

# HW4

19 ຖຸນາພັນ 2567 5:10



Student\_ye  
s\_gini

Gini					
age	income	student	credit_rating	buys_computer	
<=30	high	no	fair	no	
<=30	high	no	excellent	no	
>30	high	no	fair	yes	
<=30	low	no	fair	yes	
<=30	high	yes	fair	yes	
>30	low	yes	excellent	no	
>30	high	yes	excellent	yes	
<=30	low	no	excellent	yes	
<=30	low	yes	fair	no	
<=30	low	yes	fair	yes	
<=30	low	yes	excellent	yes	
>30	low	no	excellent	yes	
>30	high	yes	fair	yes	
>30	low	no	excellent	no	

$\text{gini}(D) = 1 - \left( \left(\frac{5}{14}\right)^2 + \left(\frac{9}{14}\right)^2 \right)$   
 $= 0.459$

$\text{gini}(\text{age}) = 1 - \left( \left(\frac{2}{8}\right)^2 + \left(\frac{6}{8}\right)^2 \right)$   
 $= 0.468$

$\text{gini}(D > 30) = 1 - \left( \left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2 \right)$   
 $= 0.444$

$\text{gini}_{\text{age}}(D) = \frac{|D_{\leq 30}| \text{gini}(D_{\leq 30}) + |D_{> 30}| \text{gini}(D_{> 30})}{|D|}$   
 $= \frac{8}{14} (0.468) + \frac{6}{14} (0.444)$   
 $= 0.457$

$\text{gini}(D) = 0.459$   
9

$$gini(D) = 0.459$$

$$gini_{age}(D) = 0.457 ; \Delta gini_{age}(D) > 0.02$$

$$gini_{income}(D) = 0.457 ; \Delta gini_{income}(D) = 0.02$$

$$gini_{credit}(D) = 0.428 ; \Delta gini_{credit}(D) = 0.031$$

$$gini_{student}(D) = 0.366 ; \Delta gini_{student}(D) = 0.093$$

income

$$\text{gini}(\text{high}) = 1 - \left( \left( \frac{2}{6} \right)^2 + \left( \frac{4}{6} \right)^2 \right)$$
$$= 0.444$$

$$\text{gini}(\text{low}) = 1 - \left( \left( \frac{5}{8} \right)^2 + \left( \frac{3}{8} \right)^2 \right)$$
$$= 0.468$$

$$\text{income}(D) = \frac{6}{14}(0.444) + \frac{8}{14}(0.468)$$
$$= 0.457$$

credit

$$\text{gini}(\text{fair}) = 1 - \left( \left( \frac{2}{8} \right)^2 + \left( \frac{6}{8} \right)^2 \right)$$
$$= 0.375$$

$$\text{gini}(\text{excellent}) = 1 - \left( \left( \frac{3}{6} \right)^2 + \left( \frac{3}{6} \right)^2 \right)$$
$$= 0.5$$

$$\text{gini}(\text{credit}(D)) = \frac{8}{14}(0.375) + \frac{6}{14}(0.5)$$
$$= 0.428$$

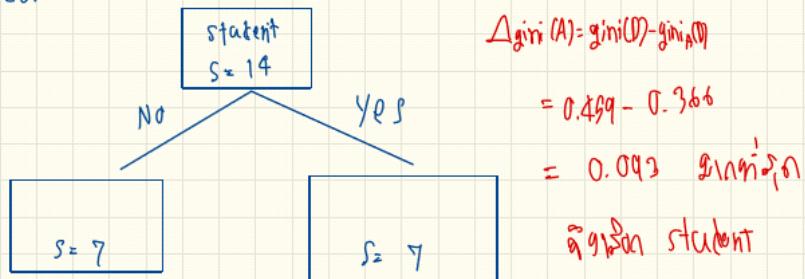
student

$$\text{gini(Yes)} = 1 - \left( \left( \frac{6}{7} \right)^2 + \left( \frac{1}{7} \right)^2 \right)$$
$$= 0.244$$

$$\text{gini(No)} = 1 - \left( \left( \frac{2}{7} \right)^2 + \left( \frac{4}{7} \right)^2 \right)$$
$$= 0.489$$

$$\text{gini(student)} = \frac{7}{14}(0.244) + \frac{7}{14}(0.489)$$
$$= 0.366$$

for student



Gini(No Node) yes

A	B	C	D	E	F	G
age	income	student	credit_rating	buys_computer		
<=30	high	no	fair	no		
<=30	high	no	excellent	no		
>30	high	no	fair	yes		
<=30	low	no	fair	yes		
<=30	high	yes	fair	yes		
>30	low	yes	excellent	no		
>30	high	yes	excellent	yes		
>30	low	no	fair	no		
<=30	low	yes	fair	yes		
<=30	low	yes	fair	yes		
<=30	low	yes	excellent	yes		
>30	low	no	excellent	yes		
>30	high	yes	fair	yes		
>30	low	no	excellent	no		

$gini(CD) = 1 - \left( \left(\frac{6}{7}\right)^2 + \left(\frac{1}{7}\right)^2 \right)$

$= 0.244$

*age*

$gini(<=30) = 1 - \left( \left(\frac{4}{4}\right)^2 + \left(\frac{0}{4}\right)^2 \right)$

$= 0$

$gini(>30) = 1 - \left( \left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2 \right)$

$= 0.444$

$gini_{age}(D) = \frac{4}{7}(0) + \frac{3}{7}(0.444)$

$= 0.19$

---

income

$$\text{gini}(\text{high}) = 1 - \left( \left( \frac{3}{3} \right)^2 + \left( \frac{0}{3} \right)^2 \right) \\ = 0$$

$$\text{gini}(\text{low}) = 1 - \left( \left( \frac{3}{4} \right)^2 + \left( \frac{1}{4} \right)^2 \right) \\ = 0.375$$

$$\text{gini}_{\text{income}}(D) = \frac{2}{7}(0) + \frac{4}{7}(0.375) \\ = 0.21$$

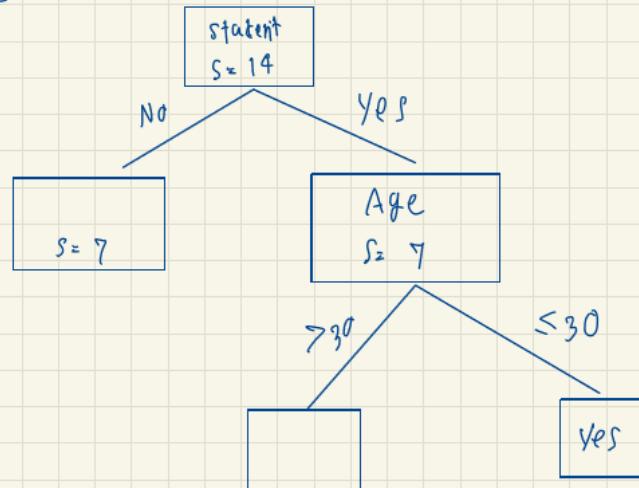
credit\_rating

$$\text{gini}(\text{fair}) = 1 - \left( \left( \frac{4}{4} \right)^2 + \left( \frac{0}{4} \right)^2 \right) \\ = 0$$

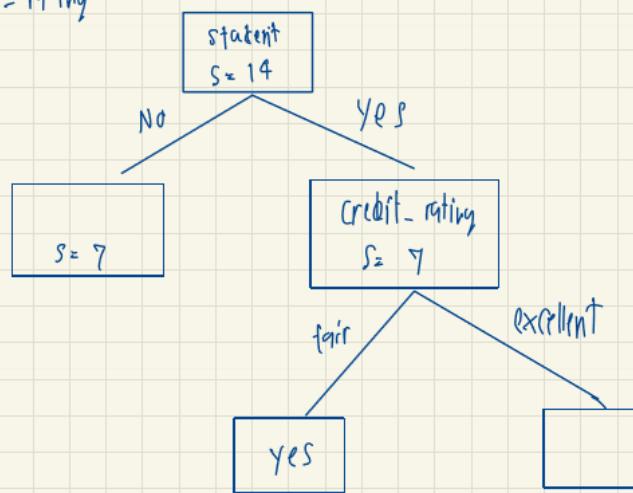
$$\text{gini}(\text{excellent}) = 1 - \left( \left( \frac{2}{3} \right)^2 + \left( \frac{1}{3} \right)^2 \right) \\ = 0.444$$

$$\text{gini}_{\text{credit}}(D) = \frac{4}{7}(0) + \frac{3}{7}(0.444) \\ = 0.19$$

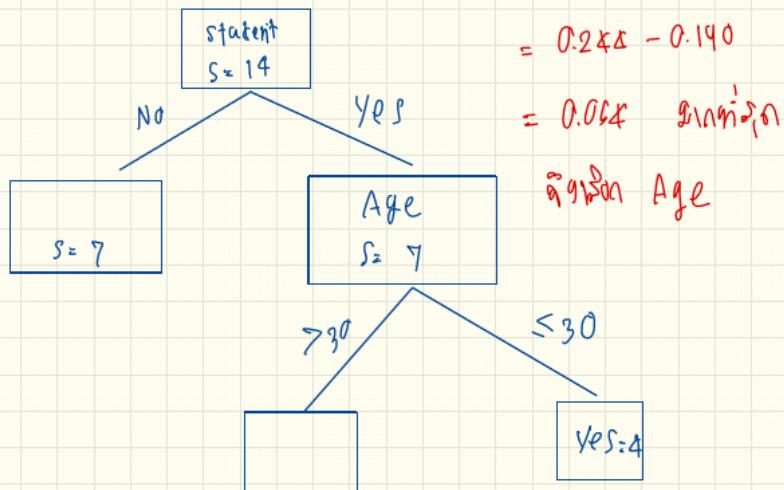
affiliation Age



affiliation credit\_rating



Decision Age



$$\Delta \text{gini}(A) = \text{gini}(D) - \text{gini}_{\text{imp}}(A)$$

$$= 0.244 - 0.190$$

$$= 0.054 \quad \text{Gain from Age}$$

Gain from Age

Home Insert Page Layout Formulas Data Review View Autom

DETACHED VIEW Be careful—files from the Internet can contain viruses. Unless you need to:

