D-Tree

Name : Aditi M. Chavan Roll. no : PB 54

Class P : buys_computer = 'yes'
Class N : buys_computer = 'no'

1. Calculate Entropy for : class labels

$$Info(D) = I(9,5) = -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) = 0.940$$

2. Calculate Information of Each Attribute

age	pi	ni	I(pi,ni)
<=30	2	3	0.971
30....40	4	0	0
>40	3	2	0.971

$$Info_{age}(D) = \frac{5}{14} \times I(2,3) + \frac{4}{14} \times I(4,0) + \frac{5}{14} \times I(3,2) = 0.694$$

Income	pi	ni	I(pi,ni)
high	2	2	1
medium	4	2	0.918
low	3	1	0.811

$$Info_{Income}(D) = \frac{4}{14} I(2,2) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1) = 0.911$$

Student	pi	ni	I(pi,ni)
yes	6	1	0.592
no	3	4	0.985

$$Info_{student}(D) = \frac{7}{14} I(6,1) + \frac{7}{14} \times I(3,4) = 0.7885$$

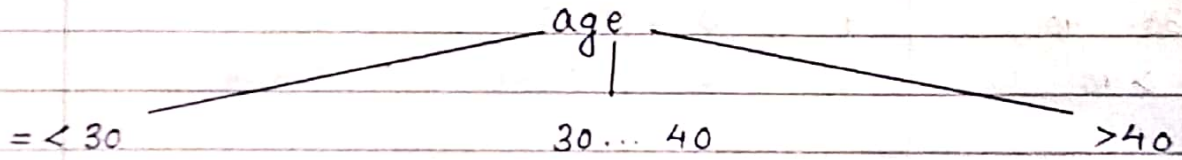
credit rating	pi	ni	I(pi, ni)	$I_{cr}(D) = \frac{6}{14} I(3,3)$
excellent	3	3	1	$+ \frac{8}{14} I(6,2) = 0.892$
fair	6	2	0.811	

Gain (age) = $-Info_{age}(D) + Info(D) = -0.694 + 0.940 = 0.246$

Gain (income) = $Info(D) - Info_{income}(D) = 0.940 - 0.911 = 0.029$

Gain (student) = $Info(D) - Info_{student}(D) = 0.940 - 0.789 = 0.151$

Gain (credit_rating) = $Info(D) - Info_{cr}(D) = 0.940 - 0.892 = 0.048$



Income	Student	credit rating	class
H	N	F	N
H	N	E	N
M	N	F	N
L	Y	F	Y
M	Y	E	Y

Income	Student	Credit rating	Class
H	N	F	Y
L	Y	E	Y
M	N	E	Y
H	Y	F	Y

Income	Student	Credit rating	class
medium	N	F	Y
low	Y	F	Y
low	Y	E	N
medium	Y	F	Y
medium	N	E	N

↳ For age between 30...40
all class labels ⇒ Y
∴ if age = 30...40
buys_computer = Yes

H ⇒ high
M ⇒ medium
L ⇒ low

F ⇒ fair
E ⇒ excellent
Y ⇒ Yes
N ⇒ No

For age > 40.

1. Cal. Entropy for Class Labels.

$$\text{Info}(D) = I(3,2) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$$

2. Cal. Information of Each Attribute.

Income	pi	ni	I(pi, ni)	$\text{Info}_{income}(D) = 0 +$
high	0	0	0	$\frac{2}{5} I(1,1) + \frac{3}{5} I(2,1)$
low	1	1	1	
medium	2	1	0.918	$= 0.951$

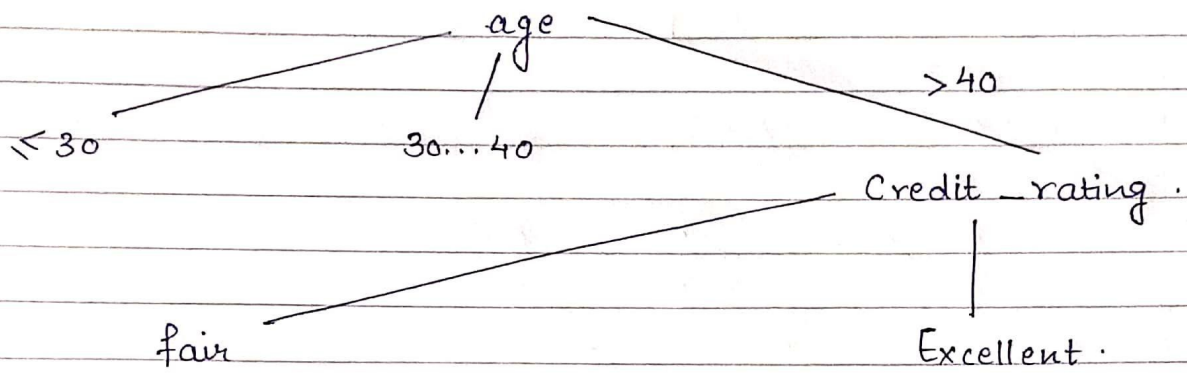
Student	pi	ni	I(pi, ni)	$\text{Info}_{student}(D) = \frac{5}{5} I(2,1)$
Yes	2	1	0.918	$+ \frac{2}{5} I(1,1) = 0.951$
No	1	1	1	

