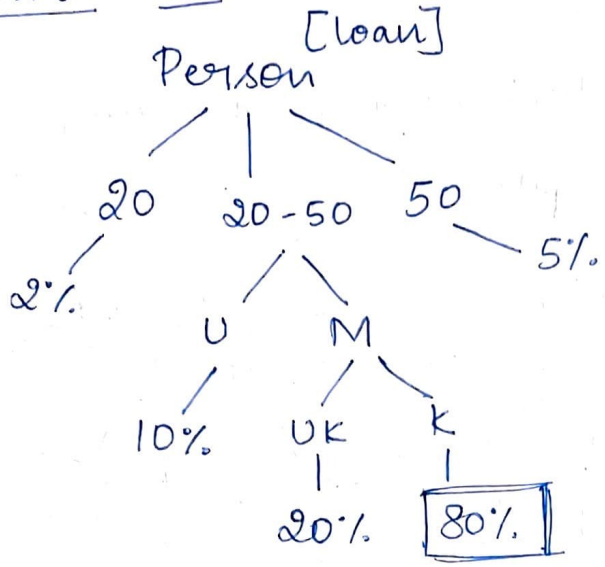
Decision Tree [Using information gain]



Age	Competition	Type	Profit
Old	Yes	S/W	Down
Old	No	S/W	Down
Old	No	H/W	Down
mid	Yes	S/W	Down
mid	Yes	H/W	Down
mid	No	H/W	Up
mid	No	S/W	Up
new	Yes	S/W	Up
new	No	H/W	Up
new	No	S/W	Up

P → posibility of Yes → N=5 [up]

N → No → N=5 [down]

Entropy of class:

$$= \frac{-P}{P+N} \log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N} \log_2\left(\frac{N}{P+N}\right)$$

$$= \frac{-5}{10} \log_2\left(\frac{5}{10}\right) - \frac{5}{10} \log_2\left(\frac{N}{P+N}\right)$$

$$= 1$$

To find the Root Node.

↳ Entropy, information gain, gain for each attribute

$$I(P_i, N_i) = \frac{-P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

$$\text{Entropy} = \frac{\sum P_i + N_i}{P+N}\left(I(P_i N_i)\right)$$

Gain

= Entropy class − Entropy attributes

Age	P_i	N_i	$I(P_i N_i)$
Old:	0	3	0
Mid	2	2	1
New	3	0	0

[P = 0 N = 3 . -- --] → Any value is zero the Information gain will be zero

→ if both value is same then it is 1

$$\text{Entropy}: \frac{0+3}{10}(0) + \frac{2+2}{5+5}(1) + \frac{3+0}{10}(0)$$

↳ class attribute

$$= \frac{4}{10} = 0.4$$

Gain: Class entropy − Entropy (Age)

$$= 1 - 0.4$$

$$= 0.6$$

Compitition

	Pi	Ni	I(Pi Ni)
Yes	1	3	0.81127
No	4	2	0.918295

$$\rightarrow I(P_i N_i) = \frac{-P}{P+N} \log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N} \log_2\left(\frac{P}{P+N}\right)$$

P=1, N=3

$$= \frac{-1}{4} \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right)$$

$$= 0.81127$$

P=4, N=2

$$= \frac{-4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right)$$

$$= 0.918295$$

$$\rightarrow \text{Entropy} \quad \frac{\Sigma P_i + N_i}{P+N} \left(I(P_i N_i)\right)$$

$$= \frac{1+3}{5+5}(0.81127) + \frac{4+2}{5+5}(0.918295)$$

$$= 0.8754$$

Gain: Class entropy − Entropy (Competition)

$$= 1 - 0.8754$$

$$= 0.124515$$

Type

	Pi	Ni	Information gain
S/W	3	3	1
H/W	2	2	1

Entropy $\frac{\sum P_i + N_i}{P+N} \left(I(P_i N_i)\right)$

$= \frac{3+3}{5+5}(1) + \frac{2+2}{5+5}(1) = 1$

Gain = Class entropy − Entropy (Type)

$= 1 - 1 = 0$

Information Gain

Age → 0.6

Competion → 0.124515

Type → 0

Age → Root Node

Old — Mid — New

Old → down

Mid → ?

