① Decision tree using ID3

Entropy of dataset =

positives = 9
negatives = 5
$\overline{14}$

$E(S) = -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right)$

$= 0.940$

Attributes

① Outlook

	Positives	Negatives	Entropy
Sunny	2	3	0.971
Rainy	3	2	0.971
Overcast	4	0	0

AVE $\Rightarrow$

I(outlook)

$= \frac{3}{2}$

$\frac{3+2}{9+5} \times 0.971$

$+$

$\frac{2+3}{9+5} \times 0.971$

$+$

0

$= 0.693$

I(outlook) $= 0.693$

Gain

Entropy (S) − I(outlook)

$= 0.940 - 0.693$

$= 0.247$ — (i)

Gain

(ii) I(temperature)

temp	positive	negative	E
hot	2	2	1
Mild	4	2	0.918
Cold	3	1	0.811

I(temp) =>

$$\left(\frac{2+2}{9+5} \times 1\right) + \left(\frac{4+2}{9+5} \times 0.918\right) + \left(\frac{3+1}{9+5} \times 0.811\right)$$

I(temp) => 0.911

Gain =>

$= 0.940 - 0.911$

$= 0.029$ — (ii)

(iii) Humidity

	positive	Negative	E
High	3	4	0.985
Normal	6	1	0.591

I(humidity)

$$\left(\frac{3+4}{9+5} + 0.985\right) + \left(\frac{6+1}{9+5} \times 0.591\right)$$

$= 0.788$

gain $\Rightarrow$ $E(S) - I(\text{humidity})$

$= \cancel{0.1}\ 0.940 - 0.788$

$= 0.152$ — (iii)

(iv) Windy

	positive	negatives	
true	3	3	1
false	6	2	0.811

$$\left(\frac{3+3}{9+5} \times 1\right) + \left(\frac{6+2}{9+5} \times 0.811\right)$$

$= 0.892$

gain

$\Rightarrow E(S) - I(\text{windy})$

$\Rightarrow 0.940 - 0.892$

$\Rightarrow 0.048$ — (iv)

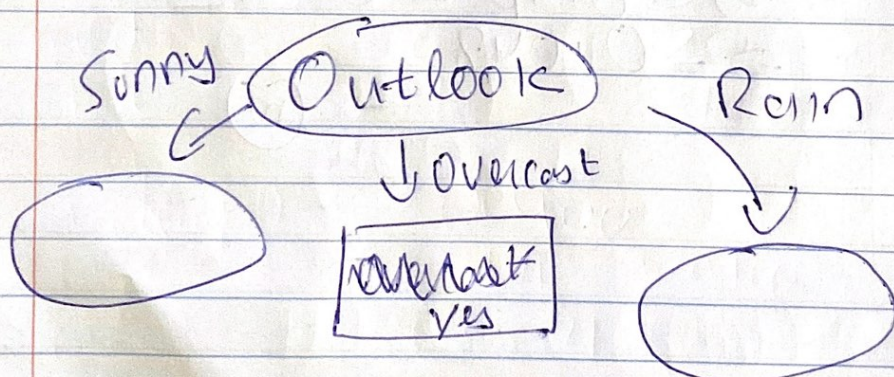
from (i), (ii), (iii) & (iv)

Gain

Attribute	gain
outlook	0.247
temp	0.029
humidity	0.152
Windy	0.048

from this, we will choose outlook as the base node

that gives us,



Under outlook Attribute

① For Sunny positive = 2
negative = 3
$\overline{5}$

$E(S) =$

$$\frac{-2}{2+3} \log_2\left(\frac{2}{2+3}\right) - \frac{3}{2+3} \log_2\left(\frac{3}{2+3}\right)$$

$\Rightarrow 0.971$

I (humidity) $\Rightarrow$

	positive	negative	Entropy
high	0	3	0
normal	2	0	0

I (humidity) = 0

gain = 0.971 − 0
= 0.971

~ ①

I(Wind)

	Positive	Negative	Entropy
True	1	1	1
False	1	2	0.918

I(Wind) = 0.951

gain = 0.971 − 0.951
= 0.020 — (ii)

I(Temp)

	positive	negative	E
cool	1	0	0
mild	0	2	0
hot	1	1	0

I(temp) = 0.400
= 0.971 — 0.400
= 0.571
— (iii)