A B C D E F G

	age	income	student	credit_rating	buys_computer	
1	<=30	high	no	fair	no	
2	<=30	high	no	excellent	no	
3	>30	high	no	fair	yes	
4	<=30	low	no	fair	yes	
5	<=30	high	yes	fair	yes	
6	>30	low	yes	excellent	no	
7	>30	high	yes	excellent	yes	
8	<=30	low	no	fair	no	
9	<=30	low	yes	fair	yes	
10	<=30	low	yes	excellent	yes	
11	>30	low	no	excellent	yes	
12	>30	high	yes	fair	yes	
13	>30	low	no	excellent	no	

12 > 30 & 13

$$\text{gini}(CD) = 1 - \left( \left( \frac{2}{3} \right)^2 + \left( \frac{1}{3} \right)^2 \right)$$

$$= 0.444$$

income

$$\text{gini}(low) = 1 - \left( \left( \frac{1}{1} \right)^2 + \left( \frac{0}{1} \right)^2 \right)$$

$$= 0$$

$$\text{gini}(high) = 1 - \left( \frac{2}{2} \right)^2 + \left( \frac{0}{2} \right)^2$$

$$= 0$$

$$gini_{income} = \frac{1}{3}(0) + \frac{2}{3}(0)$$

$$= 0$$

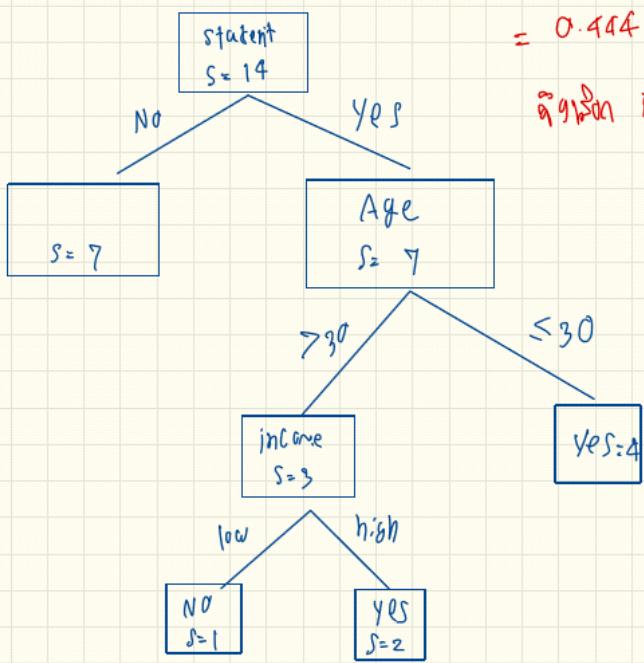
low income  $S=9$

$$\Delta gini(A) = gini(D) - gini(P)$$

$$= 0.444 - 0$$

$$= 0.444 \text{ gini gain}$$

high income



Student\_No

with the student

n student = no.

$$\text{gini}(\text{student} = \text{no}) = 1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2 \\ = 0.490$$

without age

$$\text{gini}_{\text{age}}(\text{student} = \text{no}) = \frac{4}{7} \left(1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2\right) + \frac{3}{7} \left(1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2\right) \\ = \frac{4}{7} (0.375) + \frac{3}{7} (0.444) \\ = 0.2143 + 0.1903 \\ = 0.4046$$

$$\Delta \text{gini}(\text{student} = \text{no}, \text{age}) = 0.49 - 0.4046$$

$$= 0.0854$$

without income

$$\text{gini}_{\text{income}}(\text{student} = \text{no}) = \frac{5}{7} \left(1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2\right) + \frac{4}{7} \left(1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2\right) \\ = \frac{5}{7} \left(\frac{8}{9}\right) + \frac{4}{7} \left(\frac{1}{2}\right) \\ = 0.1905 + 0.2857 \\ = 0.4762$$

$$\Delta \text{gini}(\text{student} = \text{no}, \text{income}) = 0.49 - 0.4762$$

$$= 0.0138$$

without Credit-rating

$$\text{gini}_{\text{credit}}(\text{student} = \text{no}) = \frac{4}{7} \left(1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2\right) + \frac{3}{7} \left(1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2\right) \\ = \frac{4}{7} \left(\frac{1}{2}\right) + \frac{3}{7} \left(\frac{4}{9}\right) \\ = 0.4762$$

$$\Delta \text{gini}(\text{student} = \text{no}, \text{Credit}) = 0.49 - 0.4762$$

$$= 0.0138$$

minimum age

\* student = no age <= 30

$$\text{gini}(\text{student}=\text{no}, \text{age} <= 30) = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{1}{4}\right)^2 \\ = 0.375$$

Family income

$$\text{gini}_{\text{income}}(\text{student}=\text{no}, \text{age} <= 30) = \frac{2}{4} \left(1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right) + \frac{2}{4} \left(1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right) \\ = 0 + \frac{1}{2}(1) \\ = 0.25$$

$$\Delta \text{gini}(\text{student}=\text{no}, \text{age} <= 30, \text{income}) = 0.375 - 0.25 \\ = 0.125 \quad \cancel{\neq} \quad 0.0557$$

Family credit

$$\text{gini}_{\text{credit}}(\text{student}=\text{no}, \text{age} <= 30) = \frac{2}{4} \left(1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2\right) + \frac{2}{4} \left(1 - \left(\frac{1}{1}\right)^2\right) \\ = \frac{2}{4} \left(\frac{4}{9}\right) + \frac{1}{4}(0) \\ = 0.333$$

$$\Delta \text{gini}(\text{student}=\text{no}, \text{age} <= 30, \text{credit}) = 0.375 - 0.333 \\ = 0.042$$

\* student = no age >= 30 minimum income

Income = high  $\Rightarrow$  income No signs

Income = low  $\Rightarrow$  0

$$\text{gini}(\text{student}=\text{no}, \text{age} >= 30, \text{income}=\text{high}) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 \\ = 0.5$$

minimum deviation feature & a boundary node

R student = no age > 30

$$\text{gini}(\text{student} = \text{no}, \text{age} > 30) = 1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2 \\ = \frac{4}{9} \\ = 0.444$$

Father income

$$\text{gini}_{\text{income}}(\text{student} = \text{no}, \text{age} > 30) = \frac{1}{3} \left(1 - \left(\frac{1}{1}\right)^2 - (0)^2\right) + \frac{2}{3} \left(1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right) \\ = 0 + 0.333 \\ = 0.333$$

$$\Delta \text{gini}(\text{student} = \text{no}, \text{age} > 30, \text{income}) = 0.444 - 0.333 \\ = 0.111 \quad \times$$

Father credit rating

$$\text{gini}_{\text{credit}}(\text{student} = \text{no}, \text{age} > 30) = \frac{1}{3} \left(1 - \left(\frac{1}{1}\right)^2 - (0)^2\right) + \frac{2}{3} \left(1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right) \\ = 0.333$$

$$\Delta \text{gini}(\text{student} = \text{no}, \text{age} > 30, \text{credit}) = 0.444 - 0.333 \\ = 0.111 \quad \times$$

0.023

with 2nd If income

i student = no age > 30 income = high

$$\text{gini}(\text{student} = \text{no}, \text{age} > 30, \text{income} = \text{high}) = 1 - \left(\frac{1}{2}\right)^2 = 0$$

= 0

i student = no age > 30 income = low

$$\text{gini}(\text{student} = \text{no}, \text{age} > 30, \text{income} = \text{low}) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$

= 0.5

without credit

i student = no age > 30 credit\_rating = fair

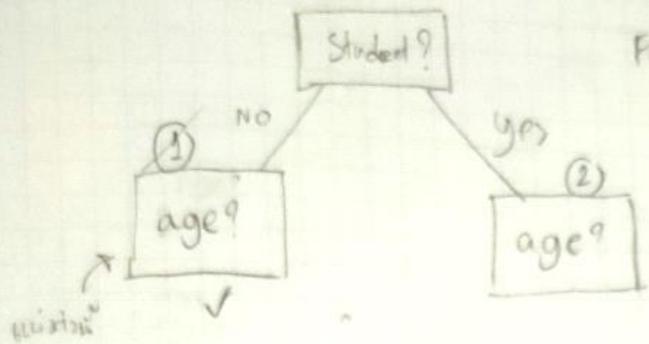
$$\text{gini}(\text{student} = \text{no}, \text{age} > 30, \text{credit\_rating} = \text{fair}) = 1 - \left(\frac{1}{2}\right)^2 = 0$$

= 0

i student = no age > 30 credit\_rating = excellent

$$\text{gini}(\text{student} = \text{no}, \text{age} > 30, \text{credit\_rating} = \text{excellent}) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$