New → up

Again calculate only Mid

P = 2, N = 2 → 2 down, 2 up

Entropy (profit) = 1

(A) Competition :

	Pi	Ni	I(Pi Ni)
Yes	0	2	0
NO	2	0	0

Entropy: $= \frac{2}{4}(0) + \frac{2}{4}(0) = 0$

Gain $= 1 - 0 = 1$

(*) Type

	Pi	Ni	I(PiNi)
S/w	1	1	1
H/w	1	1	1

Entropy:

$$= \frac{\Sigma Pi + Ni}{P+N} \left(I(PiNi)\right)$$

$$= \frac{1+1}{2+2}(1) + \frac{1+1}{2+2}(1) = 1$$

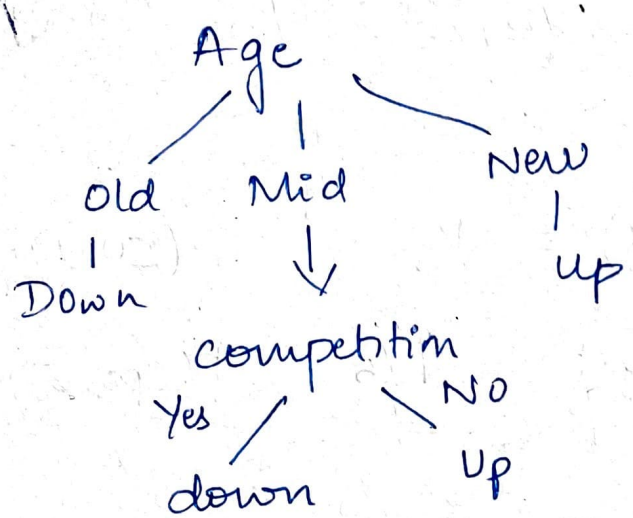
Gain: $1 - 1 = 0$

(Class) Entropy — (Entropy Attribute) =

Gain

Competition 1 ← Node.

Type 0



Decision Tree using Gini Index

Weekend	Weather	Parents	Money	Decision
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay In
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

Gini Index $_{\text{Attribute}}$
↳ having minimum Gini Index is having the maximum information

→ Calculate Gini Index for all attributes

→ 6 → Cinema
2 → Tennis
1 → Stay in
1 → Shopping.

$$\text{Gini}(S) = 1 - \left[\left(\frac{6}{10}\right)^2 + \left(\frac{2}{10}\right)^2 + \left(\frac{1}{10}\right)^2 + \left(\frac{1}{10}\right)^2 \right]$$

$$= 0.58$$

Decision tree : Gini Index

* Money

7 → Rich

3 → Poor

Money = Poor → Cinema

$$Gini(S) = 1 - \left[\left(\frac{3}{3}\right)^2\right] = 0$$

Money = Rich

→ 2 → Tennis 1 → stay in

3 → Cinema 1 → shopping

$$Gini(S) = 1 - \left[\left(\frac{2}{7}\right)^2 + \left(\frac{3}{7}\right)^2 + \left(\frac{1}{7}\right)^2 + \left(\frac{1}{7}\right)^2\right]$$

↳ [2+3+1+1]

$$= 0.694$$

* Weighted Average (Money)

$$= 0 * \left(\frac{3}{10}\right) + 0.694 * \left(\frac{7}{10}\right) = 0.486$$

↓ Gini Index ← Money = Poor ← Money = Rich

---

* Parents

Yes → 5

No → 5

Parent → Yes

$$Gini(S) = 1 - \left[\left(\frac{5}{5}\right)^2\right] = 0$$

↓

Cinema

[5 examples]

Parents = NO

2 → Tennis
1 → stay in
1 → shopping
1 → Cinema

} 2+1+1+1 [5]

$$Gini(S) = 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2\right]$$

$$= 0.72$$

Weighted Average (Parents)

$$= 0 * \left[\frac{5}{10}\right] + [0.72 * \left(\frac{5}{10}\right)]$$

5 example Cinema

5 examples

$$= 0.36$$

Weather

Sunny — 3 examples
Rainy — 3 examples
Windy — 4 examples.

