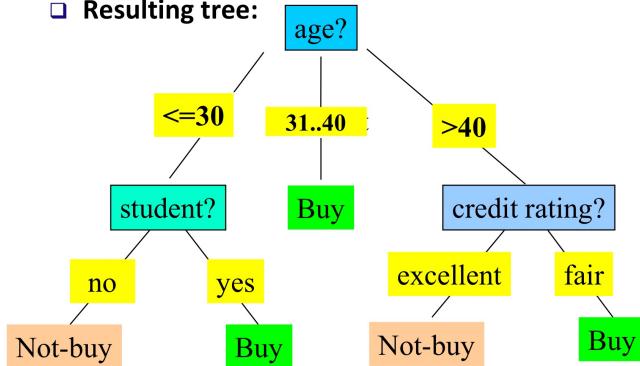


## HW 5

## Resulting tree:



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

Class  $\begin{cases} \text{yes} = 9 \\ \text{No} = 5 \end{cases}$

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

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

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

feature

$$\begin{aligned} \text{Info}_{\text{age}}(D) &= \frac{5}{14} I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2) \\ &= \frac{5}{14} \left[ -\frac{2}{5} \log_2 \left( \frac{2}{5} \right) - \frac{3}{2} \log_2 \left( \frac{3}{2} \right) \right] + \frac{4}{14} \left[ -\frac{4}{4} \log_2 \left( \frac{4}{4} \right) \right] + \frac{5}{14} \left[ -\frac{3}{5} \log_2 \left( \frac{3}{5} \right) - \frac{2}{5} \log_2 \left( \frac{2}{5} \right) \right] \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{income}}(D) &= \frac{4}{14} I(2,2) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1) \\ &= \frac{4}{14} \left[ -\frac{2}{4} \log_2 \left( \frac{1}{4} \right) - \frac{2}{4} \log_2 \left( \frac{1}{4} \right) \right] + \frac{6}{14} \left[ -\frac{4}{6} \log_2 \left( \frac{4}{6} \right) - \frac{2}{6} \log_2 \left( \frac{2}{6} \right) \right] + \frac{4}{14} \left[ -\frac{3}{4} \log_2 \left( \frac{3}{4} \right) - \frac{1}{4} \log_2 \left( \frac{1}{4} \right) \right] \\ &= 0.911 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{student}}(D) &= \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4) \\ &= \frac{7}{14} \left[ -\frac{6}{7} \log_2 \left( \frac{6}{7} \right) - \frac{1}{7} \log_2 \left( \frac{1}{7} \right) \right] + \frac{7}{14} \left[ -\frac{3}{7} \log_2 \left( \frac{3}{7} \right) - \frac{4}{7} \log_2 \left( \frac{4}{7} \right) \right] \\ &= 0.789 \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{credit-rating}}(D) &= \frac{8}{14} I(6,2) + \frac{6}{14} I(3,3) \\ &= \frac{8}{14} \left[ -\frac{6}{8} \log_2 \left( \frac{1}{8} \right) - \frac{2}{8} \log_2 \left( \frac{2}{8} \right) \right] + \frac{6}{14} \left[ -\frac{3}{6} \log_2 \left( \frac{3}{6} \right) - \frac{3}{6} \log_2 \left( \frac{3}{6} \right) \right] \\ &= 0.892 \end{aligned}$$

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

$$\begin{aligned} \text{Gain}(\text{age}) &= 0.940 - 0.694 = 0.246 \quad \# \text{ជាការការពួក}^{\dagger} \text{ (root node)} \\ \text{Gain}(\text{income}) &= 0.940 - 0.911 = 0.029 \\ \text{Gain}(\text{student}) &= 0.940 - 0.789 = 0.151 \\ \text{Gain}(\text{credit-rating}) &= 0.940 - 0.892 = 0.048 \end{aligned}$$

ផែកក្នុងទីនេះ feature មានតារ root node  $\Rightarrow$  age

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

$\angle = 30$

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

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

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

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

$$\text{Gain}(\text{income}) = 0.971 - 0.4 = 0.571$$

$$\text{Gain}(\text{student}) = 0.971 - 0 = 0.971 \# \text{node}$$

$$\text{Gain}(\text{credit-rating}) = 0.971 - 0.951 = 0.02$$

$30 \dots 40$

$$\text{Info}(D) = I(4,0)$$

# ព័ត៌មាន  $31 \dots 40$  buys\_computer នូវ yes ទៅនេះ

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

$> 40$

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

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

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

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

$$\text{Gain}(\text{income}) = 0.971 - 0.951 = 0.02$$

$$\text{Gain}(\text{student}) = 0.971 - 0.951 = 0.02$$

$$\text{Gain}(\text{credit-rating}) = 0.971 - 0 = 0.971 \# \text{node}$$