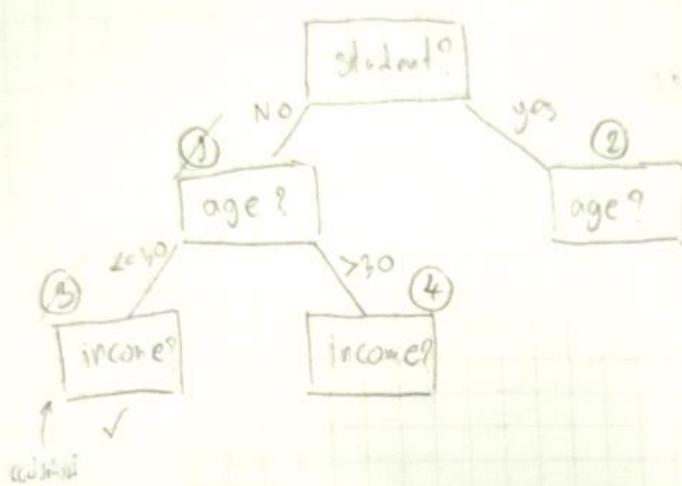
= 0.5



$$NL, N=0, lead=2$$

$$FN + \Delta gini: ① \times \frac{2}{74} = 0.0427$$

$$\Delta gini: ② \times \frac{4}{74} = 0.027$$

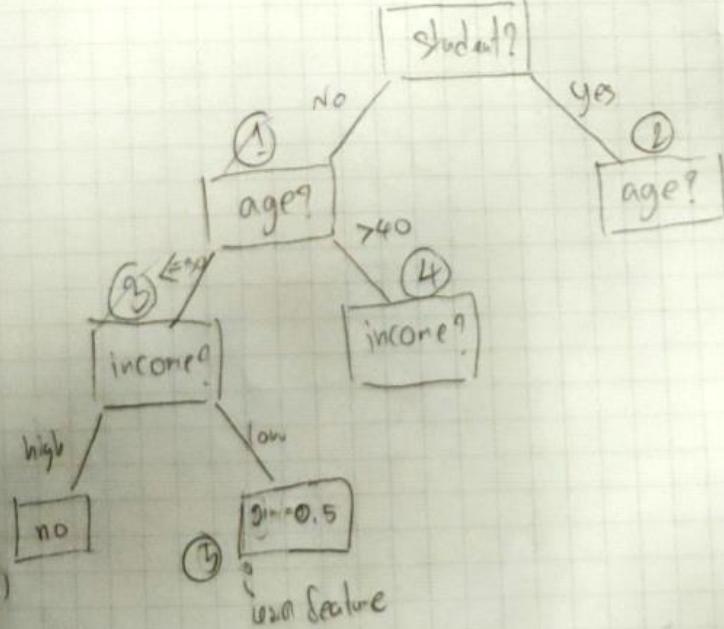


$$NL, N=2, lead=3$$

$$FN + \Delta gini: ③ \times \frac{2}{74} = 0.0357$$

$$\Delta gini: ② \times \frac{4}{74} = 0.027$$

$$\Delta gini: ④ \times \frac{2}{74} = 0.024$$

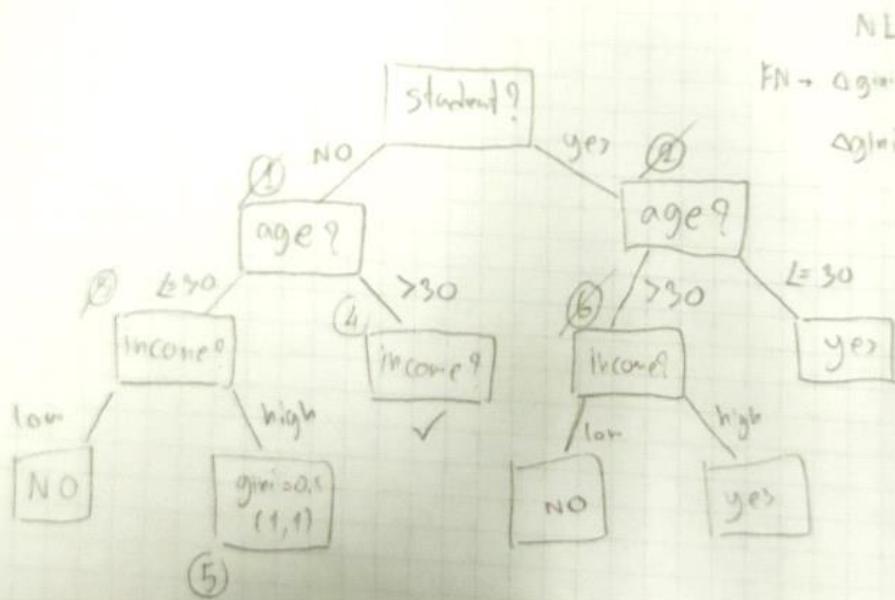
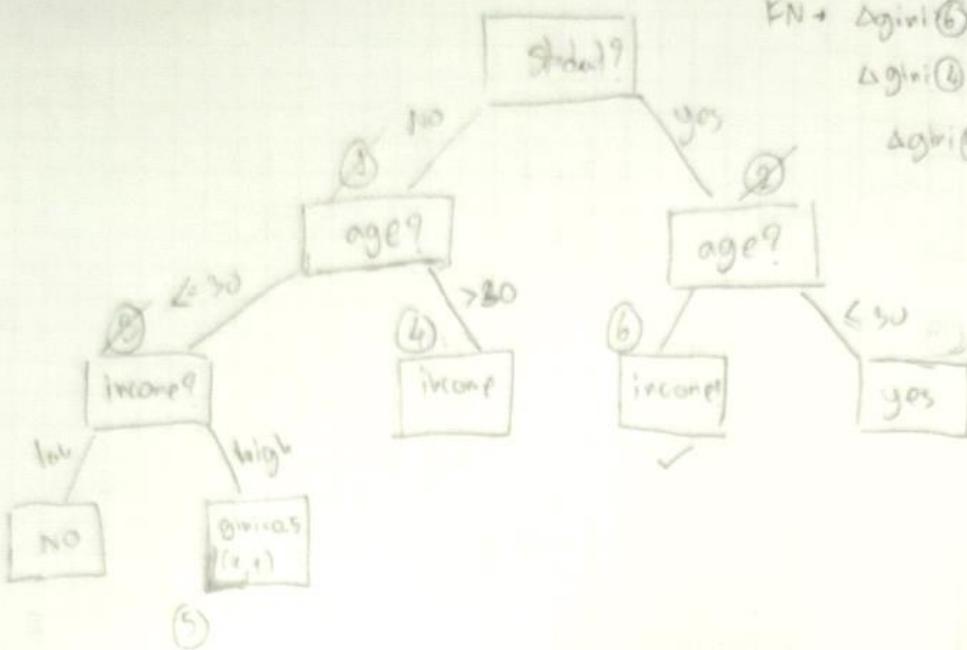


$$NL, N=2, lead=4$$

$$FN \rightarrow \Delta gini: ② \times \frac{2}{74} = 0.027$$

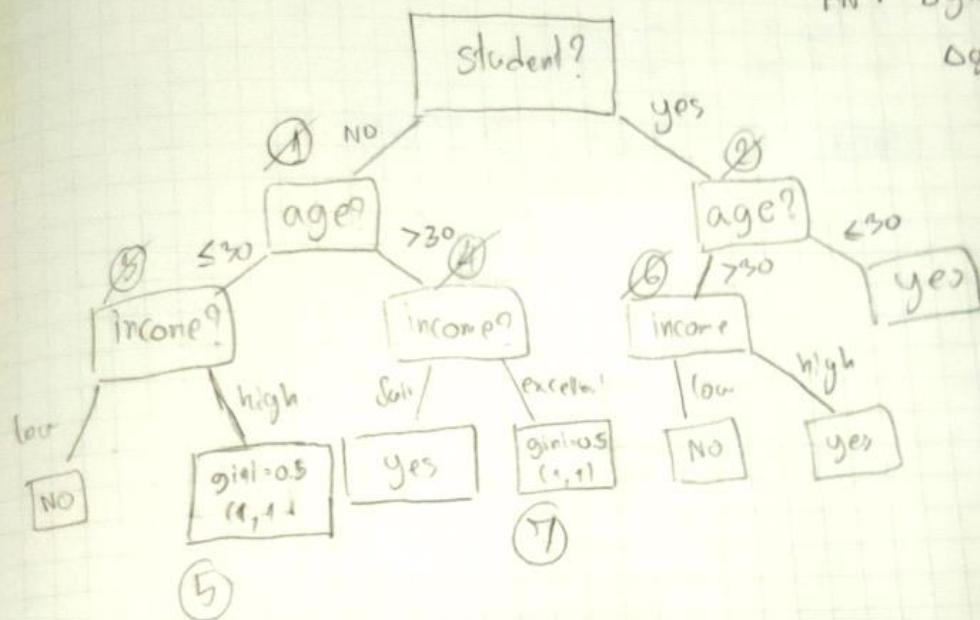
$$\Delta gini: ④ \times \frac{3}{74} = 0.024$$

$$\Delta gini: ⑤ = 0$$



NL N=5 leaf: 7

FN  $\rightarrow$  Dgiri(6) = 0 &  
Dgiri(7) = 0



# MAX LEAF NODES

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

Step 1 : ນີ້ແມ່ນເປົ້າໄຫວ້ວ ສະບັບກະລຸນາຍ

- Class P: buys\_computer = "yes" → 9
- Class N: buys\_computer = "no" → 5

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

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

Step 2 : ນີ້ແມ່ນເປົ້າໄຫວ້ວ ສະບັບກະລຸນາຍຂອງຕ�性

age	p <sub>i</sub>	n <sub>i</sub>	p <sub>i</sub> +n <sub>i</sub>
<=30	5	3	8
>30	4	2	6

$$\text{Info}_{\text{age}}(D) = \frac{8}{14} I(5, 3) + \frac{6}{14} I(4, 3) \\ = \frac{8}{14} \left[ -\frac{5}{8} \log_2 \left( \frac{5}{8} \right) - \frac{3}{8} \log_2 \left( \frac{3}{8} \right) \right] + \\ \frac{6}{14} \left[ -\frac{4}{6} \log_2 \left( \frac{4}{6} \right) - \frac{2}{6} \log_2 \left( \frac{2}{6} \right) \right] \\ = \frac{8}{14} (0.954) + \frac{6}{14} (0.918) \\ = 0.938$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

income	$P_i$	$n_i$	$P_i + n_i$
high	4	2	6
low	5	3	8

Info<sub>income</sub>(D)

$$\begin{aligned}
 &= \frac{6}{14} I(4, 2) + \frac{8}{14} I(5, 3) \\
 &= \frac{6}{14} \left[ -\frac{4}{6} \log_2 \left( \frac{4}{6} \right) - \frac{2}{6} \log_2 \left( \frac{2}{6} \right) \right] + \\
 &\quad \frac{8}{14} \left[ -\frac{5}{8} \log_2 \left( \frac{5}{8} \right) - \frac{3}{8} \log_2 \left( \frac{3}{8} \right) \right] \\
 &= \frac{6}{14} (0.918) + \frac{8}{14} (0.954) \\
 &= 0.938
 \end{aligned}$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

student	$P_i$	$n_i$	$P_i + n_i$
yes	6	1	7
no	3	4	7

Info<sub>student</sub>(D)

$$\begin{aligned}
 &= \frac{7}{14} I(6, 1) + \frac{7}{14} I(3, 4) \\
 &= \frac{7}{14} \left[ -\frac{6}{7} \log_2 \left( \frac{6}{7} \right) - \frac{1}{7} \log_2 \left( \frac{1}{7} \right) \right] + \\
 &\quad \frac{7}{14} \left[ -\frac{3}{7} \log_2 \left( \frac{3}{7} \right) - \frac{4}{7} \log_2 \left( \frac{4}{7} \right) \right] \\
 &= \frac{7}{14} (0.591) + \frac{7}{14} (0.985) \\
 &= 0.788
 \end{aligned}$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

credit_rating	P <sub>i</sub>	n <sub>i</sub>	P <sub>i</sub> + n <sub>i</sub>
excellent	3	3	6
fair	6	2	8

