

## Hw.2 Decision tree

morning group 11/30/2023-4

Training data set: Who buys computer?

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

$\{x \in \{\text{age, income, student, credit\_rating}\} \cup \{\text{buys\_computer}\}$

1) Information class in  $\text{Info}(D)$ :  $(9, 5)$

$$\begin{aligned}\text{Info}(D) &= -\sum_{i=1}^2 p_i \log_2(p_i) \\ &= I(9, 5) \\ &= -\frac{9}{14} \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \log_2\left(\frac{5}{14}\right) \\ &= 0.41 + 0.53 \\ &= 0.944 \quad \times\end{aligned}$$

	Y	N
age:	5/6	3/6
31..40	4/6	0/6
>40	3/6	2/6
income:	2/6	2/6
med	1/6	2/6
low	3/6	1/6
student:	6/10	1/10
no	3/10	5/10
credit:	3/10	3/10
fair	6/10	2/10

$$\begin{aligned}2.) \text{ feature in } \text{Info}_f(D) &= \sum_{j=1}^5 p_j \text{Info}(D_j) \\ 2.1 > \text{Info}_{\text{age}}(D) &= \frac{5}{14} I(2, 3) + \frac{3}{14} I(4, 0) \\ &\quad + \frac{2}{14} I(3, 2) \\ &= \frac{5}{14} \left[ -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) \right] \\ &\quad + \frac{3}{14} \left[ -\frac{4}{5} \log_2\left(\frac{4}{5}\right) - \frac{1}{5} \log_2\left(\frac{1}{5}\right) \right] \\ &\quad + \frac{2}{14} \left[ -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) \right] \\ &= \frac{5}{14} [0.6267 + 0.11217] \\ &\quad + \frac{3}{14} [0.11217 + 0.04877] \\ \text{Info}_{\text{age}}(D) &= 0.694 \quad \times\end{aligned}$$

$$2.2) \text{ Info}_{\text{income}}(D) = \frac{3}{10} I(1,2) + \frac{6}{10} I(3,1)$$

$$= \frac{3}{10} \left[ -\frac{1}{2} \log_2 \left( \frac{2}{3} \right) - \frac{1}{2} \log_2 \left( \frac{1}{3} \right) \right] + \frac{6}{10} \left[ -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \left( \frac{1}{2} \right) \right] + \frac{1}{10} \left[ \frac{2}{3} \log_2 \left( \frac{3}{2} \right) - \frac{1}{3} \log_2 \left( \frac{1}{2} \right) \right]$$

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

$$2.3) \text{ Info}_{\text{student}}(D) = \frac{3}{10} I(6,1) + \frac{7}{10} I(3,2)$$

$$= \frac{3}{10} \left[ -\frac{1}{3} \log_2 \left( \frac{1}{3} \right) - \frac{2}{3} \log_2 \left( \frac{1}{2} \right) \right] + \frac{7}{10} \left[ -\frac{3}{7} \log_2 \left( \frac{2}{7} \right) - \frac{4}{7} \log_2 \left( \frac{1}{3} \right) \right]$$

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

$$2.4) \text{ Info}_{\text{credit}}(D) = \frac{3}{10} I(3,3) + \frac{6}{10} I(6,2)$$

$$= \frac{3}{10} \left[ -\frac{3}{5} \log_2 \left( \frac{2}{5} \right) - \frac{2}{5} \log_2 \left( \frac{3}{5} \right) \right] + \frac{6}{10} \left[ -\frac{1}{3} \log_2 \left( \frac{1}{3} \right) - \frac{2}{3} \log_2 \left( \frac{2}{3} \right) \right]$$

$$\text{Info}_{\text{credit}}(D) = 0.812 \quad \times$$

3.) Gain w/ gain Root node

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

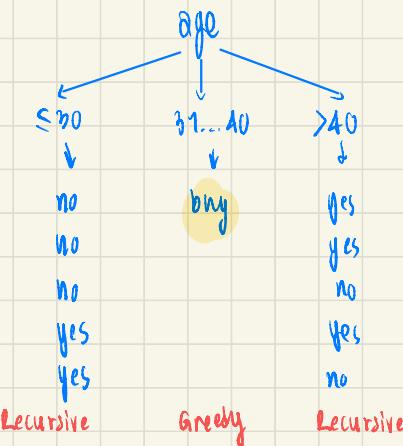
$$= 0.910 - 0.909$$

$$\text{Gain}(\text{age}) = 0.910 - 0.9035 = 0.0245 \quad \times \longrightarrow \text{Gain age als root node}$$

$$\text{Gain}(\text{income}) = 0.910 - 0.909 = 0.001$$

$$\text{Gain}(\text{student}) = 0.910 - 0.788 = 0.1516$$

$$\text{Gain}(\text{credit\_rating}) = 0.910 - 0.812 = 0.098$$



3.1.1 > Recursive age:  $\leq 30$  ( $\frac{4}{5}, \frac{1}{5}$ )

$$-\text{Info}(D) = I(\text{high}) = -\frac{2}{5} \log_2\left(\frac{1}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) = 0.971$$

$$-\text{Info}_{\text{income}}(D) = \frac{2}{5} I(\text{low}, 2) + \frac{2}{5} I(\text{med}, 1) + \frac{1}{5} I(\text{high}, 0) = 0.4$$

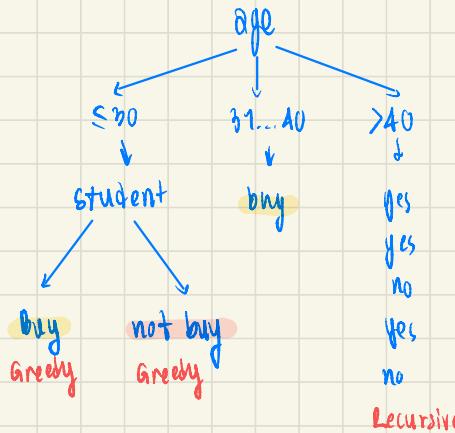
$$-\text{Info}_{\text{student}}(D) = \frac{2}{5} I(\text{yes}, 0) + \frac{3}{5} I(\text{no}, 1) = 0$$

$$-\text{Info}_{\text{credit-rating}}(D) = \frac{2}{5} I(\text{excellent}, 1) + \frac{3}{5} I(\text{fair}, 2) = 0.9509$$

$$-\text{Gain}(\text{income}) = 0.9710 - 0.4 = 0.5910$$

$$\text{Gain}(\text{student}) = 0.9710 - 0 = 0.9710 \text{ KK}$$

$$\text{Gain}(\text{credit-rating}) = 0.9710 - 0.9509 = 0.0201$$



3.1.2 > Recursive age:  $> 40$  ( $\frac{4}{5}, \frac{1}{5}$ )

$$-\text{Info}(D) = -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) = 0.9710$$

$$-\text{Info}_{\text{income}}(D) = \frac{2}{5} I(\text{med}, 1) + \frac{3}{5} I(\text{low}, 0) = 0.9509$$

$$-\text{Info}_{\text{student}}(D) = \frac{2}{5} I(\text{yes}, 1) + \frac{3}{5} I(\text{no}, 0) = 0.4$$

$$-\text{Info}_{\text{credit-rating}}(D) = \frac{2}{5} I(\text{high}, 2) + \frac{3}{5} I(\text{med}, 1) = 0$$

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

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

$$\text{Gain}(\text{credit-rating}) = 0.9710 - 0 = 0.9710 \text{ KK}$$

## Figure 2-10 Decision tree

