HW 2

นางสาว พิมพ์พิกา ยอดศรี
643020511-3

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
31...40	low	yes	excellent	yes
<=30	medium	no	fair	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

①

buys_computer

$$\text{Info}(D) = I\overset{Y\ N}{(9,5)}$$

$$= \frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right)$$

$$\therefore \text{Info}(D) = 0.940$$

$$\text{Info}_{age}(D) = \frac{5}{14}I(2,3) + \frac{4}{14}I(4,0) + \frac{5}{14}I(3,2)$$

$$= \frac{5}{14}\left[-\frac{3}{2}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right] + \frac{4}{14}\left[-\frac{4}{4}\log_2\left(\frac{4}{4}\right) - \frac{0}{4}\log_2\left(\frac{0}{4}\right)\right] +$$

$$\frac{5}{14}\left[-\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right]$$

$$\therefore \text{Info}_{age}(D) = 0.694$$

$$\text{Info}_{income}(D) = \frac{4}{14}I(2,2) + \frac{6}{14}I(4,2) + \frac{4}{14}I(3,1)$$

$$= \frac{4}{14}\left[-\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right] + \frac{6}{14}\left[-\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right] +$$

$$\frac{4}{14}\left[-\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right]$$

$$\therefore \text{Info}_{income}(D) = 0.911$$

$$\text{Info}_{student}(D) = \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4)$$

$$= \frac{7}{14}\left[-\frac{6}{7}\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right)\right]$$

$$\therefore \text{Info}_{student}(D) = 0.892$$

$$\text{Info}_{credit}(D) = \frac{6}{14}I(3,3) + \frac{8}{14}I(6,2)$$

$$= \frac{6}{14}\left[-\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right] + \frac{8}{14}\left[-\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right)\right]$$
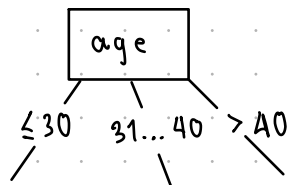
$$\therefore \text{Info}_{credit}(D) = 0.892$$

② Gain (age) = Info(D) − $\text{Info}_{age}(D)$ = 0.940 − 0.694 = 0.246 — เยอะสุด ∴ จึงเป็น root node

Gain (income) = Info(D) − $\text{Info}_{income}(D)$ = 0.940 − 0.911 = 0.029

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

Gain(credit_rating) = Info(D) − $\text{Info}_{credit\_rating}(D)$ = 0.048

จึง age แบ่ง

① age ≤ 30

age
≤30   31...40   >40

$$\text{Info}(D) = I(\overset{Y\ N}{2,3})$$

$$= -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)$$

$$\therefore \text{Info}(D) = 0.9710$$

$$\text{Info}_{income}(D) = \frac{2}{5}I(0,2) + \frac{2}{5}I(1,1) + \frac{1}{5}I(1,0)$$

$$= \frac{2}{5}\left[-\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] +$$

$$\frac{1}{5}\left[-\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log_2\left(\frac{0}{1}\right)\right]$$

$$\therefore \text{Info}_{income}(D) = 0.4$$

$$\text{Info}_{\text{student}}(D) = \frac{2}{5} I(2,0) + \frac{3}{5} I(0,3)$$

$$= \frac{2}{5}\left[-\frac{2}{2}\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\log_2\left(\frac{0}{2}\right)\right] + \frac{3}{5}\left[-\frac{0}{3}\log_2\left(\frac{0}{3}\right) - \frac{3}{3}\log_2\left(\frac{3}{3}\right)\right]$$

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

$$\text{Info}_{\text{credit\_r}}(D) = \frac{3}{5} I(1,2) + \frac{2}{5}(1,1)$$

$$= \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] + \frac{3}{5}\left[-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right]$$
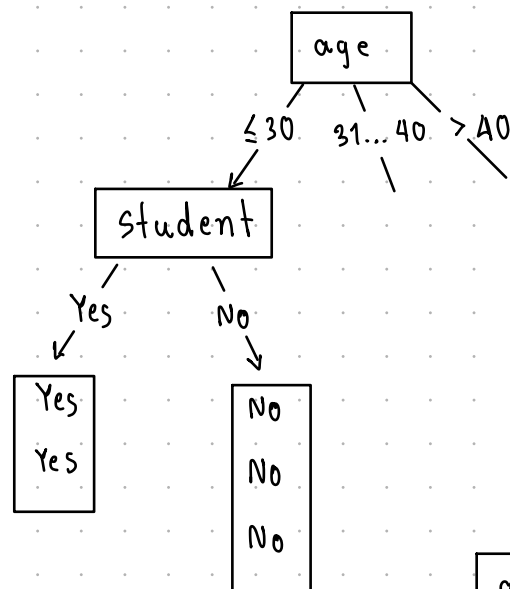
$$\text{Info}_{\text{credit\_r}}(D) = 0.9509$$