Info<sub>credit\_rating</sub>(D)

$$\begin{aligned}
 &= \frac{6}{14} I(3,3) + \frac{8}{14} I(6,2) \\
 &= \frac{6}{14} \left[ -\frac{3}{7} \log_2 \left( \frac{3}{7} \right) - \frac{3}{7} \log_2 \left( \frac{3}{7} \right) \right] + \\
 &\quad \frac{8}{14} \left[ -\frac{6}{7} \log_2 \left( \frac{6}{7} \right) - \frac{2}{7} \log_2 \left( \frac{2}{7} \right) \right] \\
 &= \frac{6}{14} (1) + \frac{8}{14} (0.811) \\
 &= 0.892
 \end{aligned}$$

Step 3 : Information gained von allen attributes

$$\text{Gain}(\text{age}) = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.940 - 0.938 = 0.002$$

$$\text{Gain}(\text{income}) = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.940 - 0.938 = 0.002$$

$$\text{Gain}(\text{student}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.940 - 0.788 = 0.152$$

$$\begin{aligned}
 \text{Gain}(\text{credit\_rating}) &= \text{Info}(D) - \text{Info}_{\text{credit\_rating}}(D) \\
 &= 0.940 - 0.892 = 0.048
 \end{aligned}$$

∴ Root Node ist student

Student

no

yes

Age	Income	Credit_rating	buys_computer
<=30	high	fair	no
<=30	high	excellent	no
>30	high	fair	yes
<=30	low	fair	yes
<=30	low	fair	no
>30	low	excellent	yes
>30	low	excellent	no

Age	Income	Credit_rating	buys_computer
<=30	high	fair	yes
>30	low	excellent	no
>30	high	excellent	yes
<=30	low	fair	yes
<=30	low	fair	yes
<=30	low	excellent	yes
>30	high	fair	yes

Student : yes

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

Step 1 : an expected info gain  
classify now

- Class P: buys\_computer = "yes" → 6
- Class N: buys\_computer = "no" → 1

Info student: yes(D)

$$= -\frac{6}{7} \log_2 \left( \frac{6}{7} \right) - \frac{1}{7} \log_2 \left( \frac{1}{7} \right)$$

$$= 0.592$$

Step 2: an info gain on attributes

age	P <sub>i</sub>	n <sub>i</sub>	P <sub>i</sub> +n <sub>i</sub>
<=30	4	0	4
>30	2	1	3

$$\begin{aligned} \text{Info}_{\text{age}}(D) &= \frac{4}{7} I(4,0) + \frac{3}{7} I(2,1) \\ &= \frac{4}{7} \left[ -\frac{4}{4} \log_2 \left( \frac{4}{4} \right) - \frac{0}{4} \log_2 \left( \frac{0}{4} \right) \right] + \\ &\quad \frac{3}{7} \left[ -\frac{2}{3} \log_2 \left( \frac{2}{3} \right) - \frac{1}{3} \log_2 \left( \frac{1}{3} \right) \right] \\ &= \frac{4}{7} (0) + \frac{3}{7} (0.918) \\ &= 0.393 \end{aligned}$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

income	$p_i$	$n_i$	$p_i + n_i$
high	3	0	3
low	3	1	4

Info<sub>income</sub>(D)

$$\begin{aligned}
 &= \frac{3}{7} I(3,0) + \frac{4}{7} I(3,1) \quad \text{琐屑忽略} \\
 &= \frac{3}{7} \left[ -\frac{3}{3} \log_2 \left( \frac{3}{3} \right) - \frac{0}{3} \log_2 \left( \frac{0}{3} \right) \right] + \\
 &\quad \frac{4}{7} \left[ -\frac{3}{4} \log_2 \left( \frac{3}{4} \right) - \frac{1}{4} \log_2 \left( \frac{1}{4} \right) \right] \\
 &= \frac{3}{7} (0) + \frac{4}{7} (0.811) \\
 &= 0.463
 \end{aligned}$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

credit_rating	$p_i$	$n_i$	$p_i + n_i$
excellent	2	1	3
fair	4	0	4

Info<sub>credit\_rating</sub>(D)

$$\begin{aligned}
 &= \frac{3}{7} I(2,1) + \frac{4}{7} I(4,0) \\
 &= \frac{3}{7} \left[ -\frac{2}{3} \log_2 \left( \frac{2}{3} \right) - \frac{1}{3} \log_2 \left( \frac{1}{3} \right) \right] + \\
 &\quad \frac{4}{7} \left[ -\frac{4}{4} \log_2 \left( \frac{4}{4} \right) - \frac{0}{4} \log_2 \left( \frac{0}{4} \right) \right] \\
 &= \frac{3}{7} (0.918) + \frac{4}{7} (0) \quad \text{琐屑忽略} \\
 &= 0.393
 \end{aligned}$$

Step 3: Information gained von Werten der Attribute

$$\text{Gain}(\text{age}) = \text{Info}_{\text{student:yes}}(D) - \text{Info}_{\text{age}}(D) = 0.592 - 0.393 = 0.199 \times$$

$$\text{Gain}(\text{income}) = \text{Info}_{\text{student:yes}}(D) - \text{Info}_{\text{income}}(D) = 0.592 - 0.463 = 0.129$$

$$\begin{aligned}\text{Gain}(\text{credit\_rating}) &= \text{Info}_{\text{student:yes}}(D) - \text{Info}_{\text{credit\_rating}}(D) \\ &= 0.592 - 0.393 = 0.199 \times\end{aligned}$$

Student : no

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

Step 1 : an expected info នៃការ  
classify នូវបាន

→ Class P : buys\_computer = "yes" → 3

→ Class N : buys\_computer = "no" → 4

Info student : yes (D)

$$= -\frac{3}{7} \log_2 \left( \frac{3}{7} \right) - \frac{4}{7} \log_2 \left( \frac{4}{7} \right)$$

$$= 0.985$$

Step 2 : និង info រវាងពាណិជ្ជកម្ម attributes

age	P <sub>i</sub>	n <sub>i</sub>	p <sub>i</sub> +n <sub>i</sub>
<=30	1	3	4
>30	2	1	3

$$\begin{aligned} \text{Info}_{\text{age}}(\text{D}) &= \frac{1}{7} I(1, 3) + \frac{3}{7} I(2, 1) \\ &= \frac{4}{7} \left[ -\frac{1}{4} \log_2 \left( \frac{1}{4} \right) - \frac{3}{4} \log_2 \left( \frac{3}{4} \right) \right] + \\ &\quad \frac{3}{7} \left[ -\frac{2}{3} \log_2 \left( \frac{2}{3} \right) - \frac{1}{3} \log_2 \left( \frac{1}{3} \right) \right] \\ &= \frac{4}{7} (0.811) + \frac{3}{7} (0.918) \\ &= 0.857 \end{aligned}$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

income	P <sub>i</sub>	n <sub>i</sub>	P <sub>i</sub> +n <sub>i</sub>
high	1	2	3
low	2	2	4

Info<sub>income</sub>(D)

$$\begin{aligned}
 &= \frac{3}{7} I(1,2) + \frac{4}{7} I(2,2) \\
 &= \frac{3}{7} \left[ -\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right) \right] + \\
 &\quad \frac{4}{7} \left[ -\frac{2}{4} \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \log_2 \left(\frac{1}{4}\right) \right] \\
 &= \frac{3}{7} (0.918) + \frac{4}{7} (-1) \\
 &= 0.965
 \end{aligned}$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

credit_rating	P <sub>i</sub>	n <sub>i</sub>	P <sub>i</sub> +n <sub>i</sub>
excellent	1	2	3
fair	2	2	4

Info<sub>credit\_rating</sub>(D)

$$\begin{aligned}
 &= \frac{3}{7} I(1,2) + \frac{4}{7} I(2,2) \\
 &= \frac{3}{7} \left[ -\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right) \right] + \\
 &\quad \frac{4}{7} \left[ -\frac{2}{4} \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \log_2 \left(\frac{1}{4}\right) \right] \\
 &= \frac{3}{7} (0.918) + \frac{4}{7} (-1) \\
 &= 0.965
 \end{aligned}$$

Step 3 : Information gained von verschiedenen attributen

$$\text{Gain}(\text{student: no}, \text{age}) = \text{Info}_{\text{student: no}}(D) - \text{Info}_{\text{age}}(D)$$
$$= 0.985 - 0.128 = 0.128 \quad \times$$

$$\text{Gain}(\text{student: no}, \text{income}) = \text{Info}_{\text{student: no}}(D) - \text{Info}_{\text{income}}(D)$$
$$= 0.985 - 0.965 = 0.02$$

$$\text{Gain}(\text{student: no}, \text{credit rating}) = \text{Info}_{\text{student: no}}(D) - \text{Info}_{\text{credit rating}}(D)$$
$$= 0.985 - 0.965 = 0.02$$

