พัณณิตา ทองบ่อ
64302050S-2

feature

class (Y = 9 , N = 5)

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. Age

age	yes	No
≤ 30	2	3
31-40	4	0
> 40	3	2

2. Income

Income	yes	No
low	3	1
medium	4	2
high	2	2

3. student

student	yes	No
yes	6	1
No	3	4

4. credit_rating

credit_rating	yes	No
fair	6	2
excellent	3	3

1. หา class

(9,5)

สูตร $\quad Info(D) = -\sum_{i=1}^{m} p_i \log_2(P_i)$

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

$$= 0.41 + 0.53$$

$$= 0.94$$

∴ $Info(D) = 0.94$

2. หา feature

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

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

$$= \frac{5}{14}\left[0.5287 + 0.44217\right] + \frac{5}{14}\left[0.44217 + 0.52877\right]$$

∴ $Info_{age}(D) = 0.894$

พัณณิตา ทองบ่อ
6430205082

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

$$= \frac{4}{14}\left[\frac{-2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right] + \frac{6}{14}\left[\frac{-4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right] + \frac{4}{14}\left[\frac{-3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right]$$

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

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

$$= \frac{7}{14}\left[\frac{-6}{7}\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right)\right] + \frac{7}{14}\left[\frac{-3}{7}\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\log_2\left(\frac{4}{7}\right)\right]$$

$\therefore$ $\text{Info}_{\text{student}}(D) = 0.788$

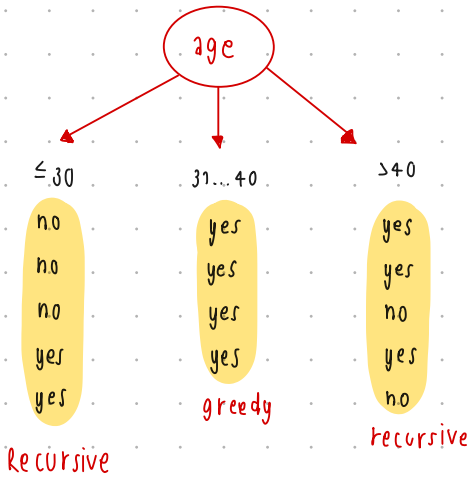
$$\text{Info}_{CR}(D) = \frac{6}{14} I(6,2) + \frac{6}{14} (3,3)$$

$$= \frac{6}{14}\left[\frac{-3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log\left(\frac{3}{6}\right)\right] + \frac{8}{14}\left[\frac{-6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right)\right]$$

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

หา $\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$ ก่อน

Gain (age)	=	0.94 - 0.694	=	0.246 (Root node)
Gain (Income)	=	0.94 - 0.911	=	0.029
Gain (student)	=	0.94 - 0.788	=	0.152
Gain (Credit_rating)	=	0.94 - 0.892	=	0.048



พัณณิตา ทองบ่อ
6430205082

age ≤ 30

คำนวณ class (I = 2,3)

$$Info(D) = \sum_{i=1}^{m} p_i \log_2(p_i)$$

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

$$= 0.5288 + 0.4422$$

$$= 0.971$$

∴ Info (D) = 0.971

คำนวณ feature

$$Info_{income}(D) = \sum_{j=1}^{v} \left|\frac{D_j}{D}\right| \times Info(D_j)$$

$$= \frac{3}{5} I(2,1) + \frac{2}{5} I(1,1)$$

$$= \frac{3}{5}\left[-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) + -\frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$$

$$= 0.551 + 0.4$$

$$= 0.951$$

∴ $Info_{income}(D)$ = 0.951

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

$$= \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] + \frac{3}{5}\left[-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right]$$

$$= 0.4 + 0.551$$

$$= 0.951$$

∴ $Info_{student}(D)$ = 0.951

$$Info_{CR}(D) = \frac{3}{5} I(3,0) + \frac{2}{5} I(0,2)$$

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

$$= 0$$

∴ $Info_{CR}(D)$ = 0

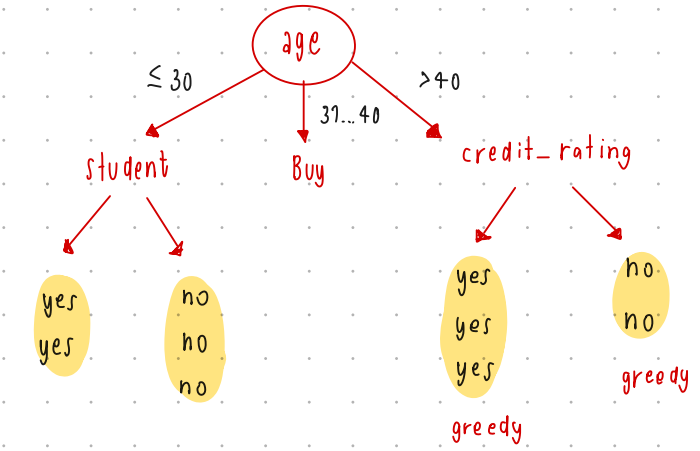
$$\text{จาก } Gain(A) = Info(D) - Info_A(D)$$

Gain (Income) = 0.971 − 0.951 = 0.2

Gain (student) = 0.971 − 0.951 = 0.2

Gain (Credit_rating) = 0.971 − 0 = 0.9710

พัณณิตา ทองบ่อ
6430205508-2



Dicision Tree