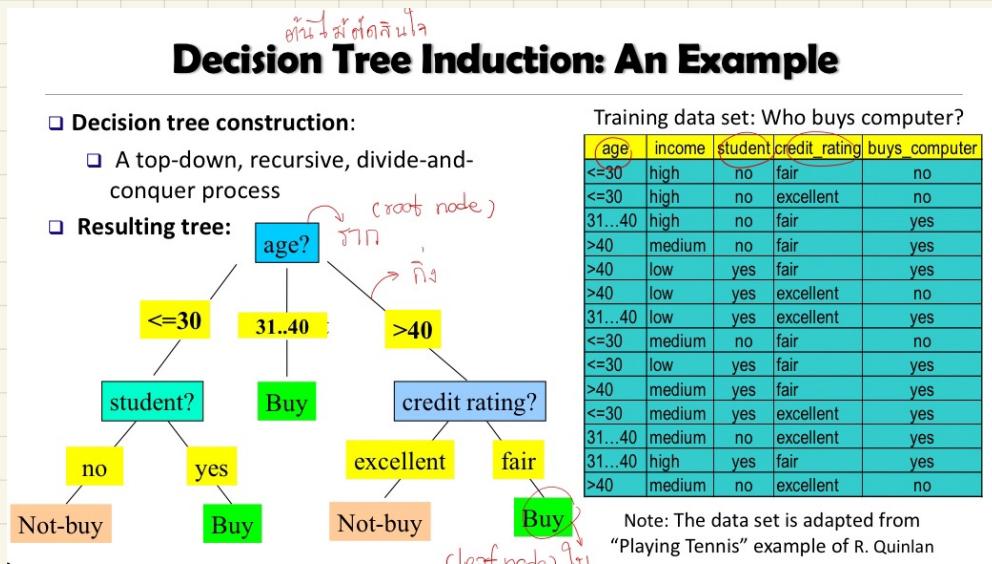


ប៊ូតុណុគារណ៍ Decision Tree : ពីរដ្ឋម៉ែត្រសិនកុំ



- Class P : buys_Computer = Yes = 9
- Class N : buys_Computer = no = 5

STEP 1

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

(entropy class)

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

STEP 2

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

(entropy feature)

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

↑ 31..40

$$= \frac{5}{14} \left[-\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \right] + \frac{4}{14} \left[-\frac{4}{4} \log_2\left(\frac{4}{4}\right) \right] + \frac{5}{14} \left[-\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) \right]$$

$$= 0.694$$

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

$$= \frac{4}{14} \left[-\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) \right] + \frac{6}{14} \left[-\frac{4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right) \right] + \frac{4}{14} \left[-\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right) \right]$$

$$= 0.911$$

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

$$= \frac{7}{14} \left[-\frac{3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right) \right] + \frac{7}{14} \left[-\frac{6}{7} \log_2\left(\frac{6}{7}\right) - \frac{1}{7} \log_2\left(\frac{1}{7}\right) \right]$$

$$= 0.789$$

- $\text{Info}_{\text{credit_rating}}(D) = \frac{8}{14} I(6,2) + \frac{6}{14} I(3,3)$

$$= \frac{8}{14} \left[-\frac{6}{8} \log_2 \left(\frac{6}{8} \right) - \frac{2}{8} \log_2 \left(\frac{2}{8} \right) \right] + \frac{6}{14} \left[-\frac{3}{6} \log_2 \left(\frac{3}{6} \right) - \frac{3}{6} \log_2 \left(\frac{3}{6} \right) \right]$$

$$= 0.892$$

STEP 3

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

សម្រាប់ទីតាំងថ្មីនៃ root node

- Gain(age) = 0.940 - 0.694 = 0.246
- 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

Gain(age) ជំនួយស្ថិតិថ្មីនៃ root node

STEP 4

នឹងករណីមាន feature នៅលើ age → នៅ root node

- $<= 30$

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

$$\text{Info}(D) = I(2,3) = 0.971$$

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

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

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

ពាក្យសារ Gain

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

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

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

- Gain(student) ជំនួយស្ថិតិថ្មីនៃ node នៅលើ $<= 30$

- 31 - 40

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

$$\text{age } 31-40 \text{ buys_computer} = \text{Yes} = 4$$

$$= \text{no} = 0$$

នូវ Yes

• >40

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
<=30...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

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

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

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

ກຳນົດ Gain

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

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

$$\text{Gain}(\text{credit_rating}) = 0.971 - 0 = 0.971$$

- Gain(credit_rating) ສະຫຼຸບຜັດໄດ້ໃນ node ຖ້າ age > 40

STEP 5

