind	Play Golf?

I/P
{overcast, mild, Normal, weak} → $\boxed{\text{model DT}}$ → Yes/NO?

Tree:	Body Temp	Gives Birth	Mammal?
	Hot		yes
	Cold		NO



cold → branches values of X

y- values.

If input {Hot, NO} → $\boxed{\text{Model}}$ → NO.

Entropy:- measure of randomness.

more Random ⇒ less certainty ⇒ more entropy.

Less Randomness ⇒ more certainty ⇒ Less entropy.

eg:-

		Randomness	Entropy
{0.5, 0.5}	→ less certainty →	High	High
{0.9, 0.1}			
{1.0, 0.0}	→	Low	Low

↳ We have total certainty, NO will win

From previous table Y, N play golf?

$\{\frac{9}{14}, \frac{5}{14}\}$
Yes's Prob, No's Prob.

Entropy of a distribution $= -\sum_{i=1}^{k} p(x_i)\log_2 p(x_i)$

$\{0.5, 0.5\} = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)$

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

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

$= -\underbrace{\log_2 1}_{0} + \underbrace{\log_2 2}_{1}$

$= 1$

$\{1.0, 0.0\} = -\left(1\cdot\log_2 1 + 0\cdot\log_2 0\right)$

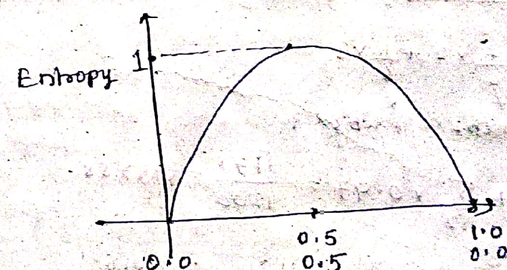
$= 0$

$\{0.9, 0.1\} = -\left(\frac{9}{10}\log_2\frac{9}{10} + \frac{1}{10}\log_2\frac{1}{10}\right)$

$= 0.03118$

$= +ve.$

Entropy plot



0.0 / 1.0	0.5 / 0.5	1.0 / 0.0
least Randomness, Least Entropy	Highest Randomness, Highest Entropy	Least Randomness, Least Entropy.

Using Entropy to select attribute.

A	B	class
Y	Y	C_1
N	N	C_2

$C_1: 6$
$C_2: 6$

A: Y → C1: 4, C2: 3; N → 2, 3

B: Y → 1, 4; N → 5, 2

Entropy for A

6 6

$-\frac{6}{12}\log_2\frac{6}{12} - \frac{6}{12}\log_2\frac{6}{12}$

$= 1$.

$\{\frac{4}{7}, \frac{3}{7}\}$ $\{\frac{1}{5}, \frac{4}{5}\}$

↓

more certainty

for children entropy calculation

A

C1: 4, C2: 3 — C1: 2, C2: 3

B

C1: 1, C2: 4 — C1: 5, C2: 2

$-\frac{4}{7}\log\frac{4}{7} - \frac{3}{7}\log\frac{3}{7}$ $= 0.98$

$-\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5}$ $= 0.97$

$-\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5}$ $= 0.72$

$-\frac{5}{7}\log_2\frac{5}{7} - \frac{2}{7}\log_2\frac{2}{7}$ $= 0.86$

$\cancel{\frac{3}{7}}$

Average entropy or Expectation,

1, 1, 1, 2, 2, 3

$= \frac{1+1+1+2+2+3}{6}$

$= \frac{3\times1 + 2\times2 + 1\times3}{6}$

$= \frac{3}{6}\times 1 + \frac{2}{6}\times 2 + \frac{1}{6}\times 3$

prob of 1 × 1 value, prob 2 × 2, prob 3 × 3

Assignment 2:
MT1 & MT2
Answer
Next Friday

$\frac{7}{12}\times 0.98 + \frac{5}{12}\times 0.97 = \frac{1171}{1200} = 0.9758$

$\frac{5}{12}\times 0.72 + \frac{7}{12}\times 0.86 = \frac{481}{600} = 0.8016$

Gain(A)
$= 1.0 - 0.975$
$= 0.025$

Gain(B)
$= 1.0 - 0.801$
$= 0.199$ (reduce randomness so B is selected because it is reducing the randomness).

Decision Tree: 23/11/2023 wednesday

outlook	Temp	Humidity	wind	play Golf!

