Problem-1

Class	size	color	shape
A	small	Yellow	Round
A	Big	yellow	Round
A	Big	Red	Round
A	small	Red	Round
B	small	Black	Round
B	Big	Black	cube
B	Big	yellow	cube
B	Big	Black	Round
B	small	yellow	Cube

Gini index $n=9$

$$Gini(S) = 1 - \sum_{i=1}^{n} P_i^2$$

$$= 1 - \left[\left(\frac{4}{9}\right)^2 + \left(\frac{5}{9}\right)^2\right] = 0.49$$

For size

size < small (4), Big (5)

size-small : $Gini(S) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right] = 0.5$

size-big : $Gini(S) = 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2\right] = 0.48$

weighted Avg $= 0.5 * \frac{4}{9} + (0.48)\left(\frac{5}{9}\right) = 0.48$

For color attribute

color < yellow (4), Red (2), Black (3)

yellow $Gini(S) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right] = 0.5$

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

Black

$$Gini(S) = 1 - \left[\left(\frac{3}{8}\right)^2\right] = 0$$

weighted average = 0.22

Shape

Shape < Round (6), Cube (3)

Round

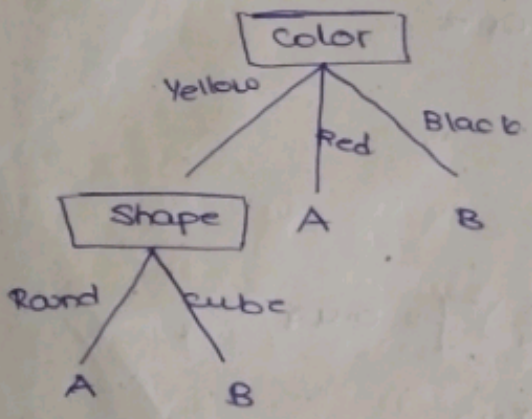
$$Gini(S) = 1 - \left[\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right] = 0.44$$

Cube

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

weighted average = 0.29

Decision tree



Problem-2

A	B	C	class
15	1	A	C1
20	3	B	C2
25	2	A	C1
30	4	A	C1
35	2	B	C2
25	4	A	C1
15	2	B	C2
20	3	B	C2

Gini index n=8

$$Gini(S) = 1 - \left[\left(\frac{4}{8}\right)^2 + \left(\frac{4}{8}\right)^2\right] = 0.5$$

For A

$$mean = \frac{15+20+26+30+35+26+15+20}{8} = 23.12$$

	C_1	C_2	
<=23.1	1	3	=4
>23.1	3	1	=4

For <=23.1

$$Gini(S) = 1 - \left[\left(\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2\right] = 0.375$$

For >23.1

$$Gini(S) = 1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right] = 0.375$$

Weighted average = 0.375

For B

mean = 2.6

	C_1	C_2	
<=2.6	2	2	=4
>2.6	2	2	=4

<=2.6

$$Gini(S) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right] = 0.5$$

>2.6

Gini(S) = 0.5

weighted average = 0.5

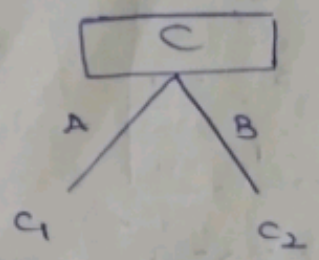
For $C \leq \begin{matrix} A \\ B \end{matrix}$

$$Gini(S) = 1 - \left[\left(\frac{4}{4}\right)^2\right] = 0$$

$$Gini(S) = 1 - \left[\left(\frac{4}{4}\right)^2\right] = 0$$

Weighted average = 0

Decision tree



Problem 3

example	colour	shape	size	class
1	Red	square	Big	+
2	Blue	square	Big	+
3	Red	circle	Big	+
4	Red	circle	small	-
5	Green	square	small	-
6	Green	square	Big	-

Gini index

$$Gini(S) = 1 - \left[\left(\frac{3}{6}\right)^2 + \left(\frac{3}{6}\right)^2\right] = 0.5$$

