

Info (D)

$$\text{Info}(D) = I(9,5) = -\frac{9}{14} \log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) \log_2\left(\frac{5}{14}\right) \\ = 0.94$$

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

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

$$I(4,0) = -\frac{4}{4} \log_2\left(\frac{4}{4}\right) - \frac{0}{4} \log_2\left(\frac{0}{4}\right) = 0$$

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

$$\text{Info}_{\text{age}}(D) = \frac{5}{14} (0.971) + \frac{4}{14} (0) + \frac{5}{14} (0.971) = 0.694$$

Gain (age)

$$\text{Gain}(\text{age}) = 0.94 - 0.694 = 0.246$$

$$\underline{\underline{\text{Info}_{\text{income}}(D)}} = \left[\frac{4}{14} I(2,2) \right] + \left[\frac{6}{14} I(4,2) \right] + \left[\frac{4}{14} I(3,1) \right]$$

$$I(2,2) = -\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) = 1$$

$$I(4,2) = -\frac{4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right) = 0.918$$

$$I(3,1) = -\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right) = 0.811$$

$$\text{Info}_{\text{income}}(D) = \frac{4}{14} (1) + \frac{6}{14} (0.918) + \frac{4}{14} (0.811) = 0.911$$

Gain (income)

$$\text{Gain}(\text{income}) = 0.94 - 0.911 = 0.029$$

Info student (D)

Y

N

2

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

$$I(6,1) = -\frac{6}{7} \log_2\left(\frac{6}{7}\right) - \frac{1}{7} \log_2\left(\frac{1}{7}\right) = 0.592$$

$$I(3,4) = -\frac{3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right) = 0.985$$

$$\text{Info}_{\text{student}}(D) = \frac{7}{14} (0.592) + \frac{7}{14} (0.985) = 0.789$$

$$\text{Gain}(\text{student}) = 0.94 - 0.789 = 0.151$$

fair

excellent

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

$$I(6,2) = -\frac{6}{8} \log_2\left(\frac{6}{8}\right) - \frac{2}{8} \log_2\left(\frac{2}{8}\right) = 0.811$$

$$I(3,3) = -\frac{3}{6} \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \log_2\left(\frac{3}{6}\right) = 1$$

$$\text{Gain}(\text{credit rating}) = 0.94 - 0.892 = 0.048$$

Gain

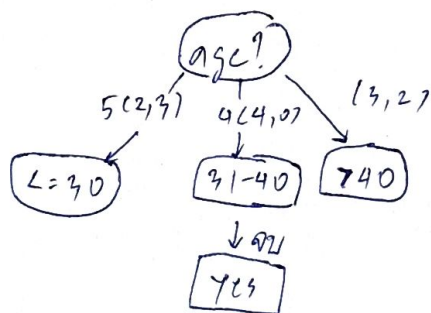
$$\text{Gain}_{\text{age}} = 0.246$$

$$\text{Gain}_{\text{income}} = 0.029$$

$$\text{Gain}_{\text{student}} = 0.151$$

$$\text{Gain}_{\text{credit rating}} = 0.048$$

เลือก Gain ที่มีค่ามากที่สุดมาทำ node Gain (age)



age (L=30)

3

Info (0) vs age (L=30)

$$\text{Info}(0) = I(2, 1) = 0.991$$

$$\text{Info}_{\text{income}}(0) \text{ vs age } (L=30) = \overset{\text{high}}{\frac{2}{5}} I(0, 2) + \overset{\text{medium}}{\frac{2}{5}} I(1, 1) + \overset{\text{low}}{\frac{1}{5}} I(1, 0)$$

$$I(0, 2) = -\frac{0}{2} \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \log_2\left(\frac{2}{2}\right) = 0$$

$$I(1, 1) = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

$$I(1, 0) = -\frac{1}{1} \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \log_2\left(\frac{0}{1}\right) = 0$$

$$\text{Info}_{\text{income}}(0) \text{ vs age } (L=30) = \frac{2}{5}(0) + \frac{2}{5}(1) + \frac{1}{5}(0) = 0.4$$

Gain (income) vs age (L=30)

$$\text{Gain (income)} \text{ vs age } (L=30) = 0.991 - 0.4 = 0.591$$

Info_{student} (0) vs age (L=30)

$$\text{Info}_{\text{student}}(0) \text{ vs age } (L=30) = \overset{\text{Yes}}{\frac{2}{5}} I(2, 0) + \overset{\text{No}}{\frac{3}{5}} I(0, 3)$$

Yes (buy - computer), No → no buy - computer

นักเรียนซื้อ student หรือ ไม่ซื้อคอมพิวเตอร์ก็ได้

age (740)

$$\text{Info}(0) \text{ vs age } (740) = I(3, 2) = 0.991$$

Info

medium

Low

$$\text{Info}_{\text{income}}(0) \text{ vs age } (740) = \frac{3}{5} I(2, 1) + \frac{2}{5} I(1, 1)$$

$$I(2, 1) = -\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) = 0.918$$

$$I(1, 1) = 1$$

$$\text{Info}_{\text{income}}(0) \text{ vs age } (740) = \frac{3}{5}(0.918) + \frac{2}{5}(1) = 0.951$$

Gain (income) vs age (740)

$$\text{Gain (income)} \text{ vs Age } (740) = 0.991 - 0.951 = 0.04$$

Info student (D) var age (740)

yes

No

$$\text{Info student (D) var age (740)} = \frac{2}{5} I(2,1) + \frac{1}{5} I(1,1)$$

$$I(2,1) = -\frac{2}{5} \log_2(2) \left(\frac{2}{5}\right) - \frac{1}{5} \log_2(2) \left(\frac{1}{5}\right) = 0.918$$

$$\text{Info student (D) var age (740)} = \frac{2}{5} (0.918) + \frac{1}{5} (1) = 0.951$$

gain (student var age (740))

$$\text{gain (student, age (740))} = 0.971 - 0.951 = 0.02$$

$$\text{Info credit-rating (D) var age (740)} = 0.971 - 0.951 = 0.02$$

fair

Excellent

$$\text{Info credit-rating (D) var age (740)} = \frac{2}{5} I(3,1) + \frac{1}{5} I(0,1)$$

fair \rightarrow yes, buy-computer, excellent \rightarrow No (buy-computer)

now, how credit-rating also associated