Gain = Entropy (parent) − Average Entropy of children

Y: 9, N: 5 } parent

Outlook: S → Y: 2, N: 3; O → 4, 0; R → 3, 2

Entropy $= -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$

for children.

$-\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.97$.

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

$-\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.97$

Average children Entropy $= \frac{5}{14}\times 0.97 + \frac{4}{14}\times 0 + \frac{5}{14}\times 0.97 = 0.69$

parent $\boxed{\text{Temp}}$ same

Temp: H → Y: 2, N: 2; m → 4, 2; C → 3, 1

$H \Rightarrow \frac{2}{2} \Rightarrow 1$

$m \Rightarrow -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6} = 0.92$

$C \Rightarrow -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} = 0.81$

Average children Entropy $= \frac{4}{14}\times 1 + \frac{6}{14}\times 0.92 + \frac{4}{14}\times 0.81 = 0.91$

Humidity: H → Y: 3, N: 4; N → 6, 1

$H \Rightarrow -\frac{3}{4}\log_2\frac{3}{4} - \frac{4}{3}\log_2\frac{4}{3} = 0.98$

$N \Rightarrow -\frac{6}{7}\log_2\frac{6}{7} - \frac{1}{7}\log_2\frac{1}{7} = 0.59$

Average children Entropy $= \frac{7}{14}\times 0.98 + \frac{7}{14}\times 0.98$
$= 0.79$

Wind: W → Y: 6, N: 2; S → 3, 3

$W \Rightarrow -\frac{6}{8}\log_2\frac{6}{8} - \frac{2}{8}\log_2\frac{2}{8} = 0.81$

$S \Rightarrow -\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6} = 1$

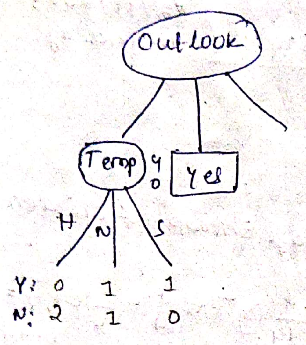
$= \frac{8}{14}\times 0.81 + \frac{6}{14}\times 1 = 0.89$.

Gain(outlook) = 0.94 − 0.69 = 0.25 ✓ select as root node

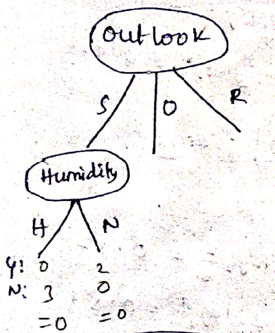
Gain(Temp) = 0.94 − 0.91 = 0.03

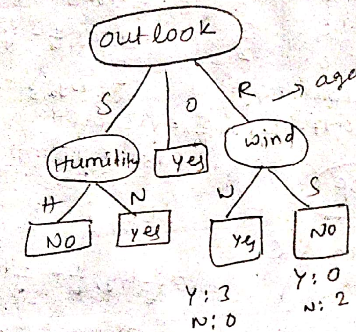
Gain(Humidity) = 0.94 − 0.79 = 0.15

Gain(wind) = 0.94 − 0.89 = 0.05.

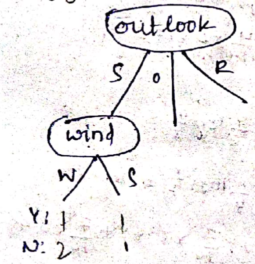


final model.





check again 3 Temp, wind, humidity



eg:- Day = {sunny, High, Normal, weak}

o/p:- yes

H S → N of Humidity

day = {Rainy, cold, high, strong}

o/p: NO.

R → S (of wind).

Stopping criteria:

① All examples belongs to same class.

Y: 0 N: 7 ⇒ $\boxed{\text{NO}}$

② Reached maximum depth. assign majority label.

Y: 9 N: 1 ⇒ $\boxed{\text{Yes}}$

Bayesian classifier:

↳ Naive Bayes classification.

Probability – Marginal
– Joint
– conditional.

eg:- Bank employee.

