HW 4

$$Info(D) = I(8,4) = -\frac{8}{12}\log_2\left(\frac{8}{12}\right) - \frac{4}{12}\log_2\left(\frac{4}{12}\right) = 0.9183$$

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

∴ จากการคำนวณ สรุปได้ว่า

Gain(age) = 0.1804 → เลือก Gain(age)

Gain(credit_rating) = 0.1686

Gain(student) = 0.0933

Gain(income) = 0.0546

age	p_i	n_i	$I(p_i, n_i)$
<=30	2	2	1
31…40	3	0	0
>40	3	2	0.971

$$I(2,2) = -\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right) = 1$$
$$I(3,0) = -\frac{3}{3}\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right) = 0$$
$$I(3,2) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$$

$$Info_{age}(D) = \frac{4}{12}I(2,2) + \frac{3}{12}I(3,0) + \frac{5}{12}I(3,2)$$
$$= \frac{4}{12}\cdot 1 + \frac{3}{12}\cdot 0 + \frac{5}{12}\cdot 0.971$$
$$= 0.7379$$

$$Gain(age) = Info(D) - Info_{age}(D)$$
$$= 0.9183 - 0.7379$$
$$= 0.1804$$

income	p_i	n_i	$I(p_i, n_i)$
High	2	2	1
Medium	4	1	0.7219
low	2	1	0.9183

$$I(2,2) = -\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right) = 1$$
$$I(4,1) = -\frac{4}{5}\log_2\left(\frac{4}{5}\right) - \frac{1}{5}\log_2\left(\frac{1}{5}\right) = 0.7219$$
$$I(2,1) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.9183$$

$$Info_{income}(D) = \frac{4}{12}I(2,2) + \frac{5}{12}I(4,1) + \frac{3}{12}I(2,1)$$
$$= \frac{4}{12}\cdot 1 + \frac{5}{12}\cdot 0.7219 + \frac{3}{12}\cdot 0.9183$$
$$= 0.8637$$

$$Gain(income) = Info(D) - Info_{income}(D)$$
$$= 0.9183 - 0.8637$$
$$= 0.0546$$

student	p_i	n_i	$I(p_i, n_i)$
Yes	5	1	0.65
No	3	3	1

$$I(5,1) = -\frac{5}{6}\log_2\left(\frac{5}{6}\right) - \frac{1}{6}\log_2\left(\frac{1}{6}\right) = 0.6500$$
$$I(3,3) = -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right) = 1$$

$$Info_{student}(D) = \frac{6}{12}I(5,1) + \frac{6}{12}I(3,3)$$
$$= \frac{6}{12}\cdot 0.6500 + \frac{6}{12}\cdot 1$$
$$= 0.825$$

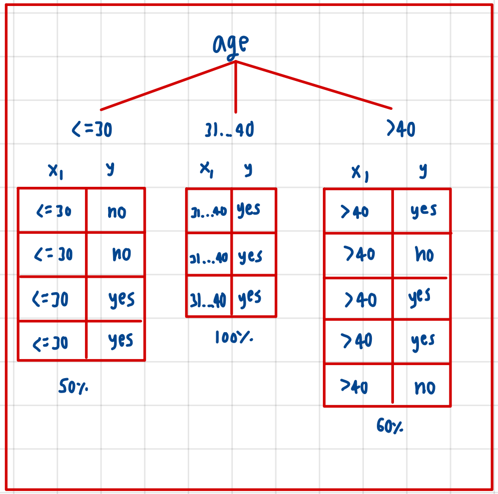
$$Gain(student) = Info(D) - Info_{student}(D)$$
$$= 0.9183 - 0.825$$
$$= 0.0933$$

Credit_rating	p_i	n_i	$I(p_i, n_i)$
fair	6	1	0.5917
excellent	2	3	0.9710

$$I(6,1) = -\frac{6}{7}\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right) = 0.5917$$
$$I(2,3) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.9710$$

$$Info_{credit\_rating}(D) = \frac{7}{12}I(6,1) + \frac{5}{12}I(2,3)$$
$$= \frac{7}{12}\cdot 0.5917 + \frac{5}{12}\cdot 0.9710$$
$$= 0.7497$$

$$Gain(credit\_rating) = Info(D) - Info_{credit\_rating}(D)$$
$$= 0.9183 - 0.7497$$
$$= 0.1686$$



age <=30 เลือก Gain(income), Gain(student)

age >30 เลือก Gain(credit_rating)

age <=30

income (x_2)	student (x_3)	credit_rating (x_4)	buys_computer (y)
high	no	fair	no
high	no	excellent	no
low	yes	fair	yes
medium	yes	excellent	yes

$Info(D_1 <= 30) = I(2,2) = -\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right) = 1$

income	p_i	n_i	$I(p_i, n_i)$
High	0	2	0
Medium	1	0	0
low	1	0	0

$I(0,2) = -\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right) = 0$

$I(1,0) = -\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log_2\left(\frac{0}{1}\right) = 0$

