7.2 Learning Decision Trees (continued)

- 3 parameters are computed:
  - **a)** **Information Content** or **Entropy** regarding the classes in the samples
  - **b)** **Remainder** or Expected **Remaining Entropy** of an **Attribute** Test
  - **c)** **Information Gain** of an **Attribute** Test

a) $I(P(C_1), P(C_2), \ldots, P(C_m)) = \Sigma_{i = 1:m} -P(C_i) \times Log_2(P(C_i))$,
where $P(C_i)$ – probability of the class C_i ,
m – number of classes.

For our example,
I(P(positive), P(negative)) =

$$-P(positive) \times Log_2(P(positive)) - P(negative) \times Log_2(P(negative))$$

ID	**Age**	**Income**	**Student**	**Credit Rating**	*Decision/ Class/ Label*
1	≤ 30	high	no	fair	negative
2	≤ 30	high	no	excellent	negative
3	31…40	high	no	fair	positive
4	> 40	medium	no	fair	positive
5	> 40	low	yes	fair	positive
6	> 40	low	yes	excellent	negative
7	31…40	low	yes	excellent	positive
8	≤ 30	Medium	no	fair	negative
9	≤ 30	low	yes	fair	positive
10	> 40	medium	yes	fair	positive
11	≤ 30	medium	yes	excellent	positive
12	31…40	medium	no	excellent	positive
13	31…40	high	yes	fair	positive
14	> 40	medium	no	excellent	negative

- That is, I(9/14, 5/14) = - 9/14 × Log_2 (9/14) - 5/14 × Log_2 (5/14) ≈ 0.940 (bits)

b. Remainder (A) =

$\Sigma_{i=1\,:\,v}$ P(Samples with ith value of A) × I(Classes in samples with ith value of A), where v = number of distinct values of A.

i. Remainder(Age) = 5/14 × I(2/5, 3/5) [≤ 30]
+ 4/14 × I(4/4, 0/4) [31 … 40]
+ 5/14 × I(3/5, 2/5) [> 40]

ID	**Age**	**Income**	**Student**	**Credit Rating**	*Decision/ Class/ Label*
1	≤ 30	high	no	fair	negative
2	≤ 30	high	no	excellent	negative
3	31…40	high	no	fair	positive
4	> 40	medium	no	fair	positive
5	> 40	low	yes	fair	positive
6	> 40	low	yes	excellent	negative
7	31…40	low	yes	excellent	positive
8	≤ 30	Medium	no	fair	negative
9	≤ 30	low	yes	fair	positive
10	> 40	medium	yes	fair	positive
11	≤ 30	medium	yes	excellent	positive
12	31…40	medium	no	excellent	positive
13	31…40	high	yes	fair	positive
14	> 40	medium	no	excellent	negative

- That is, Remainder(Age) = 2 ×5/14 × I(2/5, 3/5) ≈ 0. 694 (bits)

ii) Remainder(Income) = ?

iii) Remainder(Student) = ?

iv) Remainder(Credit Rating) = ?

c) Gain(A) = I(Classes in all samples in the table) – Remainder(A)

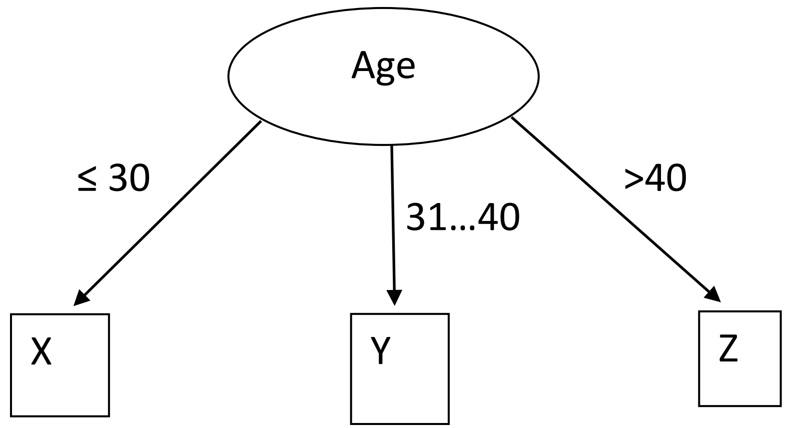
Gain(Age) = 0.940 – 0.694 = 0.246 (bits)

Check that, Gain(Income) = 0.029 bits,

Gain(Student) = 0.151 bits, and

Gain(Credit Rating) = 0.048 bits.