Gen	Rank R1	R2	R3	Total
Male	30	80	90	200
female	20	40	40	100
total	50	120	130	300

Marginal: $P(\text{Gen} = \text{male}) = \frac{200}{300}$

$P(\text{Rank} = R_2) = \frac{120}{300}$

Joint: $P(\text{Gen} = \text{male and Rank} = R_2) = \frac{80}{300}$

Marginal using Joint: $P(\text{Gen} = \text{male}) =$

$P(\text{Gen} = M \text{ and Rank} = R_1) + P(\text{Gen} = M \text{ and Rank} = R_2) + P(\text{Gen} = M \text{ and Rank} = R_3)$.

$= \frac{30}{300} + \frac{80}{300} + \frac{90}{300}$

$= \frac{200}{300}$

conditional: $P(\text{Rank} = R_2 \mid \text{Gen} = M) = \frac{80}{200}$

can be expressing using Joint and probability.

$= \frac{P(\text{Rank} = R_2 \text{ and Gen} = m)}{P(\text{Gen} = m)}$

$= \frac{\frac{80}{300}}{\frac{200}{300}} = \frac{80}{200}$

Formulae:

$$P(Y/X) = \frac{P(X \text{ and } Y)}{P(X)}$$

$$P(X \text{ and } Y) = P(Y/X)\,P(X)$$

$$P(X/Y) = \frac{P(X \text{ and } Y)}{P(Y)}$$

$$P(X \text{ and } Y) = P(X|Y)\,P(Y)$$

$$P(Y|X)\cdot P(X) = P(X|Y)\,P(Y)$$

$$\boxed{P(Y|X) = \frac{P(X|Y)\,P(Y)}{P(X)}}$$ Bayes theorem.

eg:- Day = {sunny, cold, High, weak} (x_1, x_2, x_3, x_4)

$P(\text{Play}=\text{yes} \mid \text{Day})$

$P(\text{play}=\text{No} \mid \text{Day})$.

$$P(Y=\text{yes} \mid \text{Day}) = \frac{P(\text{Day} \mid \text{yes})\,P(\text{yes})}{P(\text{Day})}$$

24/11/2023 Friday.

Bayesian classification

- Naive Bayes classifier.
- Marginal – $P(X)$
- Joint – $P(X \text{ and } Y)$
- conditional – $P(Y|X)$

X = {Rainy, Cold, High, Strong}

$P(Y=\text{yes} \mid X) = ?$

$P(Y=\text{No} \mid X) = ?$

$$P(X|Y) = \frac{P(Y \text{ and } X)}{P(Y)} \qquad P(Y \text{ and } X) = P(X|Y)P(Y)$$

From $P(Y=\text{yes}|X) = \dfrac{P(Y=\text{yes and } X)}{P(X)}$

$$= \frac{P(X|Y=\text{yes})\,P(Y=\text{yes})}{P(X)}$$

$$= \frac{P(X|Y=\text{yes})\,P(Y=\text{yes})}{P(X \text{ and } Y=\text{yes}) + P(X \text{ and } Y=\text{NO})}$$

$$= \frac{P(X|Y=\text{yes})\,P(Y=\text{yes})}{P(X|Y=\text{yes})\,P(Y=\text{yes}) + P(X|Y=\text{NO})\,P(Y=\text{NO})}$$

Bayes! continuation

This is Bayes's thm.

To simplify, we use Naive Bayes Assumption: Input variables are independent given class label. something you are oversimplifying

If independent variables

$P(x_1 \cdot x_2) = P(x_1)\cdot P(x_2)$.

$P(X=\{x_1, x_2, x_3, x_4\} \mid Y=\text{yes})$

$= P(X=x_1|Y=\text{yes})\cdot P(X=x_2|Y=\text{yes})\cdot P(X=x_3|Y=\text{yes})\cdot P(X=x_4|Y=\text{yes})$

Assumption: Input variables does not depend on each others, they are independent.

Rewrite using Naive Bayes Assumption.