Weather → Sunny:

2 – Cinema
1 – Tennis

$$Gini(S) = 1 - \left[\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2\right]$$

$$= 0.444$$

$$Gini(Rainy) = 1 - \left[\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2\right]$$

$$= 0.444$$

Weather: Rainy

$$Gini(windy) = 1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right]$$

Weather: Windy

$$= 0.375$$

Weighted Average:

$$= 0.444 * \left(\frac{3}{10}\right) + 0.444 * \left(\frac{3}{10}\right) + 0.375\left(\frac{4}{10}\right)$$

$$= 0.416$$

For weather → Gini Index → 0.416

Parents → Gini Index → 0.36 → Minimum

Money → Gini Index → 0.486

Max. info

Parents has the Root Node

Two possibility

Yes ↙ (cinema) ; No ↘ ?

W1	Sunny	Yes	Rich	Cinema	W2	Sunny	No	Rich	Tennis
W3	windy	Yes	Rich	Cinema	W5	Rainy	No	Rich	Stay in
W4	Rainy	Yes	Poor	Cinema	W7	Windy	No	Poor	Cinema
W6	Rainy	Yes	Poor	Cinema	W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema	W10	Sunny	No	Rich	Tennis

Sunny (2 examples) [Weather Attribute]

2 → Tennis

$$Gini(S) = 1 - \left[\left(\frac{2}{2}\right)^2\right] = 0$$

Rainy (1 example)

1 → Stay in

$$Gini(S) = 1 - \left[\left(\frac{1}{1}\right)^2\right] = 0$$

Windy (2 example)

1 → Cinema

1 → Shopping

$$Gini(S) = 1 - \left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right] = 0.5$$

Weighted Average

$$= 0 * \left(\frac{2}{5}\right) + 0 * \left(\frac{1}{5}\right) + 0.5 * \left(\frac{2}{5}\right)$$

$$= 0.2$$

Parents → NO → Money Attribute

Rich : 4 example.

1 → stay in, 1 → Shopping, 2 - Tennis.

$$Gini(S) = 1 - \left[\left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right]$$

$$= 0.625$$

Poor: 1 → Cinema → $Gini(S) = 1 - \left[\left(\frac{1}{1}\right)^2\right] = 0$

Weighted Average

(Parents = NO / Money)