Gain(student: yes)	0.199
Gain(student: no)	0.128

→ First Node

Student

no

yes

Age	Income	Credit_rating	buys_computer
<=30	high	fair	no
<=30	high	excellent	no
>30	high	fair	yes
<=30	low	fair	yes
<=30	low	fair	no
>30	low	excellent	yes
>30	low	excellent	no

Age	Income	Credit_rating	buys_computer
<=30	high	fair	yes
>30	low	excellent	no
>30	high	excellent	yes
<=30	low	fair	yes
<=30	low	fair	yes
<=30	low	excellent	yes
>30	high	fair	yes

Student

no

yes

no  
no  
yes  
yes  
no  
yes

yes  
no  
yes  
yes  
yes  
yes

Student : Yes

From age

student	yes	age	<=30	
age	income	student	credit_rating	buys_computer
<=30	high	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes

∴

student	yes	age	>30	
age	income	student	credit_rating	buys_computer
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
>30	high	yes	fair	yes

Step 1 : an expected info នៅរដ្ឋមន្ត្រី  
Classify them

- Class P : buys\_computer = "yes" → 4
- Class N : buys\_computer = "no" → 0

Info student : yes, age  $\leq 30$  (D)  
 $= -\frac{4}{4} \log_2 \left(\frac{4}{4}\right) - \frac{0}{4} \log_2 \left(\frac{0}{4}\right)$   
 $= 0$

Step 1 : an expected info នៅរដ្ឋមន្ត្រី  
Classify them

- Class P : buys\_computer = "yes" → 2
- Class N : buys\_computer = "no" → 1

Info student : yes, age  $> 30$  (D)  
 $= -\frac{2}{3} \log_2 \left(\frac{2}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right)$   
 $= 0.918$

## Step 2: m info von beiden attributes

income	$P_i$	$n_i$	$P_i + n_i$
high	2	0	2
low	0	1	1

credit_rating	$P_i$	$n_i$	$P_i + n_i$
excellent	1	1	2
fair	1	0	1

Info student: yes, age > 30, income  $\text{(D)}$

$$\begin{aligned}
 &= \frac{2}{3} I(2,0) + \frac{1}{3} I(0,1) \\
 &= \frac{2}{3} \left[ -\frac{2}{2} \log_2 \left( \frac{2}{2} \right) - \frac{0}{2} \log_2 \left( \frac{0}{2} \right) \right] + \\
 &\quad \frac{1}{3} \left[ -\frac{0}{1} \log_2 \left( \frac{0}{1} \right) - \frac{1}{1} \log_2 \left( \frac{1}{1} \right) \right] \\
 &= \frac{2}{3} (0) + \frac{1}{3} (0) \\
 &= 0
 \end{aligned}$$

Info student: yes, age > 30, credit\_rating  $\text{(D)}$

$$\begin{aligned}
 &= \frac{2}{3} I(1,1) + \frac{1}{3} I(1,0) \\
 &= \frac{2}{3} \left[ -\frac{1}{2} \log_2 \left( \frac{1}{2} \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) \right] + \\
 &\quad \frac{1}{3} \left[ -\frac{1}{1} \log_2 \left( \frac{1}{1} \right) - \frac{0}{1} \log_2 \left( \frac{0}{1} \right) \right] \\
 &= \frac{2}{3} (1) + \frac{1}{3} (0) \\
 &= 0.667
 \end{aligned}$$

## Step 3: m information gained von beiden attributes

Gain(student: yes, age > 30, income)

$$\begin{aligned}
 &= \text{Info}_{\text{student: yes, age } > 30} \text{ (D)} - \text{Info}_{\text{student: yes, age } > 30, \text{ income}} \text{ (D)} \\
 &= 0.918 - 0 = 0.918
 \end{aligned}$$

Gain(student: yes, age > 30, credit\_rating)

$$\begin{aligned}
 &= \text{Info}_{\text{student: yes, age } > 30} \text{ (D)} - \text{Info}_{\text{student: yes, age } > 30, \text{ credit\_rating}} \text{ (D)} \\
 &= 0.918 - 0.667 = 0.251
 \end{aligned}$$

## ສິນຄາ credit\_rating

student	yes	credit	excellent	
age	income	student	credit_rating	buys_computer
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	yes	excellent	yes

Info  
S: yes, C = excellent  $I(D) = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) = 0.918$

Step 1: ມີ info ພະນັກລະ attributes

age	$P_i$	$n_i$	$P_i + n_i$
$\leq 30$	1	0	1
$> 30$	1	1	2

Info  
S: yes, C = excellent, age  $I(D)$

$$\begin{aligned}
 &= \frac{1}{3} I(1, 0) + \frac{2}{3} I(1, 1) \\
 &= \frac{1}{3} \left[ -\frac{1}{1} \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \log_2\left(\frac{0}{1}\right) \right] + \\
 &\quad \frac{2}{3} \left[ -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] \\
 &= \frac{1}{3} (0) + \frac{2}{3} (-1) \\
 &= -0.667
 \end{aligned}$$

income	$P_i$	$n_i$	$P_i + n_i$
high	1	0	1
low	1	1	2

Info  
S: yes, C = excellent, income  $I(D)$

$$\begin{aligned}
 &= \frac{1}{3} I(1, 0) + \frac{2}{3} I(1, 1) \\
 &= \frac{1}{3} \left[ -\frac{1}{1} \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \log_2\left(\frac{0}{1}\right) \right] + \\
 &\quad \frac{2}{3} \left[ -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] \\
 &= \frac{1}{3} (0) + \frac{2}{3} (-1) \\
 &= -0.667
 \end{aligned}$$

Step 1: ມີ expected info ກໍາລັງ  
classify ຜອນ

↗ Class P: buys\_computer = "yes"  $\rightarrow 2$

↗ Class N: buys\_computer = "no"  $\rightarrow 1$

Step 3: Information gained von kategorialen attributen

$\text{Gain}(s: \text{yes}, c = \text{excellent}, \text{age})$

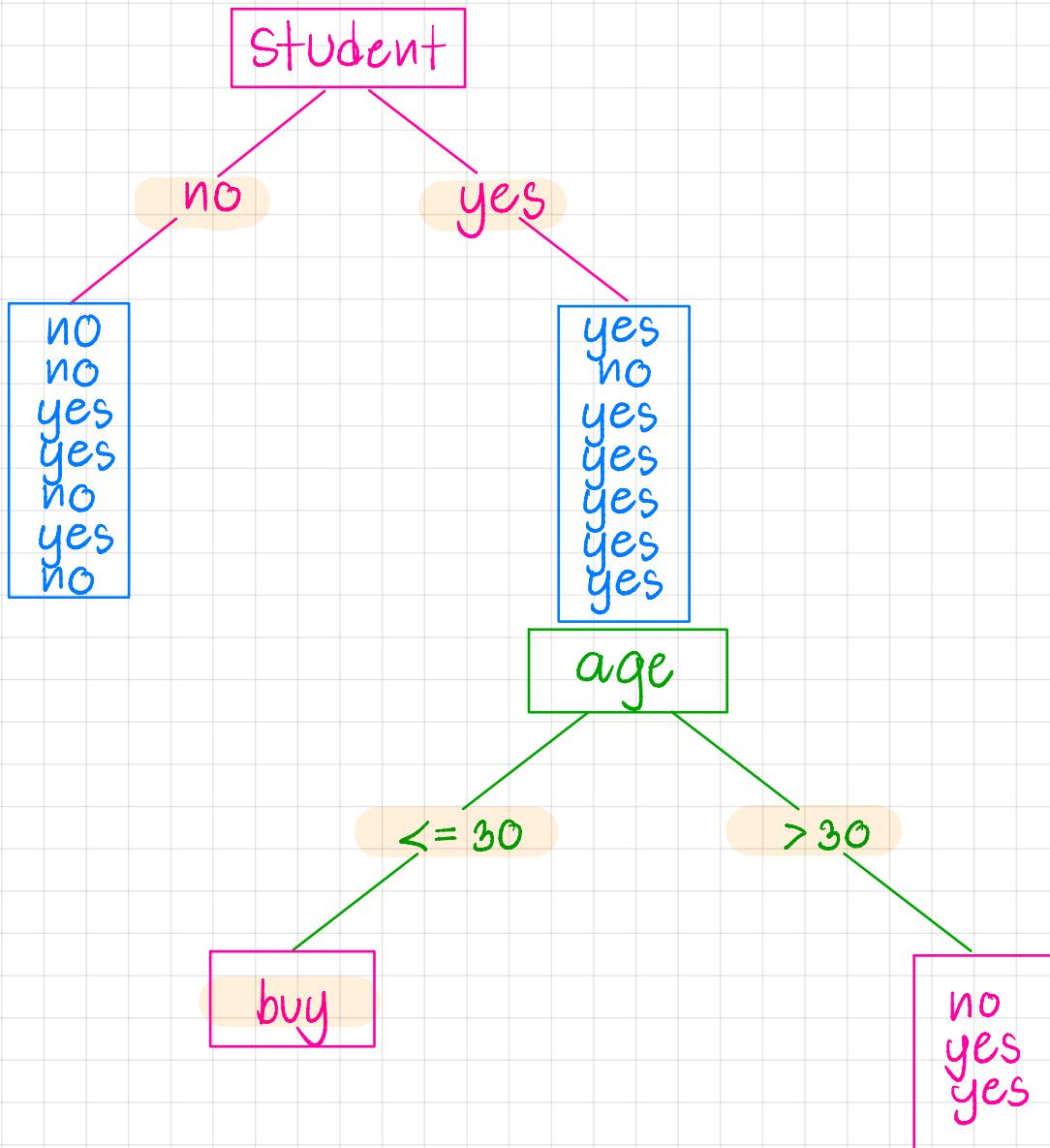
$$= \text{Info}_{s: \text{yes}, c = \text{excellent}}(D) - \text{Info}_{s: \text{yes}, c = \text{excellent}, \text{age}}(D)$$
$$= 0.918 - 0.667 = 0.251$$