Colour

Red

$$Gini(S) = 1 - \left[\left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2\right] = 0.44$$

Blue

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

Green

Gini(S) = 0

weighted average = 0.22

Shape

Square

$Gini(S) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2 \right] = 0.5$

Circle

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

Gini(S) = 0

weighted average = 0.83

Size

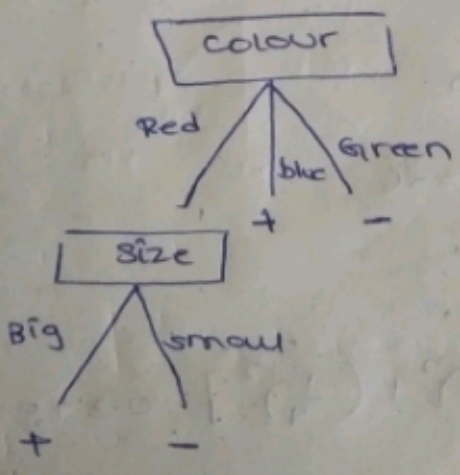
Big

$Gini(S) = 1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2 \right] = 0.375$

Small

Gini(S) = 0

weighted average = 0.25



Problem-4

A	B	class label
T	F	+
T	T	+
T	T	+
T	F	-
T	T	+
F	F	-
F	F	-
F	F	-
T	T	-
T	F	-

Gini index $n=10$

$$Gini(S) = 1 - \left[\left(\frac{4}{10}\right)^2 + \left(\frac{6}{10}\right)^2\right] = 0.48$$

A

T

$$Gini(S) = 1 - \left[\left(\frac{4}{7}\right)^2 + \left(\frac{3}{7}\right)^2\right] = 0.48$$

F

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

weighted average: 0.336

B

T

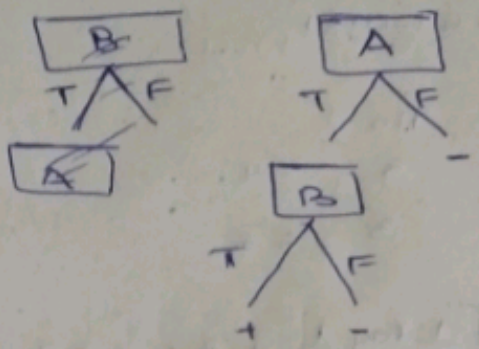
$$Gini(S) = 1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right] = 0.375$$

F

$$Gini(S) = 1 - \left[\left(\frac{3}{6}\right)^2 + \left(\frac{3}{6}\right)^2\right] = 0.5$$

weighted average: 0.45

Decision tree



Problem-5

Outlook	wind	play
rain	strong	No
sunny	weak	Yes
overcast	weak	yes
rain	weak	yes
sunny	strong	yes
rain	strong	No
overcast	strong	No

n=7

Information gain

$$Info(D) = -\left[\frac{4}{7}\log_2\frac{4}{7} + \frac{3}{7}\log_2\frac{3}{7}\right] = 0.98$$

outlook

	yes	No	
rain	1	2	=3
sunny	2	0	=2
overcast	1	1	=2

$$Info(rain) = -\left[\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right] = 0.9$$

$Info(sunny) = 0$

$Info(overcast) = 1$

Info(outlook) = $\frac{3}{7} \times 0.9 + \frac{2}{7} \times 0 + \frac{2}{7} \times 1 = 0.67$

Gain = 0.98 − 0.67 = 0.31

Wind

	Yes	No	
strong	1	3	= 4
weak	0	3	= 3

Info(strong) = $-\left[\frac{1}{4}\log\frac{1}{4} + \frac{3}{4}\log\frac{3}{4}\right] = 0.81$

Info(weak) = $-\left[\frac{3}{3}\log_2\frac{3}{3}\right] = 0$

Info(wind) = $\frac{4}{7} \times 0.81 = 0.46$

Gain = 0.98 − 0.46 = 0.52

Decision tree