

## Decision Tree or ໂັດໄນ້ມີຕົກສັນໃຈ

ກົດເລືອກ attribute ກ່ອນທີ່ສູງຈາກ Information gain ສັງເກດແລ້ວ

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

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

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

### Decision Tree Induction : Example

**data set**

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

Feature

class

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

- Class P : buys\_computer = "yes"
- Class N : buys\_computer = "no"

ຂະໜາດ

$$Info(D) = I(4,5) = -\frac{1}{14} \log_2(\frac{1}{14}) - \frac{3}{14} \log_2(\frac{3}{14})$$

$$= 0.940$$

step : 1

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

	Yes	No	
High	2	1	$= \frac{3}{14}$
Medium	4	2	$= \frac{6}{14}$
Low	3	1	$= \frac{4}{14}$

$$Info_{Income}(D) = \frac{4}{14} I(2,2) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1) = 0.911$$

	Yes	No	
Yes	1	1	$= \frac{2}{14}$
No	3	1	$= \frac{4}{14}$

$$Info_{student}(D) = \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4) = 0.789$$

	Yes	No	
Fair	1	1	$= \frac{2}{14}$
Excellent	3	3	$= \frac{6}{14}$

$$Info_{credit}(D) = \frac{8}{14} I(6,2) + \frac{6}{14} I(3,3) = 0.892$$

## Step : 2

un information gained តើខែន Gain ដែលធ្វើតុលានជារដ្ឋបាល Root Node

- gain (age) = 0.940 - 0.911 = 0.246 តាមរាយការណ៍
- gain (income) = 0.940 - 0.911 = 0.029
- gain (student) = 0.940 - 0.789 = 0.151
- gain (credit-rating) = 0.940 - 0.892 = 0.048

## Step : 4

រាយការណ៍ការចូល Feature សម្រាប់ root node

### 1 Age <= 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

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

yes	no
high	low
medium	medium

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

yes	no
high	low
medium	medium

$$\text{Info student}(D) = \frac{2}{5} I(1,0) + \frac{3}{5} I(0,3) = 0$$

yes	no
high	medium
medium	low

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

fair	excellent
medium	high

un Gain ដែលការពិភាក្សា

- gain (income) = Info(D) - Info income(D) = 0.971 - 0.4 = 0.571
- gain (student) = Info(D) - Info student(D) = 0.971 - 0 = 0.971
- gain (credit-rating) = Info(D) - Info credit(D) = 0.971 - 0.951 = 0.02

### 2 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

3

Age &gt; 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

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

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

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

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

### III Gain ကိုမျှတော်းခြင်း

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

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

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

### Decision tree

- Resulting tree:

