653020212-4 พรชนิตว์ เหล่าโยธี

Example: Attribute Selection with Information Gain

- ❑ Class P: buys_computer = “yes”
- ❑ Class N: buys_computer = “no”

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

age	p_i	n_i	$I(p_i, n_i)$
<=30	2	3	0.971
31…40	4	0	0
>40	3	2	0.971

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

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

$\frac{5}{14}I(2,3)$ means “age <=30” has 5 out of 14 samples, with 2 yes’es and 3 no’s.

Hence 4/14

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

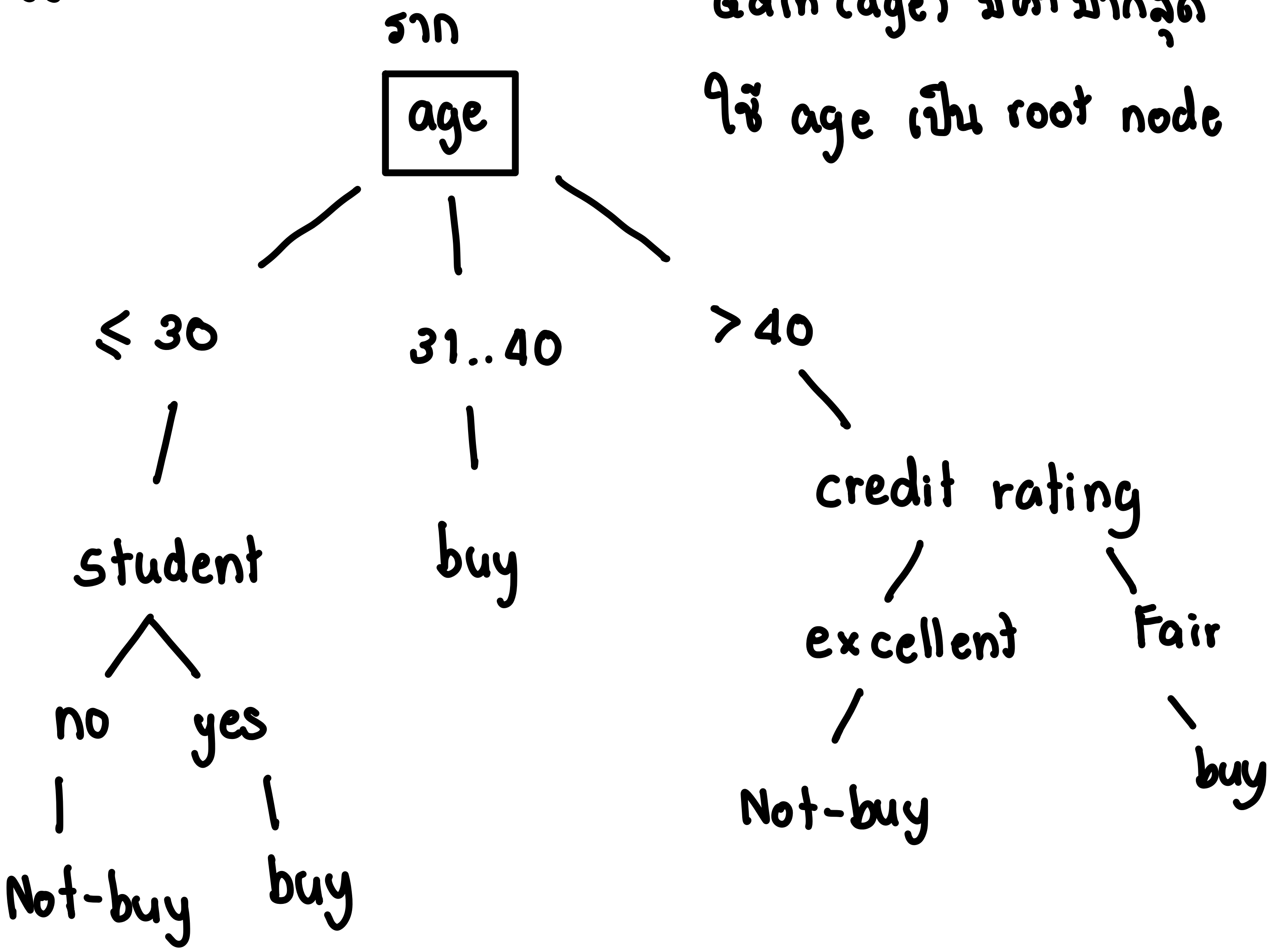
Similarly, we can get

$$Gain(income) = 0.029$$

$$Gain(student) = 0.151$$

$$Gain(credit\_rating) = 0.048$$

Resulting tree :



Gain (age) มีค่ามากสุด

ใช้ age เป็น root node

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

$= 0.9710$

653020212-4 พรชนิตว่ เหล่าโยธี

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

$$\text{Info}(D) = 0.9183$$

$$\text{Info}_{age}(D) = \frac{4}{12} I(2,2) + \frac{3}{12} I(3,0) + \frac{5}{12} I(3,2) = 0.738$$

$$\text{Gain}(age) = \text{Info}(D) - \text{Info}_{age}(D) = 0.9183 - 0.738 = 0.1803$$

$$\text{Info}_{income}(D) = \frac{4}{12} I(2,2) + \frac{3}{12} I(2,1) + \frac{5}{12} (4,1) = 0.863$$

$$\text{Gain}(income) = 0.0553$$

$$\text{Info}_{student}(D) = \frac{6}{12} I(3,3) + \frac{6}{12} I(5,1) = 0.825$$

$$\text{Gain}(student) = 0.0933$$

$$\text{Info}_{credit\_rating}(D) = \frac{6}{12} I(3,3) + \frac{6}{12} I(5,1) = 0.825$$

$$\text{Gain}(credit\_rating) = 0.1703$$

เลือก Gain(age) เพราะมีค่ามากที่สุด คือ 0.1803 #

653020212-4 พรชนิตว์ เหล่าโยธี