↳ continuation.

$$= \frac{P(X=x_1|Y=\text{yes})\cdot P(X=x_2|Y=\text{yes})\cdot P(X=x_3|Y=\text{yes})\cdot P(X=x_4|Y=\text{yes})\,P(Y=\text{yes})}{P(X=x_1|Y=\text{yes})\cdot P(X=x_2|Y=\text{yes})\cdot P(X=x_3|Y=\text{yes})\cdot P(X=x_4|Y=\text{yes})\,P(Y=\text{yes}) + P(X=x_1|Y=\text{NO})\cdot P(X=x_2|Y=\text{NO})\cdot P(X=x_3|Y=\text{NO})\cdot P(X=x_4|Y=\text{NO})\,P(Y=\text{NO})}$$

This is Naive Bayes theorem.

from X = {Rainy, cold, High, Strong}

$P(Y=\text{yes} \mid X=\{\text{Rainy, cold, High, strong}\})$

$$= \frac{P(\text{Rainy}|\text{yes})\cdot P(\text{cold}|\text{yes})\cdot P(\text{High}|\text{yes})\cdot P(\text{Strong}|\text{yes})}{P(\text{Rainy}|\text{yes})\cdot P(\text{cold}|\text{yes})\cdot P(\text{High}|\text{yes})\cdot P(\text{Strong}|\text{yes}) + P(\text{Rainy}|\text{No})\cdot P(\text{Cold}|\text{No})\cdot P(\text{High}|\text{No})\cdot P(\text{Strong}|\text{No})}$$

$P(\text{yes}) = \frac{9}{14}$

$P(\text{No}) = \frac{5}{14}$

P(outlook | Y)

	yes	No	P(·|yes)	P(·|No)
Sunny	2	3	2/9	3/5
overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5

P(temp | Y)

	Yes	No	P(·|yes)	P(·|No)
Hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cold	3	1	3/9	1/5

P(Humidity | Y)

	Yes	No	P(·|Yes)	P(·|No)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5

P(wind | Y)

	Yes	No	P(·|yes)	P(·|No)
Strong	3	3	3/9	3/5
weak	6	2	6/9	2/5

$$= \frac{\frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}}{\frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14} + \frac{2}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14}}$$

$= 0.37$

$P(Y=No|X) = 1 - P(Y=yes|X) = 1 - 0.37$
$= 0.63$

output: $P(Y=No|X) > P(Y=yes|X)$

"No" → will not play the Golf.

Email classification.

Spam

Not spam.

content (↱ set of words)	label
$w_1 w_2 w_3 w_4$	No → not spam
$w_3 w_5 w_6 w_7$	yes

New email = { w_1 w_2 w_7 w_9 }
= { you won Rs 10L }

90%. It will correctly implement
→ you need data for tables programs

= P(Yes | you won Rs 10L)

$$= \frac{P(you|yes) \cdot P(won|yes) \cdot P(Rs|yes) \cdots p(yes)}{P(you|yes) \cdots p(yes) + P(you|No) \cdots p(no)}$$

(2) $L = \{a^n b^m c^n \mid m, n \geq 1\}$ $\underline{aabcc}$

$L = \{abc, aabcc, aabbcc, aaabccc, \ldots\}$

$\delta(q_0, a, z_0) = (q_0, az_0)$

$\delta(q_0, a, a) = (q_0, aaz_0)$

$\delta(q_0, b, a) = (q_1, aaz_0)$

$\delta(q_1, c, a) = (q_2, \epsilon)$

$\delta(q_2, \epsilon, z_0) = (q_3, \epsilon)$

$(q_0, aabcc, z_0) \vdash (q_0, abcc, az_0)$

$\vdash (q_0, bcc, aaz_0)$

$\vdash (q_1, cc, aaz_0)$

$\vdash (q_2, c, az_0)$

$\vdash (q_2, \epsilon, z_0)$

Accepted $\boxed{\vdash' (q_3, \epsilon)}$

23/11/23

Pumping Lemma for Regular languages

Let 'L' be a regular language. There will be a Natural no. 'n' (pumping Lemma constant) for every $z \in L$ & $|z| \geq n$, z can be written as uvw satisfying

i) $|v| \geq 1$

ii) $|uv| \leq n$

iii) $uv^iw \in L$ for all $i \geq 0$.

⇒ Pumping Lemma is necessary but not sufficient condition for RLs.

⇒ So this can be used to prove a language as Non-regular if it doesn't satisfy pumping Lemma.

* If RL only pumping Lemma, we cannot say it is compulsary R.L

Pumping Lemma for Context free languages

Let 'L' be a Context free language. there will be a Natural 'n' (pumping Lemma constant) depending only on L, for every $z \in L$ & $|z| \leq n$, z can be written as uvwxy satisfying

i) $|vx| \geq 1$

ii) $|uvw| \leq n$

iii) $uv^iwx^iy \in L$ for all $i \geq 0$.