$I(1,0) = -\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log_2\left(\frac{0}{1}\right) = 0$

$Info_{income}(D <= 30) = \frac{2}{4} I(0,2) + \frac{1}{4} I(1,0) + \frac{1}{4} I(1,0)$

$= \frac{2}{4} \cdot 0 + \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 0$

$= 0$

$Gain(income) = Info(D_1 <= 30) - Info_{income}(D <= 30)$

$= 1 - 0$

$= 1$

student	p_i	n_i	$I(p_i, n_i)$
Yes	2	0	0
No	0	2	0

$I(2,0) = -\frac{2}{2}\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\log_2\left(\frac{0}{2}\right) = 0$

$I(0,2) = -\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right) = 0$

$Info_{student}(D_1 <= 30) = \frac{2}{4} I(2,0) + \frac{2}{4} I(0,2)$

$= \frac{2}{4} \cdot 0 + \frac{2}{4} 0$

$= 0$

$Gain(student) = Info(D_1 <= 30) - Info_{student}(D_1 <= 30)$

$= 1 - 0$

$= 1$

Credit_rating	p_i	n_i	$I(p_i, n_i)$
fair	1	1	1
excellent	1	1	1

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

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

$Info_{credit\_rating}(D <= 30) = \frac{2}{4} I(1,1) + \frac{2}{4} I(1,1)$

$= \frac{2}{4} + \frac{2}{4}$

$= 1$

$Gain(credit\_rating) = Info(D <= 30) - Info_{credit\_rating}(D <= 30)$

$= 1 - 1$

$= 0$

age >40

income (x_2)	student (x_3)	credit_rating (x_4)	buys_computer (y)
medium	no	fair	yes
low	yes	fair	yes
low	yes	excellent	no
medium	yes	fair	yes
medium	no	excellent	no

$Info(D_1 > 40) = I(3,2) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$

income	p_i	n_i	$I(p_i, n_i)$
medium	2	1	0.971
low	1	1	1

$I(2,1) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.9183$

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

$Info_{income}(D_1 > 40) = \frac{3}{5} I(3,1) + \frac{2}{5} I(1,1)$

$= \frac{3}{5} \cdot 0.971 + \frac{2}{5} \cdot 1$

$= 0.9826$

$Gain(income) = Info(D_1 > 40) - Info_{income}(D > 40)$

$= 0.971 - 0.9183$

$= 0.0527$

student	p_i	n_i	$I(p_i, n_i)$
Yes	2	1	0.971
No	1	1	1

$I(2,1) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.9183$

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

$Info_{student}(D_1 > 40) = \frac{3}{5} I(2,1) + \frac{2}{5} I(1,1)$

$= \frac{3}{5} \cdot 0.971 + \frac{2}{5}(1)$

$= 0.9826$

$Gain(student) = Info(D_1 > 40) - Info_{student}(D_1 > 40)$

$= 0.971 - 0.9183$

$= 0.0527$

Credit_rating	p_i	n_i	$I(p_i, n_i)$
fair	3	0	0
excellent	0	2	0

$I(3,0) = -\frac{3}{3}\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right) = 0$