So, we have the attribute ‘Age’ with the highest Gain, and this leads to what follows.



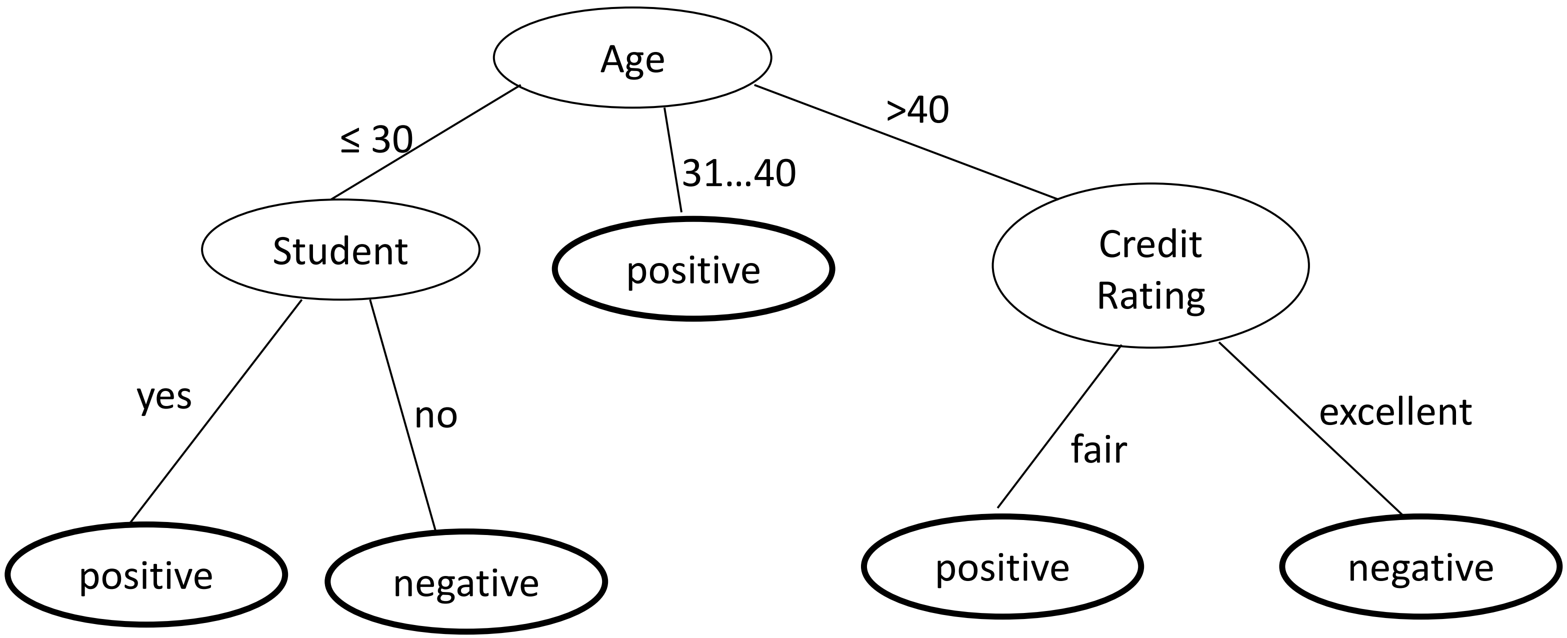
X =

Age	**Income**	**Student**	**Credit Rating**	*Decision/ Class/ Label*
≤ *30*	high	no	fair	negative
≤ *30*	high	no	excellent	negative
≤ *30*	medium	no	fair	negative
≤ *30*	low	yes	fair	positive
≤ *30*	medium	yes	excellent	positive

Y = ?

Z = ?

Finally, we have the following tree. [Self study]



And it means that we have **learned the following 5 rules**.

1. If ‘Age’ = ‘≤ 30’ and ‘Student’ = ‘yes’, then ‘Class’ = ‘Buys a computer’.
2. If ‘Age’ = ‘≤ 30’ and ‘Student’ = ‘no’, then ‘Class’ = ‘Does not buy a computer’.
3. …..
4. …..
5. If ‘Age’ = ‘>40’ and ‘Credit Rating’ = ‘excellent’, then ‘Class’ = ‘Does not buy a computer’.