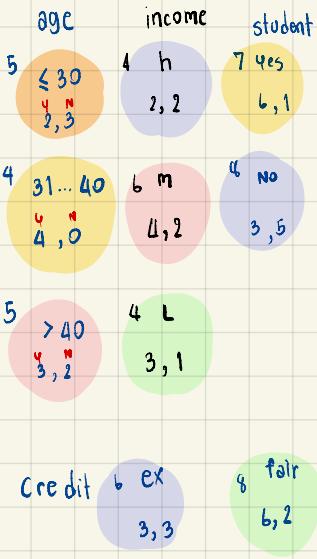


HW 5 : Decision Tree



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

\times : features | : class

1 คำนวณ class

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

$$= I(9,5)$$

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

$$= 0.41 + 0.53$$

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

2 คำนวณ Feature

$$\text{Info}_A(D)$$

$$\text{สูตร} \quad \text{Info}_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times \text{Info}(D_j)$$

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

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

$$= \frac{5}{14} [0.5287 + 0.44217] + \left[\frac{4}{14} \times \text{มากที่สุด} \right] + \frac{5}{14} [0.44217 + 0.52877]$$

$$= \left[\frac{5}{14} \times 0.9709 \right] + \left[\frac{5}{14} \times 0.9709 \right]$$

$$= 0.34676 + 0.34676 = 0.69352$$

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

$$\begin{aligned}
 \text{Info}_{\text{income}}(D) &= \frac{4}{14} I(0,2) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1) \\
 &= \frac{4}{14} \left[-\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right] + \frac{6}{14} \left[-\frac{4}{6} \log_2 \left(\frac{4}{6} \right) - \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right] + \frac{4}{14} \left[-\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right] \\
 &\approx \frac{4}{14} [0.5 + 0.5] + \frac{6}{14} [0.3899 + 0.5283] + \frac{4}{14} [0.3112 + 0.5] \\
 &\approx \left[\frac{4}{14} \times 1 \right] + \left[\frac{6}{14} \times 0.9192 \right] + \left[\frac{4}{14} \times 0.8112 \right] \\
 &\approx 0.2857 + 0.3935 + 0.2317 = 0.9109
 \end{aligned}$$

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

$$\begin{aligned}
 \text{Info}_{\text{student}}(D) &= \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4) \\
 &= \frac{7}{14} \left[-\frac{6}{7} \log_2 \left(\frac{6}{7} \right) - \frac{1}{7} \log_2 \left(\frac{1}{7} \right) \right] + \frac{7}{14} \left[-\frac{3}{7} \log_2 \left(\frac{3}{7} \right) - \frac{4}{7} \log_2 \left(\frac{4}{7} \right) \right] \\
 &\approx \frac{7}{14} [0.1906 + 0.4010] + \frac{7}{14} [0.5238 + 0.4613] \\
 &\approx \left[\frac{7}{14} \times 0.5916 \right] + \left[\frac{7}{14} \times 0.9851 \right] \\
 &\approx 0.2958 + 0.4925 = 0.7883
 \end{aligned}$$

$$\text{Info}_{\text{student}}(D) \approx 0.7883$$

$$\begin{aligned}
 \text{Info}_{\text{credit}}(D) &= \frac{6}{14} I(3,3) + \frac{8}{14} I(6,2) \\
 &= \frac{6}{14} \left[-\frac{3}{6} \log_2 \left(\frac{3}{6} \right) - \frac{3}{6} \log_2 \left(\frac{3}{6} \right) \right] + \frac{8}{14} \left[-\frac{6}{8} \log_2 \left(\frac{6}{8} \right) - \frac{2}{8} \log_2 \left(\frac{2}{8} \right) \right] \\
 &\approx \frac{6}{14} [0.5 + 0.5] + \frac{8}{14} [0.3113 + 0.5] \\
 &\approx \left[\frac{6}{14} \times 1 \right] + \left[\frac{8}{14} \times 0.9113 \right] \\
 &\approx 0.4285 + 0.4636 = 0.8921
 \end{aligned}$$