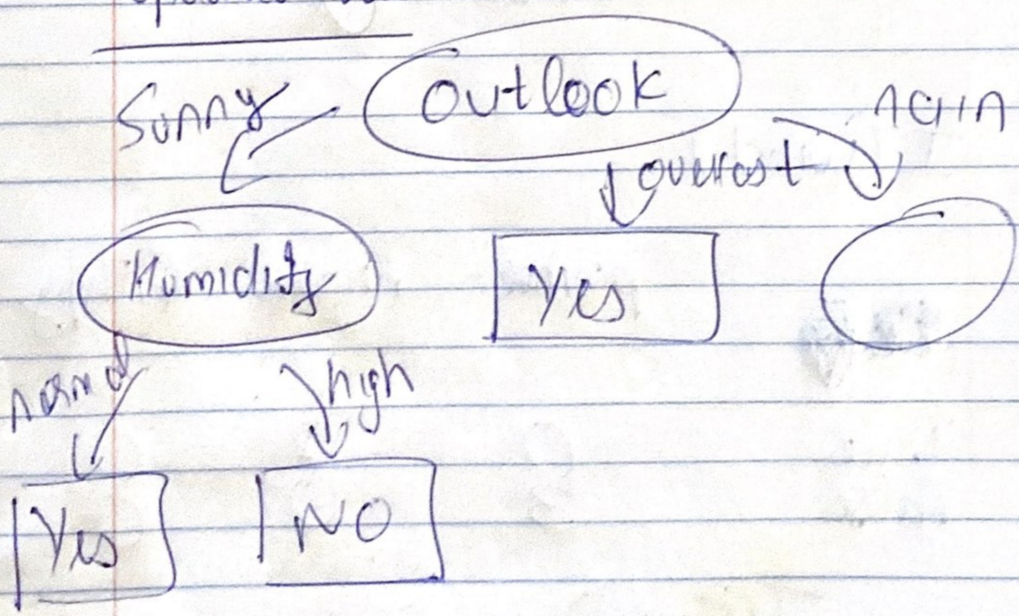
from (i) & (ii) & (iii)

Attribute	Gain
temp	0.571
humidity	0.971
Windy	0.02

humidity is highest so its the next node

~~udat~~

updated tree



11

Fouth Outlook Rainy

I(humidity)

	positive	negative	Entropy
high	1	1	1
normal	2	1	0.918

I(humidity) = 0.951

gain = 0.971 - 0.951
= 0.020 — 1

I(wind)

	positive	negative	Entropy
True	0	2	0
False	3	0	0

I(Windy) = 0
gain = 0.971

11

I(temp)

	positive	negative	Entropy
cool	1	1	1
mild	2	1	0.918

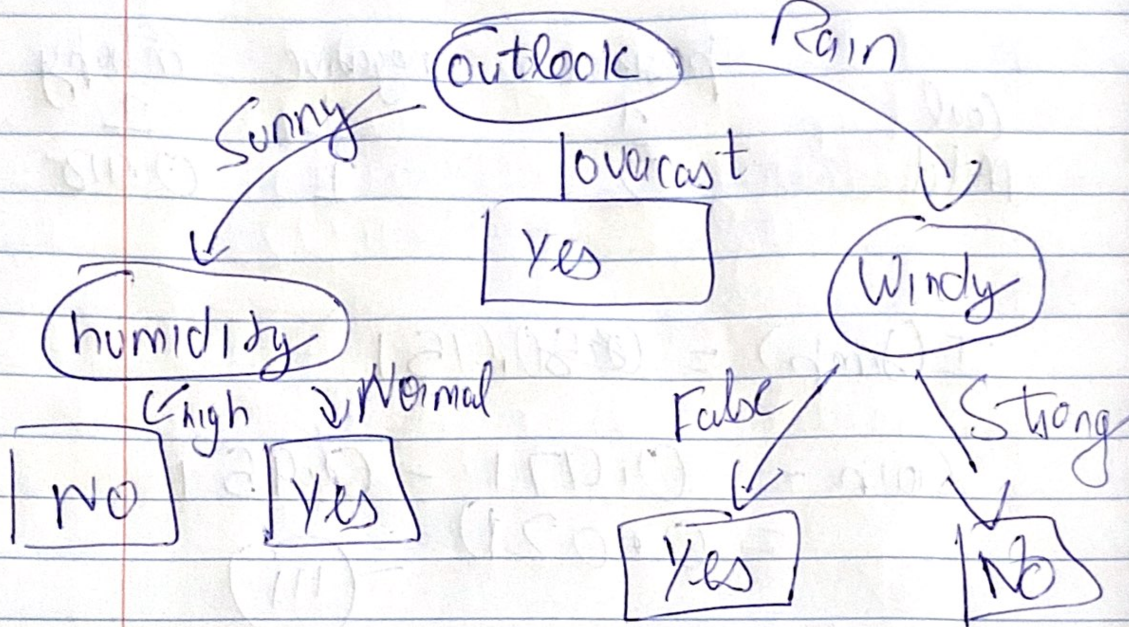
I(temp) = ~~0.5~~ 0.951

Gain = 0.971 − 0.951
= 0.020 — (III)

from (I), (II) & (III)