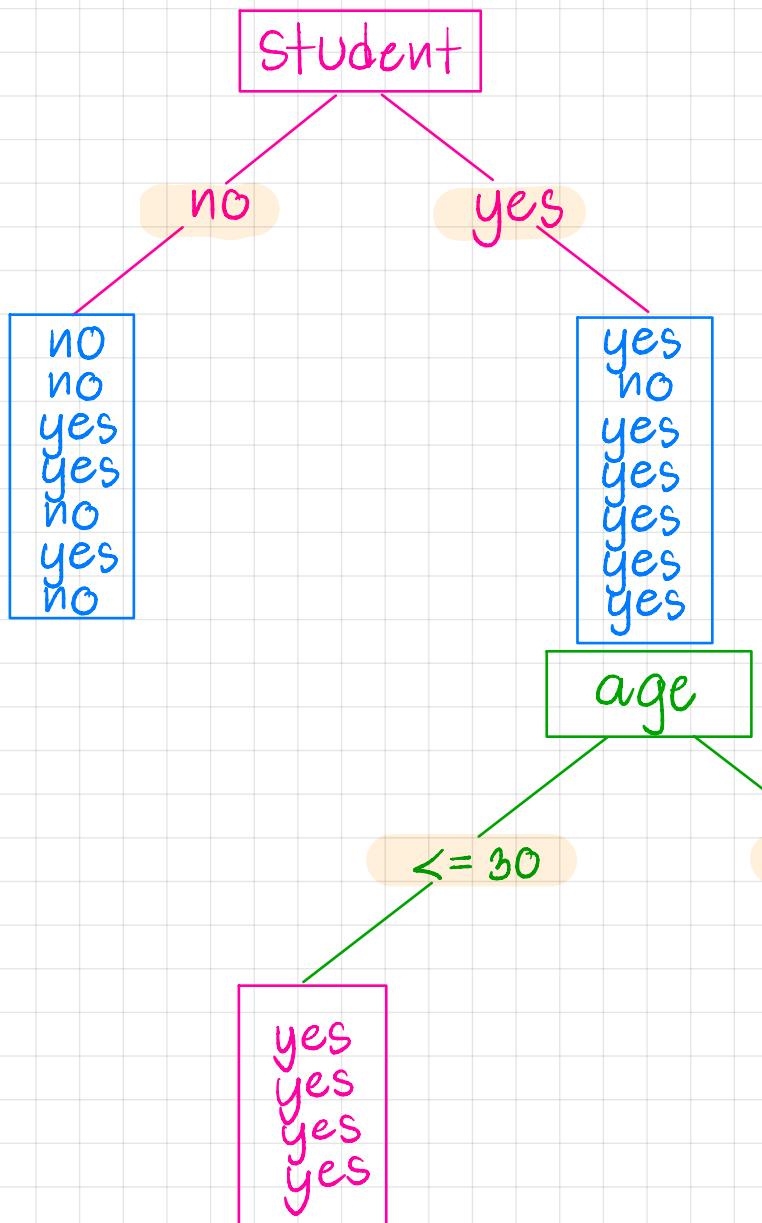
$\text{Gain}(s: \text{yes}, c = \text{excellent}, \text{income})$

$$= \text{Info}_{s: \text{yes}, c = \text{excellent}}(D) - \text{Info}_{s: \text{yes}, c = \text{excellent}, \text{income}}(D)$$
$$= 0.918 - 0.667 = 0.251$$

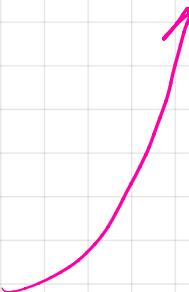
$\text{Gain}(\text{student: yes, age} > 30, \text{income})$	0.918
$\text{Gain}(\text{student: yes, age} > 30, \text{credit\_rating})$	0.251
$\text{Gain}(s: \text{yes}, c = \text{excellent}, \text{age})$	0.251
$\text{Gain}(s: \text{yes}, c = \text{excellent}, \text{income})$	0.251