$$\text{Info}_{\text{credit}}(D) \approx 0.892$$

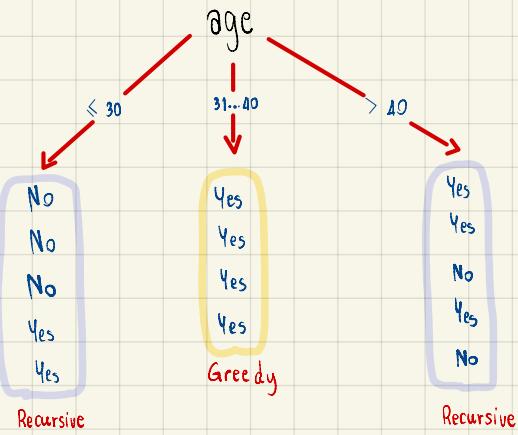
$$\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$$

$$\text{Gain}(\text{age}) = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.940 - 0.694 = 0.246 \rightarrow \text{Gain} \text{ หมายความว่า} \text{ เส้นทาง root node}$$

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

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

$$\text{Gain}(\text{credit_rating}) = \text{Info}(D) - \text{Info}_{\text{credit_rating}}(D) = 0.940 - 0.892 = 0.048$$



Recursive age < 30

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 คำนวณ class

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

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

$$= 0.5288 + 0.4422$$

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

2 คำนวณ Feature

$$\begin{aligned}
 \text{Info}_{\text{income}}(D) &= \frac{2}{5} I(0,2) + \frac{2}{5} I(1,1) + \frac{1}{5} I(1,0) \\
 &= \frac{2}{5} \left[-\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{2} \log_2 \left(\frac{1}{2}\right) \right] + \frac{2}{5} \left[-\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{2} \log_2 \left(\frac{1}{2}\right) \right] + \frac{1}{5} \left[-\frac{1}{1} \log_2 \left(\frac{1}{1}\right) - \frac{0}{1} \log_2 \left(\frac{0}{1}\right) \right] \\
 &= \left[\frac{2}{5} \times \text{min} \right] + \frac{2}{5} [0.5 + 0.5] + \left[\frac{1}{5} \times \text{min} \right] \\
 &= \frac{2}{5} \times 1
 \end{aligned}$$

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

$$\begin{aligned}
 \text{Info}_{\text{student}}(D) &= \frac{2}{5} I(1,0) + \frac{3}{5} I(0,3) \\
 &= \frac{2}{5} \left[-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{0}{2} \right) \right] + \frac{3}{5} \left[-\frac{0}{3} \log_2 \left(\frac{0}{3} \right) - \frac{3}{3} \log_2 \left(\frac{3}{3} \right) \right] \\
 &= \left[\frac{2}{5} \times \text{information} \right] + \left[\frac{3}{5} \times \text{information} \right]
 \end{aligned}$$

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

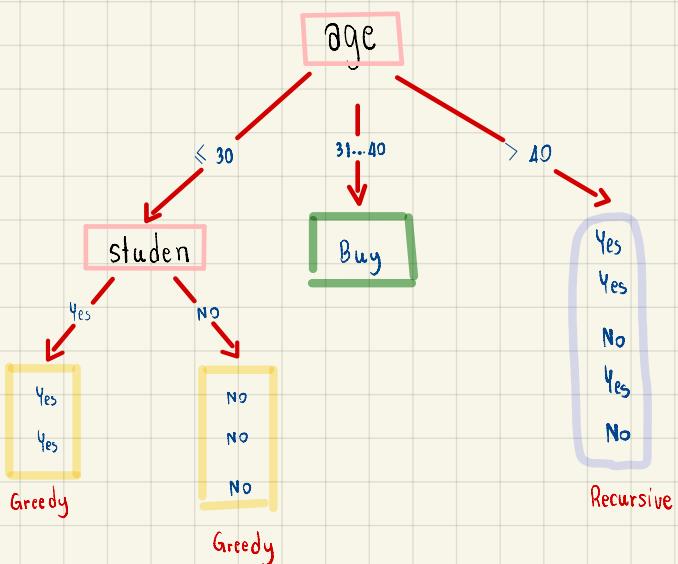
$$\begin{aligned}
 \text{Info}_{\text{credit}}(D) &= \frac{2}{5} I(1,1) + \frac{3}{5} I(1,2) \\
 &= \frac{2}{5} \left[-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right] + \frac{3}{5} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right] \\
 &= \frac{2}{5} [0.5 + 0.5] + \frac{3}{5} [0.5283 + 0.3899] \\
 &= \left[\frac{2}{5} \times 1 \right] + \left[\frac{3}{5} \times 0.9182 \right] \\
 &= 0.4 + 0.5509
 \end{aligned}$$