$= 0.625 * \left(\frac{4}{5}\right) + 0 * \left(\frac{1}{5}\right)$

$= \underline{0.5}$

For parents = NO / Weather – Gini Index = 0.2

For parents = NO / Money – Gini Index = 0.5

Weather is selected as it has smallest Gini index

Parents = NO, Weather = Sunny

we have all instance → Tennis

W2	Sunny	NO	Rich	Tennis
W10	Sunny	NO	Rich	Tennis

Parent = NO & weather – Rainy

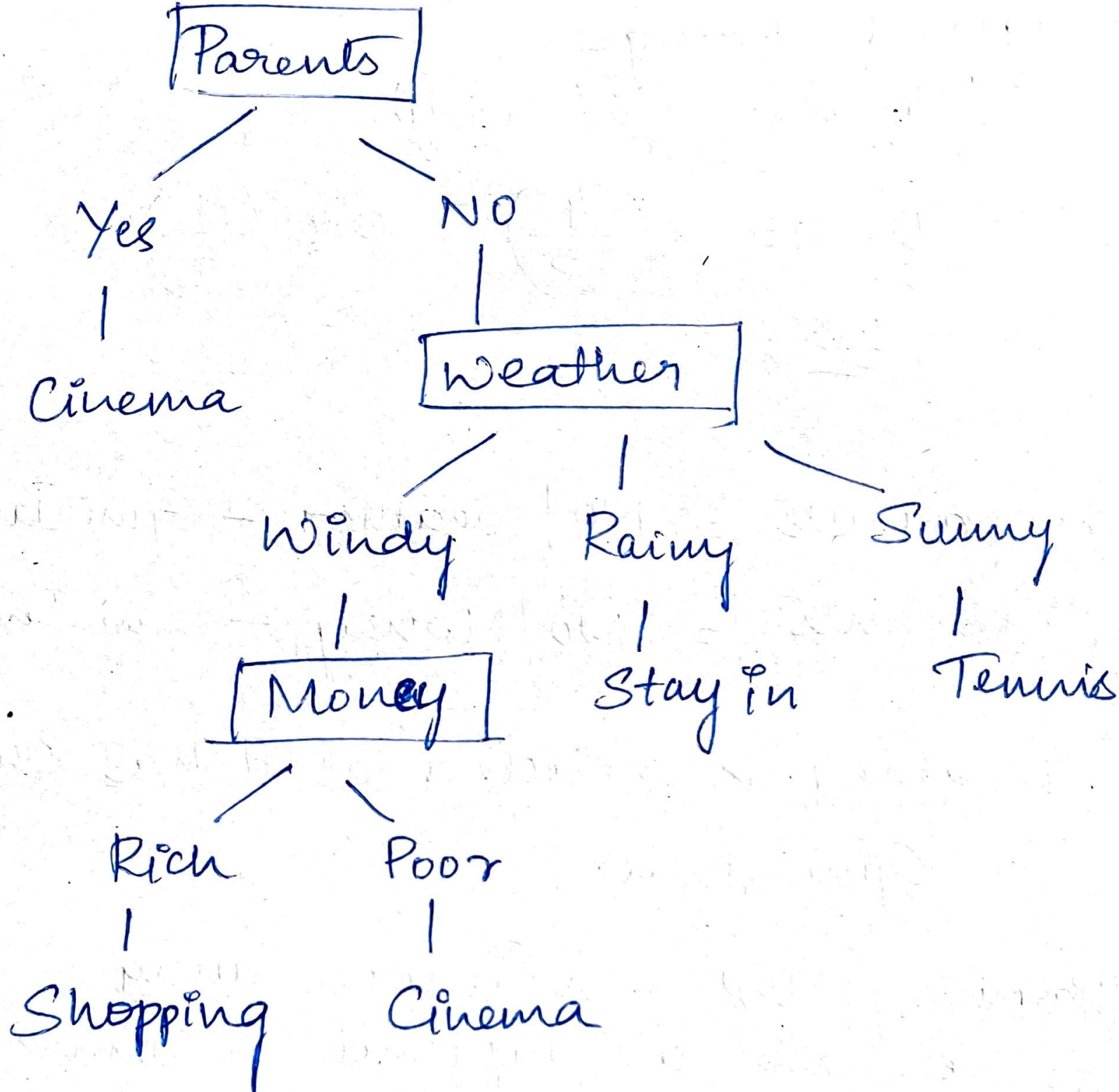
W5	Rainy	NO	Rich	Stay in

Parents = NO ⇒ Windy

W7	Windy	NO	Poor	Cinema
W8	Windy	NO	Rich	Shopping.

∴ Final decision Tree

P.T.O


Parents
Yes
NO
Cinema
Weather
Windy
Rainy
Sunny
Money
Stay in
Tennis
Rich
Poor
Shopping
Cinema

Naive Bayes

Car No.	Color	Type	Origin	Stolen
1	Red	Sports	DOM	Yes
2	Red	Sports	DOM	No
3	Red	Sports	DOM	Yes
4	Yellow	Sports	DOM	No
5	Yellow	Sports	IMP	Yes
6	Yellow	SUV	IMP	No
7	Yellow	SUV	IMP	Yes
8	Yellow	SUV	DOM	No
9	Red	SUV	IMP	No
10	Red	Sports	IMP	Yes

Naive Bayes

$P(Yes)$ 5/10

$P(No)$ 5/10

Color:

$P(Red/Yes) = 3/5$ $P(Red/No) = 2/5$

$P(Yellow/Yes) = 2/5$ $P(Yellow/No) = 3/5$

Type:

$P(SUV/Yes) = 1/5$ $P(SUV/No) = 3/5$

$P(Sports/Yes) = 4/5$ $P(Sports/No) = 2/5$

Origin:

$P(DOM/Yes) = 2/5$ $P(DOM/No) = 3/5$

$P(IMP/Yes) = 3/5$ $P(IMP/No) = 2/5$

unlabelled sample – ?

Red SUV DOM → [redacted] → NO

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

$$P(X|Yes) \cdot P(Yes) = P(Red|Yes) \cdot P(SUV|Yes) \cdot P(DOM|Yes) \cdot P(Yes)$$

$$= 3/5 \cdot 1/5 \cdot 2/5 = 0.024$$

$$P(X|NO) \cdot P(NO) = P(Red|NO) \cdot P(SUV|NO) \cdot P(DOM|NO) \cdot P(NO)$$

$$= 2/5 \cdot 3/5 \cdot 3/5$$

$$P(X|NO) = 0.072$$

$$P(X|NO) > P(X|Yes) \rightarrow P(X|NO)$$