

5

## Decision Tree Induction

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

Class P : buys\_Computer = "yes"

Class N : buys\_Computer = "no"

age	$p_i$	$n_i$	$I(p_i, n_i)$
<=30	2	3	0.971
31...40	4	0	0
>40	3	2	0.971

$$\text{Info income}(D) = \frac{4}{14} I(2,2) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1) = 0.911 \text{ bits}$$

Info income high : yes 2 / no 2  $\Rightarrow$  1  
 medium : yes 4 / no 2  $\Rightarrow$  6  
 low : yes 3 / no 1  $\Rightarrow$  4

□ Expected information (entropy) needed to classify a tuple in D:

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

□ Information needed (after using A to split D into v partitions) to classify D:

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j)$$

□ Information gained by branching on attribute A

$$Gain(A) = Info(D) - Info_A(D)$$

Attribute និមួយនាពេលវេលាដីជាអត្ថបទ នៅរបស់ទាំងអស់ខ្លួន  
 no age , income , Student និង Credit\_rating

តូចរាប់មិនមែនជាប្រភេទ Class yes ឬ no ដែល

$$Info(D) = I(9,5) = -\frac{9}{14} \log_2(\frac{9}{14}) - \frac{5}{14} \log_2(\frac{5}{14}) = 0.910$$

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

$$+ \frac{5}{14} I(3,2)$$

$$= 0.694 \text{ bits}$$

Info age <= 30 : yes 2 / no 3  $\Rightarrow$  5

31...40 : yes 4 / no 0  $\Rightarrow$  1

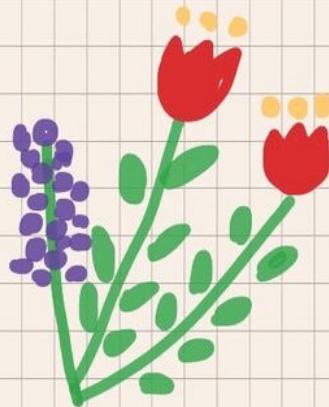
>40 : yes 3 / no 2  $\Rightarrow$  5

$$Info_{student}(D) = \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4) = 0.989 \text{ bits}$$

Info student no : yes 6 / no 1  $\Rightarrow$  7  
 yes : yes 3 / no 4  $\Rightarrow$  7

HAPPY DAY

HAPPY LIFE



# HOMEWORK



$$\text{Info credit\_rating}(D) = \frac{8}{14} I(6,2) + \frac{6}{14} I(3,3)$$

$$= 0.892 \text{ bits}$$

$\text{Info credit\_rating}$  fair : yes 6 / no 2  $\Rightarrow 8$   
excellent: yes 3 / no 3  $\Rightarrow 6$

$$\text{Gain}(\text{age}) = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.246$$

Similarly, we can get

$$\text{Gain}(\text{income}) = 0.029$$

$$\text{Gain}(\text{student}) = 0.151$$

$$\text{Gain}(\text{credit\_rating}) = 0.048$$

Find gain nomination using attribute  $\text{age}$

$$\text{gain}(\text{age}) = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.940 - 0.694 = 0.246$$

from age nomination

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

$$\text{gain}(\text{student}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.940 - 0.789 = 0.151$$

$$\text{gain}(\text{credit\_rating}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.940 - 0.892 = 0.048$$

① If  $\text{age} \leq 30$

age	income	student	credit rating	buys computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
<=30	medium	yes	excellent	yes

high: 2  
medium: 2  
low: 1  
yes: 2  
no: 3  
fair: 3  
excellent: 2

$$\text{Info}(D) = J(2,3) = -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) = 0.971$$

$$\text{Info}_{\text{income}}(D) = \frac{2}{5} J(0,2) + \frac{3}{5} J(1,1) + \frac{1}{5} J(1,0) = 0.4$$

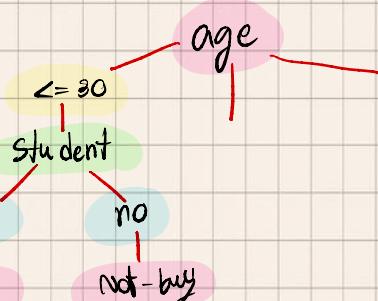
$$\text{Info}_{\text{student}}(D) = \frac{2}{5} J(2,0) + \frac{3}{5} J(0,3) = 0$$

$$\text{Info}_{\text{credit\_rating}}(D) = \frac{2}{5} J(1,2) + \frac{3}{5} J(1,1) = 0.951$$

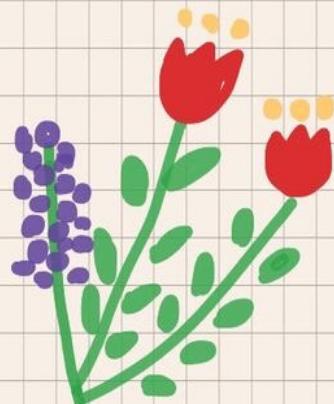
$$\text{gain}(\text{income}) = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.971 - 0.4 = 0.571$$

$$\text{gain}(\text{student}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.971 - 0 = 0.971 \text{ max}$$

$$\text{gain}(\text{credit\_rating}) = \text{Info}(D) - \text{Info}_{\text{credit\_rating}}(D) = 0.971 - 0.951 = 0.02$$



HAPPY DAY HAPPY LIFE



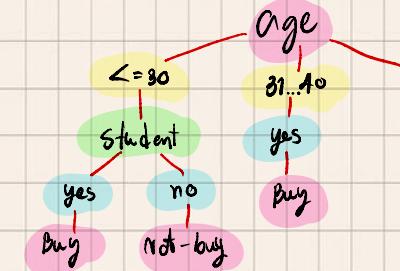
# HOMEWORK

(2) តើវិវាទ age 31...40

age	income	student	credit_rating	buys_computer
31...40	high	no	fair	yes
31...40	low	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes

ព័ត៌មានទាំងអស់នេះ 31-40 មាន ៤ yes នៃ ៦ នៅលើ ៩

នៅលើ ៣ នេះ yes នៅរបស់ខ្លួន



(3) តើវិវាទ age > 40

age	income	student	credit rating	buys computer
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
>40	medium	yes	fair	yes
>40	medium	no	excellent	no

medium 3  
low 2  
no 2  
yes 3  
fair 3  
excellent 2  
Info 3  
Info 2  
Info 3  
Info 2

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

$$\text{Info}_{\text{income}}(D) = \frac{3}{5} I(2,1) + \frac{2}{5} I(1,1) = 0.951$$

$$\text{Info}_{\text{student}}(D) = \frac{3}{5} I(2,1) + \frac{2}{5} I(1,1) = 0.951$$

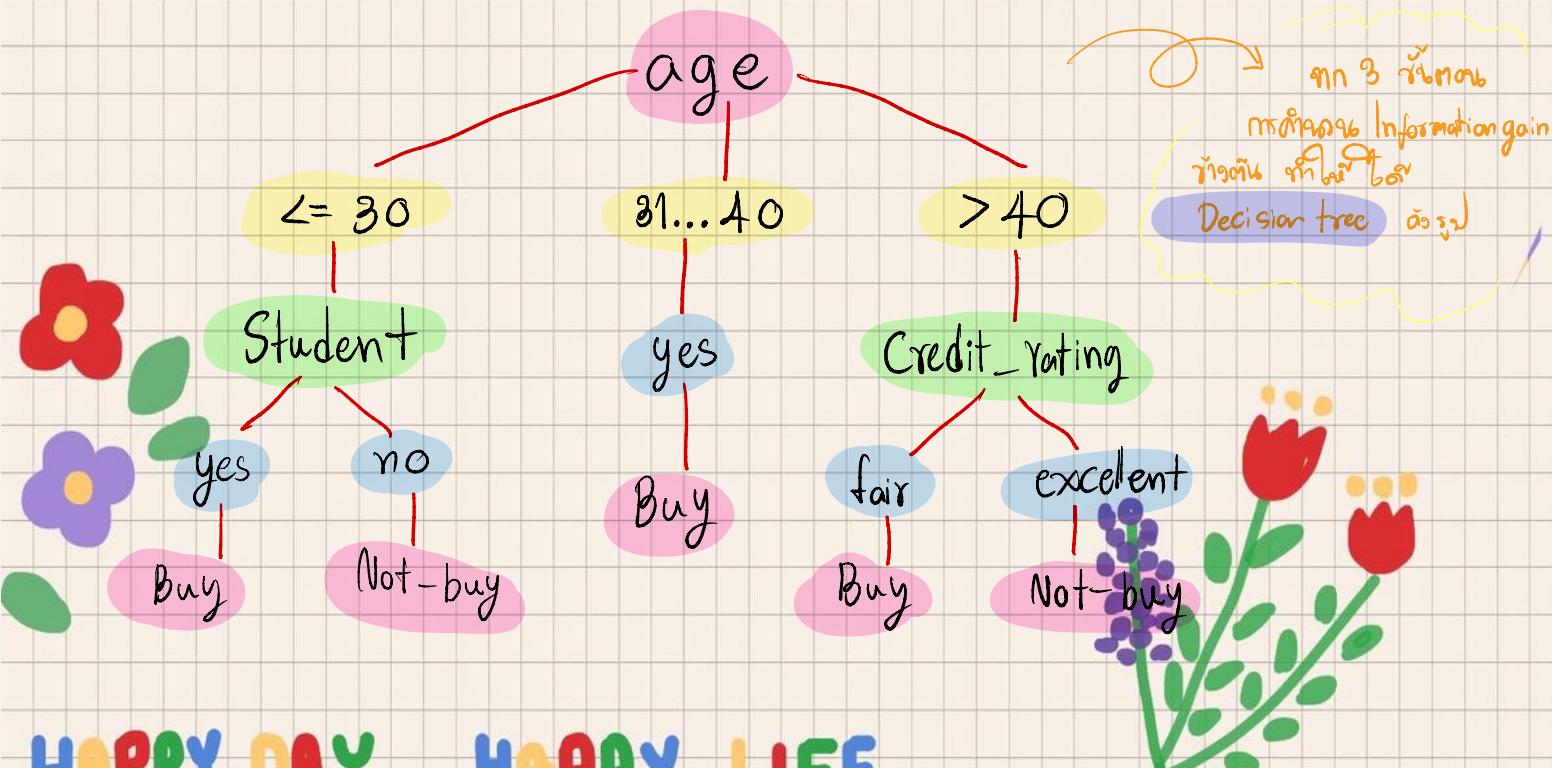
$$\text{Info}_{\text{credit\_rating}}(D) = \frac{3}{5} I(3,0) + \frac{2}{5} I(0,2) = 0$$

នៃ gain នឹង max

$$\text{gain}(\text{income}) = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.971 - 0.951 = 0.02$$

$$\text{gain}(\text{student}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.971 - 0.951 = 0.02$$

$$\text{gain}(\text{credit\_rating}) = \text{Info}(D) - \text{Info}_{\text{credit\_rating}}(D) = 0.971 - 0 = 0.971 \quad \text{max}$$



HAPPY DAY

HAPPY LIFE