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

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

$$\text{Gain}(\text{student}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.9710 - 0 = 0.9710 \rightarrow \text{Gain} \text{ မရရ } \text{ ငြေ။}$$

$$\text{Gain}(\text{credit_rating}) = \text{Info}(D) - \text{Info}_{\text{credit_rating}}(D) = 0.9710 - 0.9509 = 0.0201$$



Recursive age > 40

X

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 คำนวณ class

$$\text{Info}(D) = I^{4, N}$$

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

$$= 0.4872 + 0.5288$$

$$\text{Info}(D) = 0.971$$

2 คำนวณ Feature

$$\text{Info}_A(D)$$

$$\text{Info}_{\text{income}}(D) = \frac{3}{5} I_{(2,1)} + \frac{2}{5} I_{(1,1)}$$

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

$$= \frac{3}{5} [0.3899 + 0.5283] + \frac{2}{5} [0.5 + 0.5]$$

$$= \left[\frac{3}{5} \times 0.9182 \right] + \left[\frac{2}{5} \times 1 \right]$$

$$= 0.5509 + 0.4$$

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

4 N

$$\text{Info}_{\text{student}}(D) = \frac{3}{5} I_{(2,1)} + \frac{2}{5} I_{(1,1)}$$

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

$$= \frac{3}{5} [0.3899 + 0.5283] + \frac{2}{5} [0.5 + 0.5]$$

$$= \left[\frac{3}{5} \times 0.9182 \right] + \left[\frac{2}{5} \times 1 \right]$$

$$= 0.5509 + 0.4$$

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

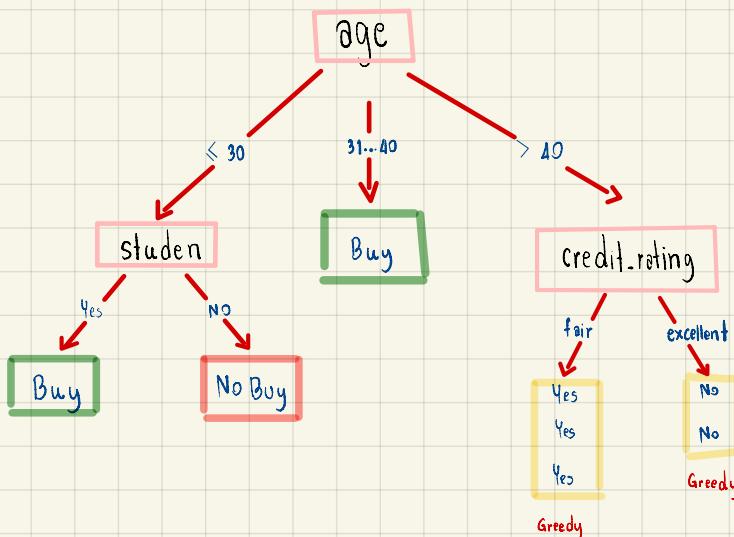
$$\begin{aligned}
 \text{Info}_{\text{credit}}(D) &= \frac{2}{5} I(0,2) + \frac{3}{5} I(3,0) \\
 &= \frac{2}{5} \left[-\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{2} \log_2 \left(\frac{1}{2}\right) \right] + \frac{3}{5} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right) \right] \\
 &= \left[\frac{2}{5} \times \text{minila} \right] + \left[\frac{3}{5} \times \text{minila} \right]
 \end{aligned}$$

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

$$\text{Gain}(\text{income}) = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.9710 - 0.9509 = 0.0201$$

$$\text{Gain}(\text{student}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.9710 - 0.9509 = 0.0201$$

$$\text{Gain}(\text{credit_rating}) = \text{Info}(D) - \text{Info}_{\text{credit_rating}}(D) = 0.9710 - 0 = 0.9710 \rightarrow \text{Gain} \text{ มาก} \text{ เลย}$$



FINAL ++