$I(0,2) = -\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right) = 0$

$Info_{credit\_rating}(D_1 > 40) = \frac{3}{5} I(3,0) + \frac{2}{5} I(0,2)$

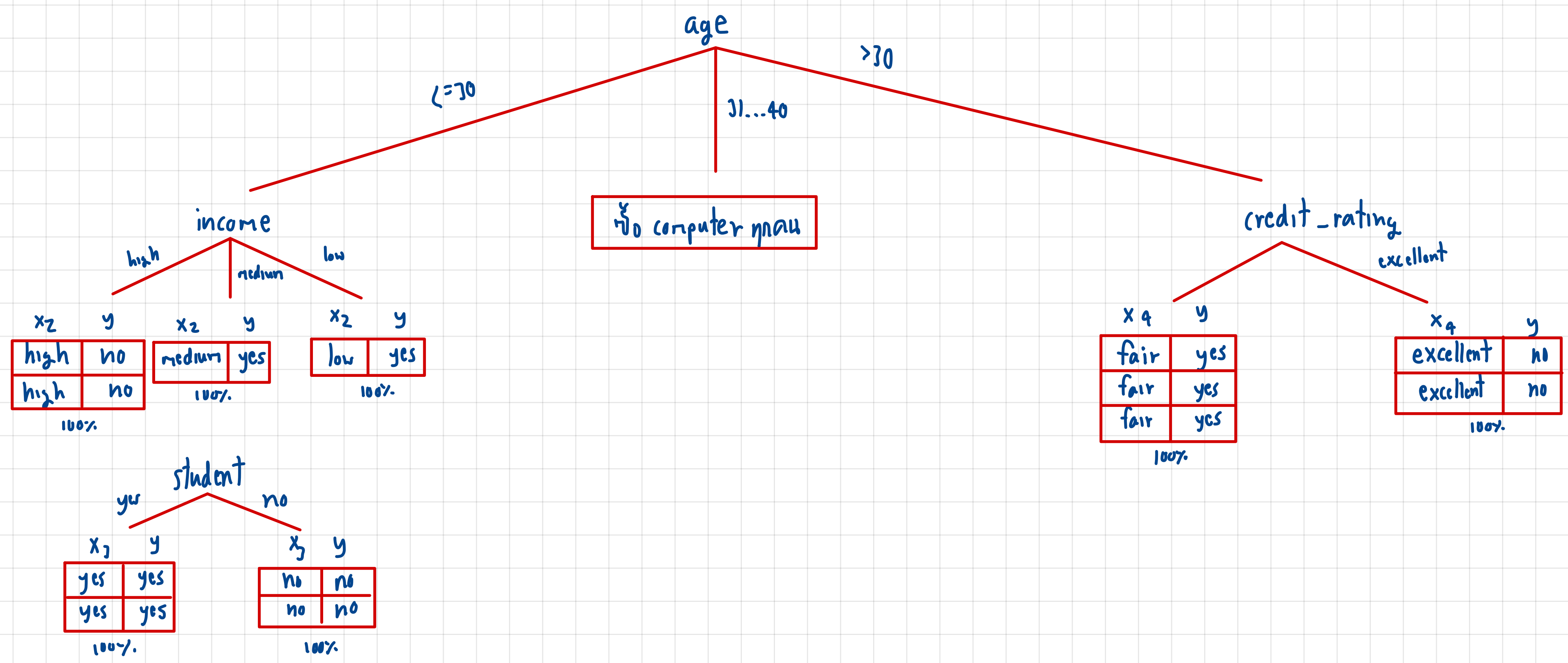
$= \frac{3}{5} \cdot 0 + \frac{2}{5} \cdot 0$

$= 0$

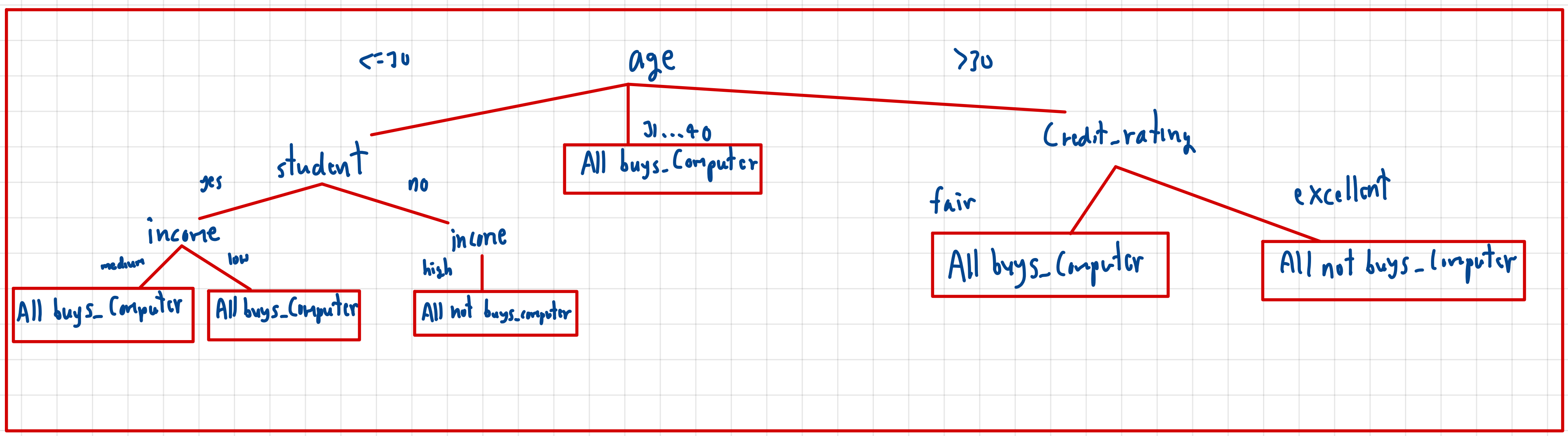
$Gain(credit\_rating) = Info(D_1 > 40) - Info_{credit\_rating}(D > 40)$

$= 0.971 - 0$

$= 0.971$



age
<=30
31...40
>30
income
ซื้อ computer ทุกคน
credit_rating
high
medium
low
excellent
x2 y
high no
high no
100%
x2 y
medium yes
100%
x2 y
low yes
100%
x4 y
fair yes
fair yes
fair yes
100%
x4 y
excellent no
excellent no
100%
student
yes
no
x3 y
yes yes
yes yes
100%
x3 y
no no
no no
100%



age
<=30
31...40
>30
student
yes
no
All buys_Computer
Credit_rating
fair
excellent
income
medium
low
income
high
All buys_Computer
All buys_Computer
All not buys_computer
All buys_Computer
All not buys_Computer