→ Pumping Lemma is necessary but not sufficient condition for CFLs.

⇒ So, this can be used to prove a language is NonCFL if it doesn't satisfy pumping lemma.

* Both 'v' and 'x' cannot be epsilon at a time.
* (or) any one can be Epsilon.

Equivalence b/w CFG & PDA

CFG ⇒ PDA

PDA ⇒ CFG

⇒ If CFG exists, then PDA will exists.

⇒ If PDA exists, then CFG will exists.

Conversion of CFG to PDA

⇒ Given a stack.

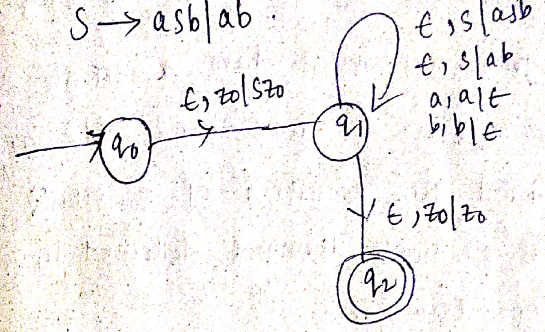
⇒ we push into stack

⇒ If top of the stack is Non-terminal then we (push) substitute the right hand side of the production

⇒ If top of the stack is terminal and if it matches the string the we pop.

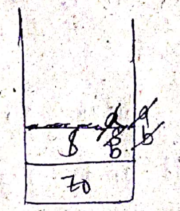
Eg: $L = \{a^m b^m \mid m \geq 1\}$

$S \rightarrow aSb \mid ab$



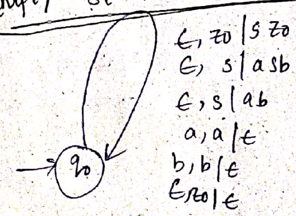
Let, aabb

$q_0\ q_1\ q_1\ q_1\ q_1\ q_2$



(or)

Empty stack acceptance



⇒ A CFG can be converted into PDA with only single state by empty stack acceptance.

(1) Type-3 Grammar:-

⇒ Type-3 grammars generate regular languages

⇒ Type-3 grammars must have a single non-Terminal on the left-hand side and a right-hand side consisting of a single terminal or single terminal followed by a single non-Terminal.

⇒ The productions must be in the form

$X \to a$
(or)
$X \to aY$

where $X, Y \in N$ (Non Terminal)

and $a \in T$ (Terminal)

∴ The rule $S \to \epsilon$ is allowed if S does not appear on the right side of any rule.

Eg:-
$X \to \epsilon$
$X \to a | aY$
$Y \to b$

(2) Type-2 Grammar:-

⇒ Type-2 Grammars generate Context-free languages.

⇒ The productions must be in the form $A \to \gamma$

where $A \in N$ (Non terminal)

and $\gamma \in (T \cup N)^*$ (String of terminal and non-terminals).

⇒ These languages generated by these grammars are be recognized by a non-deterministic pushdown automation.

Eg:-
$S \to Xa$
$X \to a$
$X \to aX$
$X \to abc$
$X \to \epsilon$

(3) Type-1 Grammar:-

⇒ Type-1 Grammars generate context-sensitive languages.

⇒ The productions must be in the form

$\alpha A \beta \to \alpha \gamma \beta$

where $A \in N$ (Non-Terminal)

and $\alpha, \beta, \gamma \in (T \cup N)^*$ (Strings of terminals and Non-terminals)

⇒ The strings α and β may be empty, but γ must be non-empty.

⇒ The rule $S \to \epsilon$ is allowed if S does not appear on the right side of any rule. The languages generated by these grammars are recognized by a linear bounded automation.

Eg:-
$AB \to AbBc$
$A \to bcA$
$B \to b$

(4) Type - 0 Grammars:-

⇒ Type-0 Grammars generate recursively enumerable languages.

⇒ The productions have no restrictions.

⇒ They are any phase structure grammar including all formal grammars.

⇒ They generate the languages that are recognized by a Turing Machine.

⇒ The productions can be in the form of $\alpha \to \beta$ where α is a string of terminals and non-terminals with atleast one non-terminal and α cannot be null. β is a string of terminals and non-terminals

Eg:-

$S \to ACaB$
$Bc \to aCB$
$CB \to DB$
$aD \to Db$

Scanned with OKEN Scanner