~~Attribute~~	Gain
humidity	0.020
wind	0.971 ✓
temp	0.020

updated tree ->



1.2 PRISM Algorithm

Two classes - "Yes" & "NO"

for play = yes

outlook	=	Sunny	2/5
outlook	=	Overcast	4/4 ① largest fraction
outlook	=	Rainy	3/5
Temp	=	hot	2/4
temp	=	mild	4/6
temp	=	cold	3/4
humidity	=	high	3/7
humidity	=	normal	6/7
Windy	=	true	3/6
Windy	=	false	6/8

from ①

outlook = overcast

For class = No

outlook = Sunny 3/5 ✓ (11) highest fraction

outlook = overcast 0/4

outlook = rainy 2/5

Temp = hot 2/4

Temp = mild 2/6

Temp = cold 1/4

humidity = high 4/7

humidity = normal 1/7

~~humidity = n~~

Windy = true 3/6

Windy = false 2/8

from (11)

outlook = Sunny

If outlook = overcast then class = Yes

outlook = Sunny then class = No

2

ID3 uses greedy approach so it is really fast, It constructs optimal decision trees for more than 50% of the datasets.

It uses Information gain to construct decision trees with less cost

Information gain heuristic makes ID3 fast and cheaper

~~Ex~~

Y

(a)

X_0	X_1	X_2	y (class)
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Step 1 –

(i) initializing weight &

(ii) learning rate

η ($0 < \eta < 1$)

Step 2 – repeat 3–5 untill convergence

Step 3 – Calculating fast net output

y in $\Sigma = b = \sum_{i=1} x_i w$

when ~~a~~ η is inputs ~~...~~

Step 4 Apply activation function

Step 5 update weight & Bias

Int/Das

$w_i = w_1 + \Delta w_1$

where

$\Delta w_i = \eta (t - o) x_i$

t = target output

o = output generated

η = learning rate

$y' = \{ \Sigma > 0 \; ; \; y' = 1 \text{ or } y = 0 \}$

$w_0 = -1 \quad w_1 = 0.4$

$w_2 = 0.5 \quad \eta = 0.6$

	x_0	x_1	x_2	y	w_0	w_1	w_2	Σ	y'	$\eta(y - y')$
1	1	0	0	0	-1	0.4	0.5	-1	0	0
2	1	0	1	1	-1	0.4	0.5	-0.5	0	0.6
3	1	1	0	1	-0.8	0.4	0.7	-0.4	0	0.6
4	1	1	1	1	-0.6	0.6	0.7	-0.7	1	0

for step ① & ④ there is no errors so we dont need to change weights bc we update ② & ③

for 2

$w_0 = -1 + 0.6 = 0.4$

$w_1 = 0.4 + 0(0.6) = 0.4$

$w_2 = 0.5 + 0.6 = 1.3$

for 3

$y' = 0, \quad y = 1$

$w_0 = -0.8 + 1(0.6) = -0.2$

$w_1 = 0.4 + (0.6)1 = 1.0$

$w_2 = 0.7 + 0(0.6) = 0.7$

x_0	x_1	x_2	y	w_0	w_1	w_2	Σ	y'	$\eta(y-y')$
1	0	0	0	-0.2	1	0.7	-0.2	0	0
1	0	1	1	-0.2	1	0.7	-0.5	0	0
1	1	0	1	-0.2	1	0.7	0.8	1	0
1	1	1	1	0.5	0.4	1.31	2.11	1	0

b $w_0 = -1$ $w_1 = 0$ $w_2 = 0.1$

$y' \begin{cases} \Sigma > 0 : y = 1 \text{ or } y = 0 \end{cases}$

$\eta = 0.2$

	x_0	x_1	x_2	y	w_0	w_1	w_2	Σ	y'	$\eta(y-y')$
1	1	0	0	0	-1	0	0.1	-1	0	0
2	1	0	1	0	-1	0	0.1	-0.9	0	0
3	1	1	0	0	-1	0	0.1	-1	0	0
4	1	1	1	1	-1	0	0.1	-0.9	0	0.9

$w_0 = -1 + 1(0.9) = -0.1$

$w_1 = 0 + 0.9 = 0.9$

$w_2 = 0.1 + 0.9 = 1.0$

	x_0	x_1	x_2	y	w_0	w_1	w_2	Σ	y'	$1(y-y')$
1	1	0	0	0	-0.1	0.9	1	-0.1	0	0
2	1	0	1	0	-0.1	0.9	1	-0.1+1	0	0
3	1	1	0	0	-0.1	0.9	1	-1	0	0
4	1	1	1	1	-0.1	0.9	1	1	1	0