Credit rating	pi	ni	I(pi, ni)	$\text{Info}_{cr}(D) = \frac{3}{5} I(3,0)$
Fair	3	0	0	$+ \frac{2}{5} I(0,2) = 0$
Excellent	0	2	0	

$$\text{Gain (income)} = \text{Info}(D) - \text{Info}_{income}(D) = 0.971 - 0.951 = 0.020$$

$$\text{Gain (student)} = \text{Info}(D) - \text{Info}_{student}(D) = 0.971 - 0.951 = 0.020$$

$$\text{Gain (credit\_rating)} = \text{Info}(D) - \text{Info}_{cr}(D) = 0.971 - 0 = 0.971.$$



Income	Student	Class
M	N	Y
L	Y	Y
M	Y	Y

Income	Student	Class.
L	Y	N
M	N	N

if age > 40 & credit_rating = fair
buys_computer = Yes

if age > 40 & credit rating = excellent
then buys_computer = NO

For age ≤ 30

1. Calc. Entropy of Class Label

$$\text{Info}(D) = I(2,3) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.971$$

2. Calc. ~~Ent~~ Information of Each Attribute.

Income	pi	ni	I(pi, ni)
high	0	2	0
medium	1	1	1
low	1	0	0

$\text{Info}_{\text{income}}(D) = \frac{2}{5} I(1,1) = 0.4$

Student	pi	ni	I(pi, ni)
yes	2	0	0
no	0	3	0

$\text{Info}_{\text{student}}(D) = 0$

credit_rating	pi	ni	I(pi, ni)
excellent	1	1	1
fair	1	2	0.918

$\text{Info}_{cr}(D) = \frac{2}{5} \times I(1,1) + \frac{3}{5} I(1,2) = 0.951$

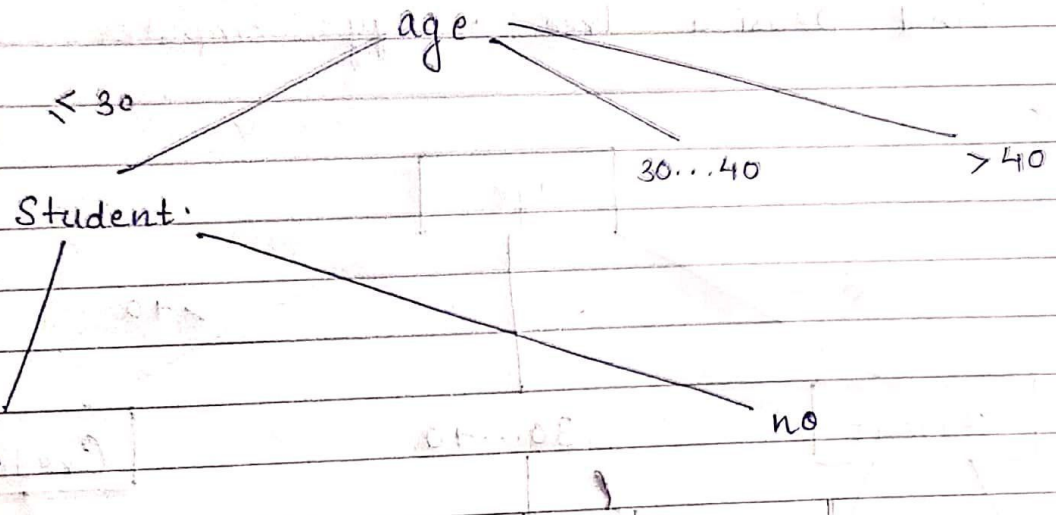
Gain (income) = 0.971 − 0.4 = 0.571

Gain (student) = 0.971 − 0 = 0.971

Gain (credit_rating) = 0.971 − 0.951 = 0.020

if age < 30 :
    if student :
        buys_computer = Yes

if age < 30 :
    if (! student) :
        buys_computer = NO



Income	Credit rating	Class
L	F	Y
M	E	Y

Income	Credit rating	class
H	F	N
H	E	N
M	F	N

Final Decision Tree - buys_computer