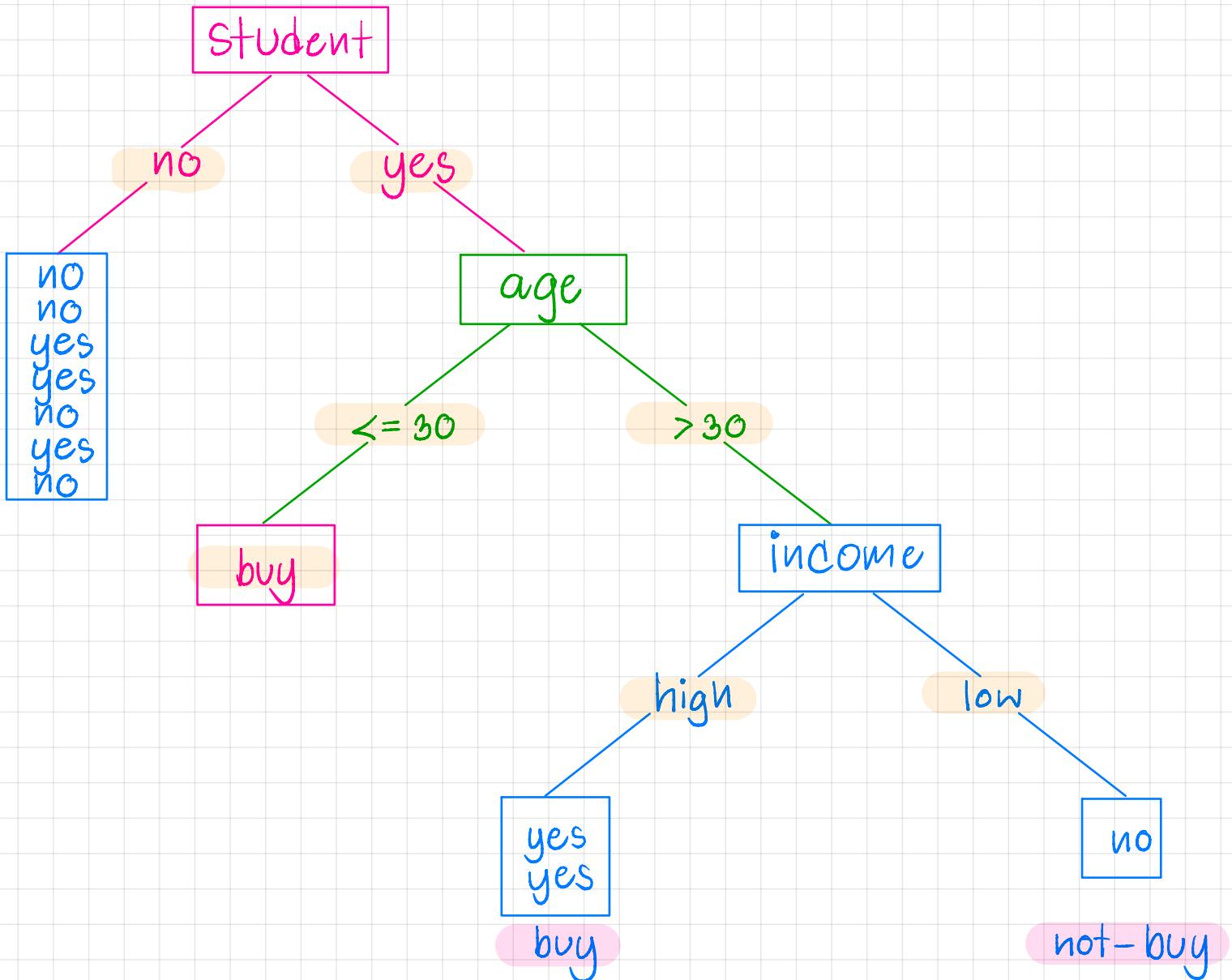
→ Second node

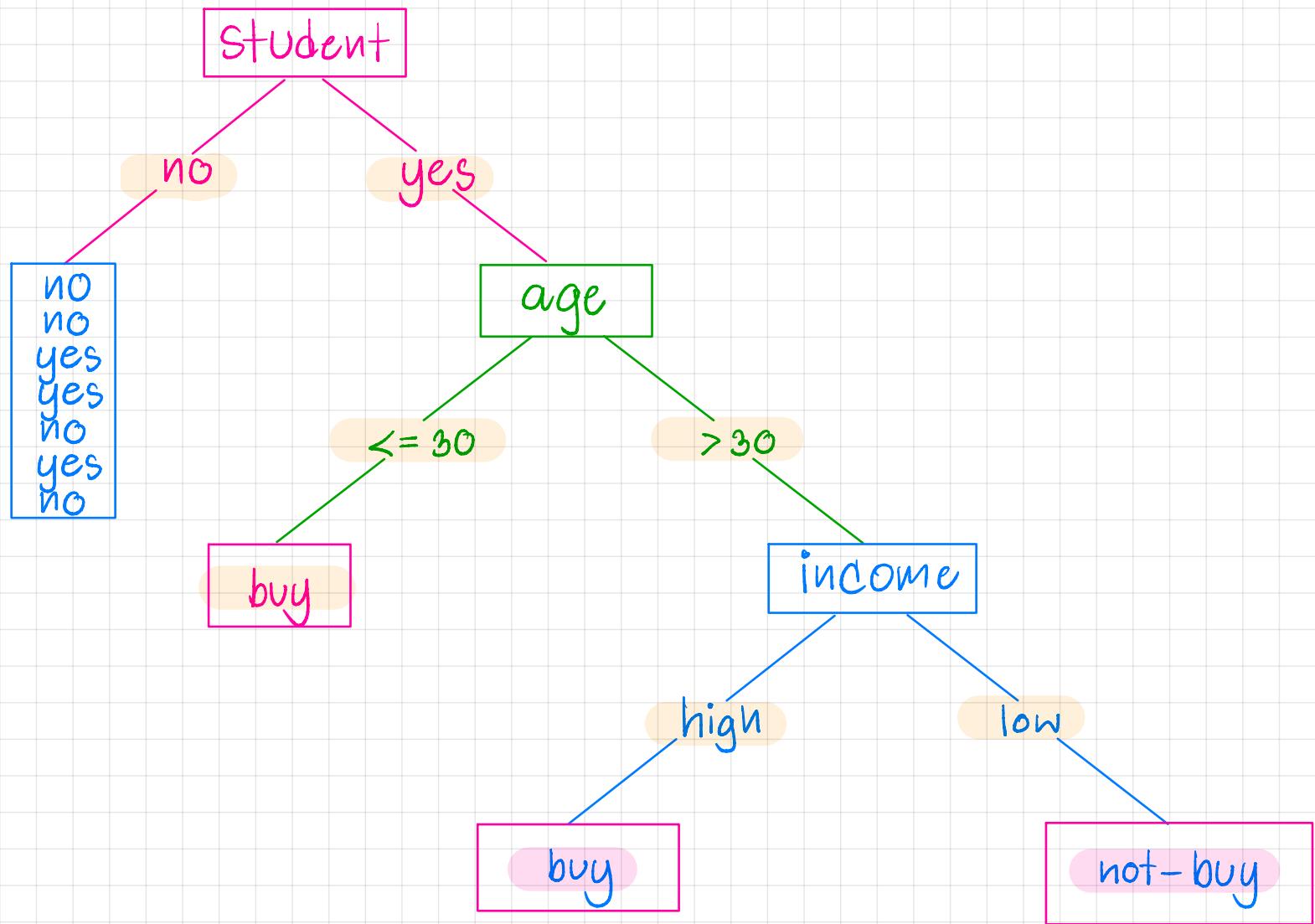




student	yes	age	>30	income = low
age	income	student	credit_rating	buys_computer
>30	low	yes	excellent	no
student	yes	age	>30	income = high
age	income	student	credit_rating	buys_computer
>30	high	yes	excellent	yes
>30	high	yes	fair	yes







Student: NO

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

Step 1: expected info 用于 classify now

↳ Class P: buy\_computer = "yes" → 3

↳ Class N: buy\_computer = "no" → 4

Info student: no(D)

$$= -\frac{3}{7} \log_2 \left( \frac{3}{7} \right) - \frac{4}{7} \log_2 \left( \frac{4}{7} \right)$$

$$= 0.9852$$

Step 2: w/ info 有关于 attributes

age	$P_i$	$n_i$	$P_i + n_i$
$\leq 30$	3	1	4
$> 30$	1	2	3

$$\text{Info}_{age}(D) = \frac{4}{7} I(1,3) + \frac{3}{7} I(2,1)$$

$$= \frac{4}{7} \left[ -\frac{1}{4} \log_2 \left( \frac{1}{4} \right) - \frac{3}{4} \log_2 \left( \frac{3}{4} \right) \right] + \frac{3}{7} \left( -\frac{2}{3} \log_2 \left( \frac{2}{3} \right) - \frac{1}{3} \log_2 \left( \frac{1}{3} \right) \right)$$

$$= \frac{4}{7} (0.8113) + \frac{3}{7} (0.91828)$$

$$= 0.8571$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

income	$p_i$	$n_i$	$p_i + n_i$
high	1	2	3
low	2	2	4

$$\begin{aligned}
 \text{Info}_{\text{income}}(D) &= \frac{3}{7} I(2,1) + \frac{4}{7} I(2,2) \\
 &= \frac{3}{7} \left[ -\frac{2}{3} \log_2 \left( \frac{2}{3} \right) - \frac{1}{3} \log_2 \left( \frac{1}{3} \right) \right] + \frac{4}{7} \left( -\frac{2}{4} \log_2 \left( \frac{2}{4} \right) - \frac{2}{4} \log_2 \left( \frac{2}{4} \right) \right) \\
 &= \frac{3}{7} (0.91828) + \frac{4}{7} (1) \\
 &\approx 0.9649
 \end{aligned}$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

credit_rating	$p_i$	$n_i$	$p_i + n_i$
excellent	1	2	3
fair	2	2	4

$$\begin{aligned}
 \text{Info}_{\text{credit\_rating}}(D) &= \frac{3}{7} I(2,1) + \frac{4}{7} I(2,2) \\
 &= \frac{3}{7} \left[ -\frac{2}{3} \log_2 \left( \frac{2}{3} \right) - \frac{1}{3} \log_2 \left( \frac{1}{3} \right) \right] + \frac{4}{7} \left( -\frac{2}{4} \log_2 \left( \frac{2}{4} \right) - \frac{2}{4} \log_2 \left( \frac{2}{4} \right) \right) \\
 &= \frac{3}{7} (0.91828) + \frac{4}{7} (1) \\
 &\approx 0.9649
 \end{aligned}$$

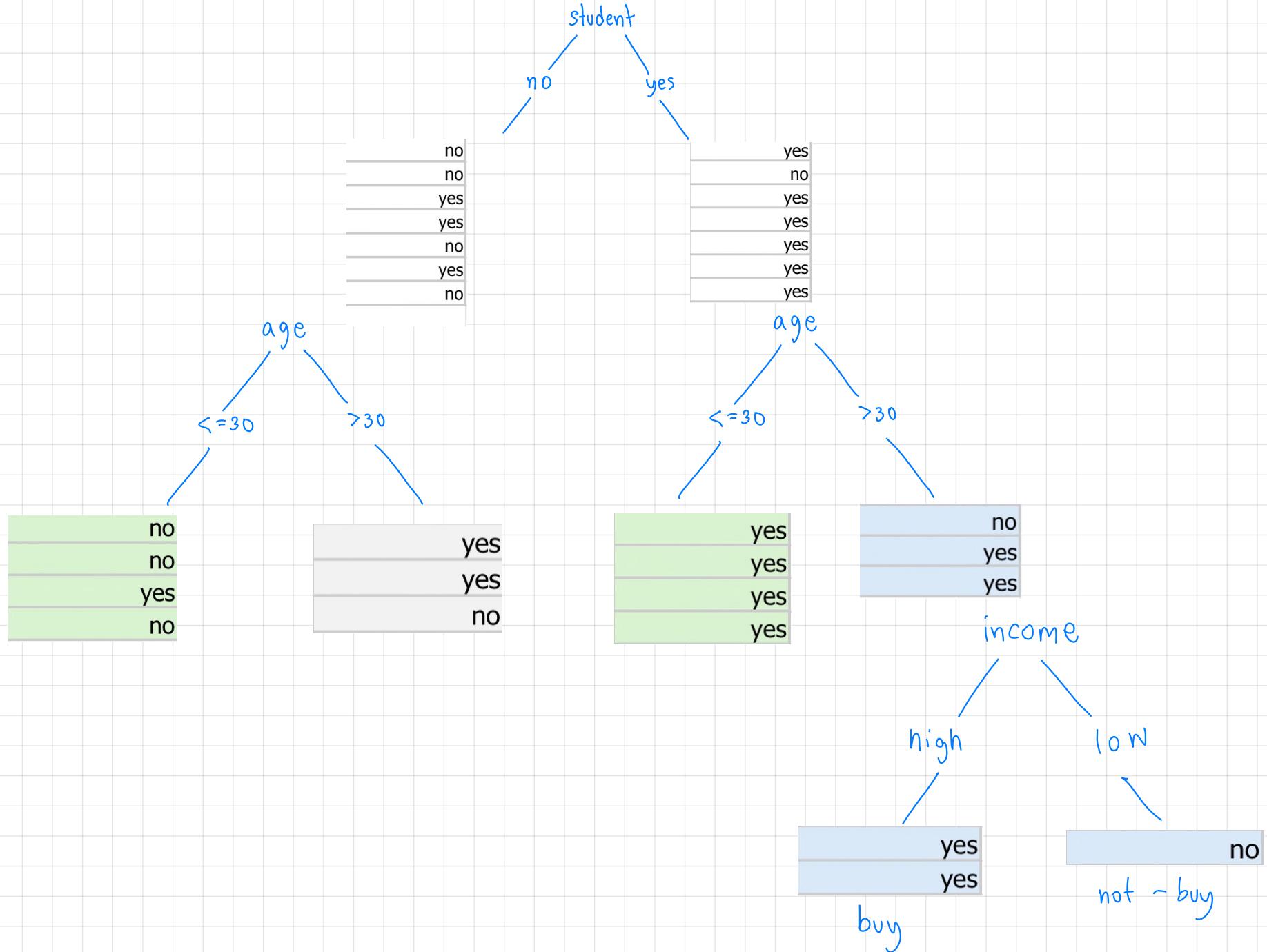
Step 3: in information gained von max attributes

$$\text{Gain}(\text{age}) = \text{Info}_{\text{Student: no}(D)} - \text{Info}_{\text{age}}(D) = 0.9852 - 0.8571 = 0.1281 \quad \checkmark$$

$$\text{Gain}(\text{income}) = \text{Info}_{\text{Student: no}(D)} - \text{Info}_{\text{income}}(D) = 0.9852 - 0.9649 = 0.0203$$

$$\text{Gain}(\text{credit-rating}) = \text{Info}_{\text{Student: no}(D)} - \text{Info}_{\text{credit-rating}}(D) = 0.9852 - 0.9649 = 0.0203$$