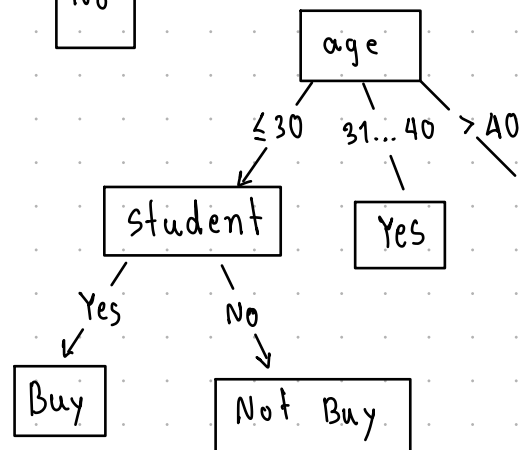$$\text{Gain (income)} = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.9710 - 0.4 = 0.5710$$

$$\text{Gain (student)} = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.9710 - \text{Gain มากสุด}$$

$$\text{Gain (credit\_rating)} = 0.0201$$

∴ เลือก เป็น node ของ age ≤ 30

② age 31...40

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
31...40	low	yes	excellent	yes
<=30	medium	no	fair	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

เนื่องจาก yes หมด ∴ จิ้วไม่ได้ คำนวณ หา node ของ 31...40

③ age > 40

$$\text{Info}(D) = I(\overset{Y\ N}{3,2})$$
$$= 0.971$$

$$\text{Info}_{\text{income}}(D) = \frac{3}{5}I(2,1) + \frac{2}{5}I(1,1)$$
$$= \frac{3}{5}\left[-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$$

$$\text{Info}_{\text{income}}(D) = 0.9509$$

$$\text{Info}_{\text{student}}(D) = \frac{3}{5}I(2,1) + \frac{2}{5}I(1,1)$$
$$= \frac{3}{5}\left[-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$$

$$\therefore \text{Info}_{\text{student}}(D) = 0.9509$$

$$\text{Info}_{\text{credit\_r}}(D) = \frac{2}{5}I(0,2) + \frac{3}{5}I(3,0)$$
$$= \frac{2}{5}\left[-\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right)\right] + \frac{3}{5}\left[-\frac{3}{3}\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right)\right]$$

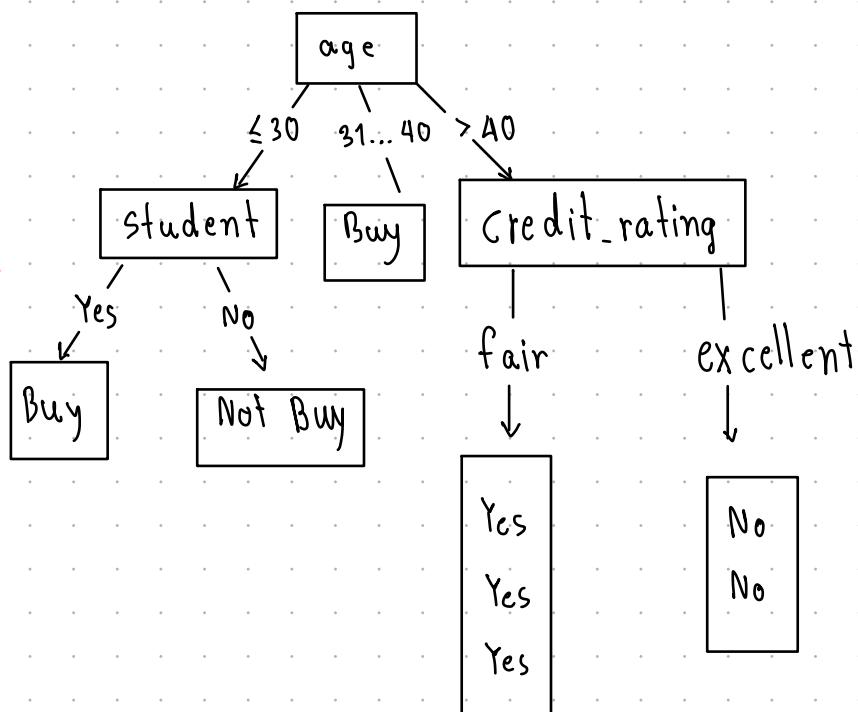$$\therefore \text{Info}_{\text{credit\_r}}(D) = 0$$

Gain (income) = 0.0201

Gain (student) = 0.0201

Gain (credit_rating) = 0.9710

มากสุด ∴ เลือก เป็น node

ของ age > 40

∴ จะได้ decision tree ดังนี้

```
                          ┌─────────┐
                          │   age   │
                          └─────────┘
             ┌─────────────────┼─────────────────┐
        ┌─────────┐        ┌─────────┐        ┌─────────┐
        │  ≤30    │        │ 31...40 │        │  > 40   │
        └─────────┘        └─────────┘        └─────────┘
             │                  │                  │
        ┌─────────┐             │          ┌──────────────┐
        │ Student │        ┌─────────┐     │ credit_rating │
        └─────────┘        │   Buy   │     └──────────────┘
          │      │         └─────────┘       │         │
        Yes      No                        fair    excellent
          │      │                           │         │
     ┌───────┐ ┌──────────┐            ┌─────────┐ ┌──────────┐
     │  Buy  │ │ Not Buy  │            │   Buy   │ │ Not Buy  │
     └───────┘ └──────────┘            └─────────┘ └──────────┘
```