

# Example: Attribute Selection with Information Gain

- Class P: buys\_computer = “yes”
- Class N: buys\_computer = “no”

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

$I(p_i, n_i)$

age	$p_i$	$n_i$	$I(p_i, n_i)$
$\leq 30$	2	3	<del>0.971</del> 0.8
31...40	3	0	0
$>40$	3	2	0.971

age	income	student	credit_rating	buys_computer
$\leq 30$	high	no	fair	no
$\leq 30$	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
$\leq 30$	medium	no	fair	no
$\leq 30$	low	yes	fair	yes
$>40$	medium	yes	fair	yes
$\leq 30$	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
$>40$	medium	no	excellent	no

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

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

Hence

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

Similarly, we can get

$$Gain(income) = 0.029$$

$$Gain(student) = 0.151$$

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

ព័ត៌មានការណែនាំ

HW4

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

13/35

## (I) គំនាល់នោង Info(D)

$$\begin{aligned}
 \text{Info}(D) &= I(8, 4) = -\frac{8}{12} \log_2 \frac{8}{12} - \frac{4}{12} \log_2 \frac{4}{12} \\
 &= \left(-\frac{8}{12}\right)(-0.5850) - \left(\frac{4}{12}\right)(-1.5850) \\
 &\approx 0.39 + 0.5283
 \end{aligned}$$

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

II තැනුමෙනු නිශ්චාලීත නිවේදන නිර්මාණ මාරුග සහ නිශ්චාලීත නිවේදන

$$\begin{aligned}
 * \text{Info}_{(\text{age})}(D) &= \frac{4}{12} I(2,2) + \frac{3}{12} I(3,0) + \frac{5}{12} I(3,2) \\
 &= \frac{4}{12} \left[ -\frac{2}{4} \log_2 \left( \frac{2}{4} \right) - \frac{2}{4} \log_2 \left( \frac{2}{4} \right) \right] + \frac{3}{12} \left[ -\frac{3}{3} \log_2 \left( \frac{3}{3} \right) - \frac{0}{3} \log_2 \left( \frac{0}{3} \right) \right] \\
 &\quad + \frac{5}{12} \left[ -\frac{3}{5} \log_2 \left( \frac{3}{5} \right) - \frac{2}{5} \log_2 \left( \frac{2}{5} \right) \right] \\
 &= \frac{4}{12} (1) + \frac{3}{12} (0) + \frac{5}{12} (0.7370)
 \end{aligned}$$

$$\text{Info}_{(\text{age})}(D) = 0.6404$$

$$\begin{aligned}
 * \text{Info}_{(\text{income})}(D) &= \frac{4}{12} I(2,2) + \frac{5}{12} I(4,1) + \frac{3}{12} I(2,1) \\
 &= \frac{4}{12} \left[ -\frac{2}{4} \log_2 \left( \frac{2}{4} \right) - \frac{2}{4} \log_2 \left( \frac{2}{4} \right) \right] + \frac{5}{12} \left[ -\frac{4}{5} \log_2 \left( \frac{4}{5} \right) - \frac{1}{5} \log_2 \left( \frac{1}{5} \right) \right] \\
 &\quad + \frac{3}{12} \left[ -\frac{2}{3} \log_2 \left( \frac{2}{3} \right) - \frac{1}{3} \log_2 \left( \frac{1}{3} \right) \right] \\
 &= \frac{4}{12} (1) + \frac{5}{12} (0.7219) + \frac{3}{12} (0.9183)
 \end{aligned}$$

$$\therefore \text{Info}_{(\text{income})}(D) = 0.8637$$

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

$$= \frac{6}{12} \left[ -\frac{5}{6} \log_2 \left( \frac{5}{6} \right) - \frac{1}{6} \log_2 \left( \frac{1}{6} \right) \right] + \frac{6}{12} \left[ -\frac{3}{6} \log_2 \left( \frac{3}{6} \right) - \frac{3}{6} \log_2 \left( \frac{3}{6} \right) \right]$$

$$= \frac{6}{12} (0.6500) + \frac{6}{12} (1)$$

$$\therefore \text{Info}_{(\text{student})}(D) = 0.825$$

$$\star \text{Info}_{(\text{credit})}(D) = \frac{7}{12} I(6,1) + \frac{5}{12} I(2,3)$$

$$= \frac{7}{12} \left[ -\frac{6}{7} \log_2 \left( \frac{6}{7} \right) - \frac{1}{7} \log_2 \left( \frac{1}{7} \right) \right] + \frac{5}{12} \left[ -\frac{2}{5} \log_2 \left( \frac{2}{5} \right) - \frac{3}{5} \log_2 \left( \frac{3}{5} \right) \right]$$

$$= \frac{7}{12} (0.5917) + \frac{5}{12} (0.9710)$$

$$\therefore \text{Info}_{(\text{credit})}(D) = 0.7497$$

## (III) ດຳເນັດໃຫຍ່ Gain

$$\ast \text{Gain (age)} = 0.9183 - 0.6404 = 0.2779$$

$$\ast \text{Gain (Income)} = 0.9183 - 0.8637 = 0.0546$$

$$\ast \text{Gain (Student)} = 0.9183 - 0.8250 = 0.0933$$

$$\ast \text{Gain (Credit rating)} = 0.9183 - 0.7497 = 0.1686$$

ຢ່າງ Gain (Age) ສຳເນົາຫຼຸດ ຖືໍ່ເກີດ Age ປິເຕ ລົດ