Gain( Student : yes )	0.199	→ First Node
Gain( Student : no )	0.128	



Student : NO แล้ว age

F) Age :  $\leq 30$

student	no	age $\leq 30$			
age	income	student	credit_rating	buys_computer	
$\leq 30$	high	no	fair	no	
$\leq 30$	high	no	excellent	no	
$\leq 30$	low	no	fair	yes	
$\leq 30$	low	no	fair	no	

age	income	student	credit_rating	buys_computer
$\leq 30$	high	no	fair	no
$\leq 30$	high	no	excellent	no
$> 30$	high	no	fair	yes
$\leq 30$	low	no	fair	yes
$\leq 30$	high	yes	fair	yes
$> 30$	low	yes	excellent	no
$> 30$	high	yes	excellent	yes
$\leq 30$	low	no	fair	no
$\leq 30$	low	yes	fair	yes
$\leq 30$	low	yes	fair	yes
$\leq 30$	low	yes	excellent	yes
$> 30$	low	no	excellent	yes
$> 30$	high	yes	fair	yes
$> 30$	low	no	excellent	no

Step 1 : expected info ที่จะ Classify นิ้ว

Class P : buy\_computer = " yes"  $\rightarrow 1$

Class N : buy\_computer = " no"  $\rightarrow 3$

Info

student : no, age  $\leq 30$  : no (D)

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

$$= 0.8113$$

Step 2 : ณ info ของลักษณะ attributes

income	$p_i$	$n_i$	$p_i + n_i$
high	0	2	2
low	1	1	2

$$\text{Info}_{\text{income}} (D) = \frac{2}{4} I(1,0) + \frac{2}{4} I(1,1)$$

$$= \frac{2}{4} \left( -\frac{1}{2} \log_2 \left( \frac{1}{2} \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) \right) + \frac{2}{4} \left[ -\frac{0}{2} \log_2 \left( \frac{0}{2} \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) \right]$$

$$= \frac{2}{4} (1) + \frac{2}{4} (0)$$

$$= 0.5$$

ดูว่า 70%

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

credit_rating	$p_i$	$n_i$	$p_i + n_i$
excellent	0	1	1
fair	1	2	3

$$\begin{aligned}
 \text{Info}_{\text{credit\_rating}}(D) &= \frac{3}{4} I(2,1) + \frac{1}{4} I(1,0) \\
 &= \frac{3}{4} \left[ -\frac{2}{3} \log_2 \left( \frac{2}{3} \right) - \frac{1}{3} \log_2 \left( \frac{1}{3} \right) \right] + \frac{1}{4} \left( -\frac{1}{1} \log_2 \left( \frac{1}{1} \right) - \frac{0}{1} \log_2 \left( \frac{0}{1} \right) \right) \\
 &= \frac{3}{4} (0.91828) + \frac{1}{4} (0) \\
 &= 0.6887
 \end{aligned}$$

s: u 77%

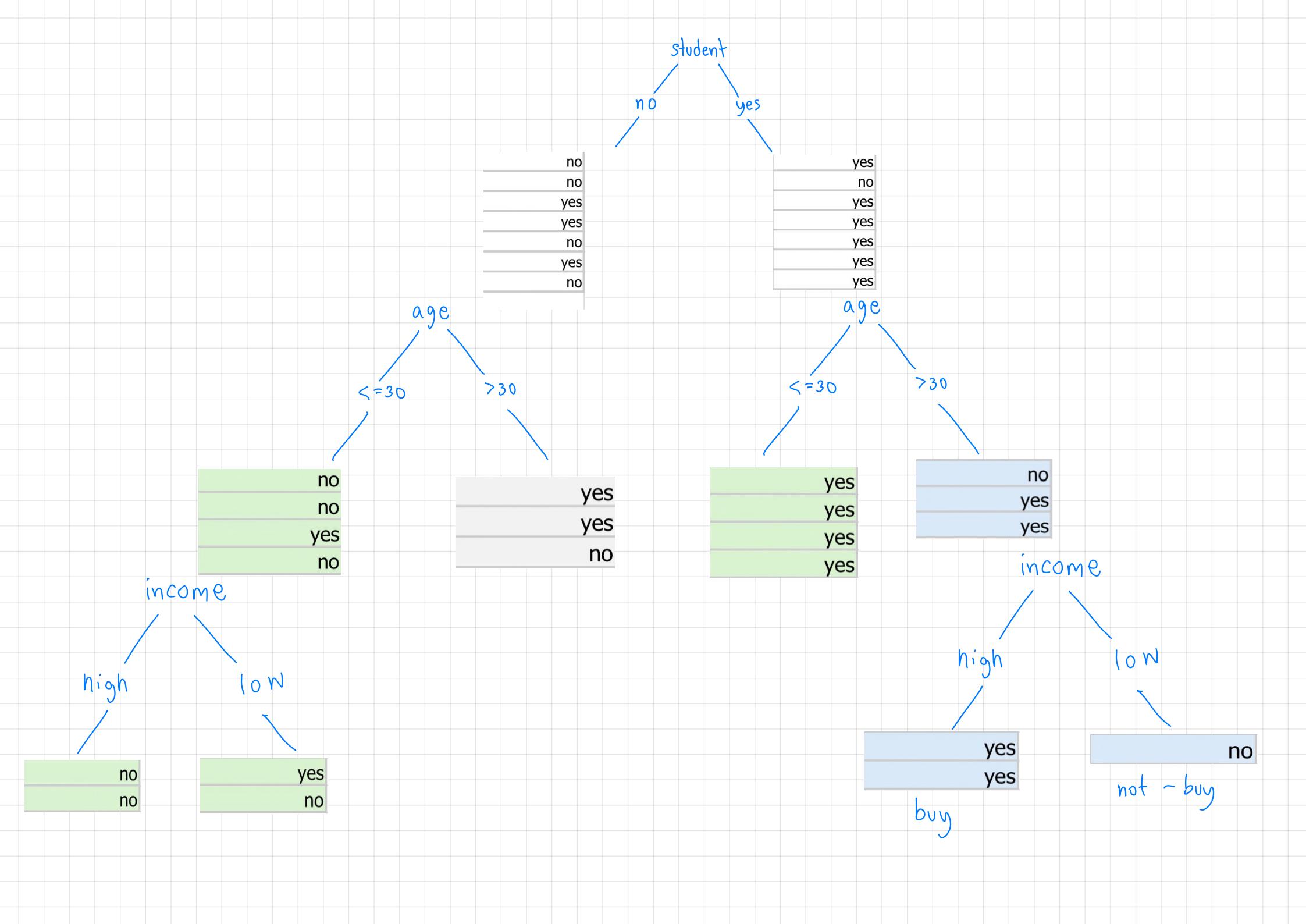
Step 3: calculate information gained using attributes

$$\text{Gain}(\text{student: no}, \text{age} \leq 30, \text{income}) = \text{Info}_{\text{student: no}, \text{age} \leq 30: \text{no}(D)}$$

$$- \text{Info}_{\text{income}}(D) = 0.8113 - 0.5 = 0.3113 \quad \checkmark$$

$$\text{Gain}(\text{student: no}, \text{age} \leq 30, \text{credit\_rating}) = \text{Info}_{\text{student: no}, \text{age} \leq 30: \text{no}(D)}$$

$$- \text{Info}_{\text{credit\_rating}}(D) = 0.8113 - 0.6887 = 0.1226$$



(F) Age : > 30

student	no	age > 30		
age	income	student	credit_rating	buys_computer
>30	high	no	fair	yes
>30	low	no	excellent	yes
>30	low	no	excellent	no

Step 1 : expected info 用于 classify now

Class P : buy\_computer = "yes" → 2

Class N : buy\_computer = "no" → 1

Info  
student: no, age > 30 : no (D)

$$= -\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right)$$

$$= 0.918296$$

Step 2 : w/ info 评估各属性

income	$p_i$	$n_i$	$p_i + n_i$
high	1	0	1
low	1	1	2

$$\begin{aligned} \text{Info}_{\text{income}}(D) &= \frac{1}{3} I(0,1) + \frac{2}{3} I(1,1) \\ &= \frac{1}{3} \left[ -\frac{0}{1} \log_2 \left(\frac{0}{1}\right) - \frac{1}{1} \log_2 \left(\frac{1}{1}\right) \right] + \frac{2}{3} \left( -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{2} \log_2 \left(\frac{1}{2}\right) \right) \\ &= \frac{1}{3}(0) + \frac{2}{3}(1) \\ &= 0.667 \end{aligned}$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
>30	high	no	fair	yes
<=30	low	no	fair	yes
<=30	high	yes	fair	yes
>30	low	yes	excellent	no
>30	high	yes	excellent	yes
<=30	low	no	fair	no
<=30	low	yes	fair	yes
<=30	low	yes	fair	yes
<=30	low	yes	excellent	yes
>30	low	no	excellent	yes
>30	high	yes	fair	yes
>30	low	no	excellent	no

credit_rating	$p_i$	$n_i$	$p_i + n_i$
excellent	1	1	2
fair	1	0	1

$$\begin{aligned}
 \text{Info}_{\text{credit\_rating}}(D) &= \frac{1}{3} I(0,1) + \frac{2}{3} I(1,1) \\
 &= \frac{1}{3} \left[ -\frac{0}{1} \log_2 \left( \frac{0}{1} \right) - \frac{1}{1} \log_2 \left( \frac{1}{1} \right) \right] + \frac{2}{3} \left( -\frac{1}{2} \log_2 \left( \frac{1}{2} \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) \right) \\
 &= \frac{1}{3}(0) + \frac{2}{3}(1) \\
 &\approx 0.667
 \end{aligned}$$

*s: u7ai7n*

Step 3: w information gained von mir: attributes

$$\text{Gain}(\text{student: no}, \text{age} > 30, \text{income}) = \text{Info}_{\text{student: no, age} > 30: \text{no}(D)} - \text{Info}_{\text{income}(D)} = 0.918 - 0.667 = 0.251$$

$$\text{Gain}(\text{student: no}, \text{age} > 30, \text{credit\_rating}) = \text{Info}_{\text{student: no, age} > 30: \text{no}(D)} - \text{Info}_{\text{credit\_rating}(D)} = 0.918 - 0.667 = 0.251$$

