

# Example: Attribute Selection with Information Gain

□ Class P: buys\_computer = "yes"

□ Class N: buys\_computer = "no"

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

$I(p_i, n_i)$

age	$p_i$	$n_i$	$I(p_i, n_i)$
$\leq 30$	2	3	0.971
31...40	4	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

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

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

Hence

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

Similarly, we can get

$$Gain(income) = 0.029$$

$$Gain(student) = 0.151$$

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

การวัดค่าความ

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

Si

①  $Gain(Age) = 0.940 - 0.7283 = 0.2117$

ค่าของ Age เป็น Type ความน่าจะเป็นของค่า Gain สูง

↑ ตัวอย่างทำแบบ

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

Si

13/85

$$\begin{aligned}
 \textcircled{1} \quad u_1 \quad \text{Info}(D) &= I(8, 4) \\
 &= -\frac{8}{12} \log_2 \frac{8}{12} - \frac{4}{12} \log_2 \frac{4}{12} \\
 &= \left(-\frac{8}{12}\right)(-0.5850) - \left(\frac{4}{12}\right)(-1.5850) \\
 &= 0.39 + 0.5283 \\
 \therefore \text{Info}(D) &= 0.9183
 \end{aligned}$$

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

$$\text{Info}_{\text{age}}(D) = 0.3237 + 0 + 0.4046$$

$$= 0.7283$$