

Homework 2

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

Yes No

$$\text{Info}(D) = I(9,5) = -\frac{9}{14} \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \log_2\left(\frac{5}{14}\right) = 0.9210$$

* Feature

- age ≤ 30 $31 \dots 40$ > 40

$$\begin{aligned} \text{Info}_{\text{age}}(D) &= \frac{5}{14} I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2) \\ &= \frac{5}{14} \left[-\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \right] + \frac{4}{14} \left[-\frac{4}{14} \log_2\left(\frac{4}{14}\right) - \frac{0}{14} \log_2\left(\frac{0}{14}\right) \right] \\ &\quad + \frac{5}{14} \left[-\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \right] = 0.694 \end{aligned}$$

- income *high* *medium* *low*

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

$$= \frac{4}{14} \left[-\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right] + \frac{6}{14} \left[-\frac{4}{6} \log_2 \left(\frac{4}{6} \right) - \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right] \\ + \frac{4}{14} \left[-\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right] = 0.911$$

- student

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

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

$$= 0.788$$

- credit rating

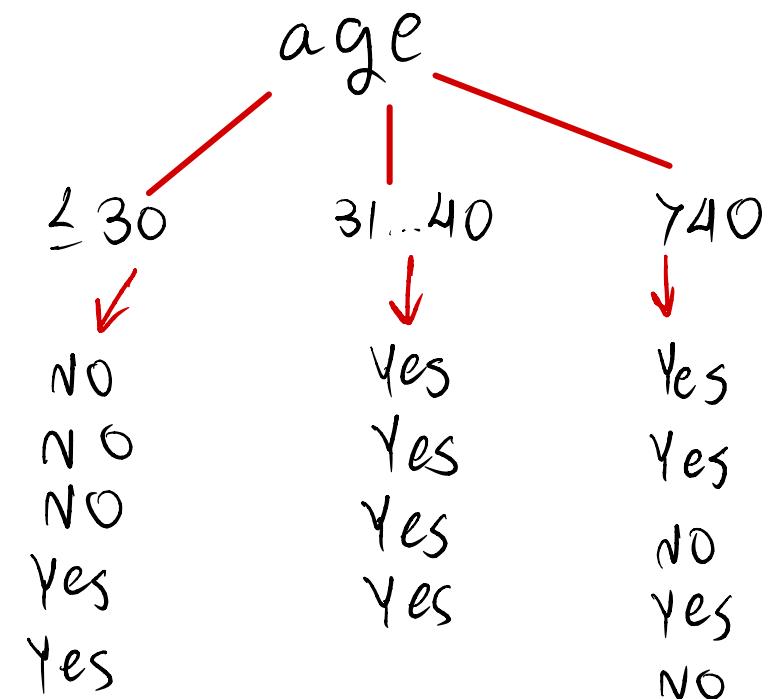
$$\begin{aligned} \text{Info}_{\text{cra.}}(0) &= \frac{6}{14} I(3,3) + \frac{8}{14} I(6,2) \\ &= \frac{6}{14} \left[-\frac{3}{6} \log_2 \left(\frac{3}{6}\right) - \frac{3}{6} \log_2 \left(\frac{3}{6}\right) \right] + \frac{8}{14} \left[-\frac{6}{8} \log_2 \left(\frac{6}{8}\right) - \frac{2}{8} \log_2 \left(\frac{2}{8}\right) \right] \\ &\approx 0.892 \end{aligned}$$

$$\text{Gain}(\text{age}) = 0.910 - 0.694 = 0.216$$

$$\text{Gain}(\text{income}) = 0.910 - 0.911 = 0.029$$

$$\text{Gain}(\text{student}) = 0.910 - 0.788 = 0.152$$

$$\text{Gain}(\text{credit-rating}) = 0.910 - 0.892 = 0.028$$



Age ≤ 30

* class

$$\text{Info}_{\leq 30}(D) = I_{(2,3)}^{\text{Yes } \text{No}} = -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) = 0.971$$

* Feature

- income high medium low

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

$$= \frac{2}{5} \left[-\frac{0}{2} \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \log_2\left(\frac{2}{2}\right) \right] + \frac{2}{5} \left[-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] \\ + \frac{1}{5} \left[-\frac{1}{1} \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \log_2\left(\frac{0}{1}\right) \right] = 0.4$$

- student

$$\text{Info}_{\text{income}}(D) = \frac{2}{5} I_{(2,0)}^{\text{Yes}} + \frac{3}{5} I_{(0,3)}^{\text{No}}$$

$$= \frac{2}{5} \left[-\frac{2}{2} \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \log_2\left(\frac{0}{2}\right) \right] + \frac{3}{5} \left[-\frac{0}{3} \log_2\left(\frac{0}{3}\right) - \frac{3}{3} \log_2\left(\frac{3}{3}\right) \right] = 0$$

- credit rating

$$Info_{\text{cra...}}(D) = \frac{2}{5} I(\text{Excellent}) + \frac{3}{5} I(\text{fair})$$

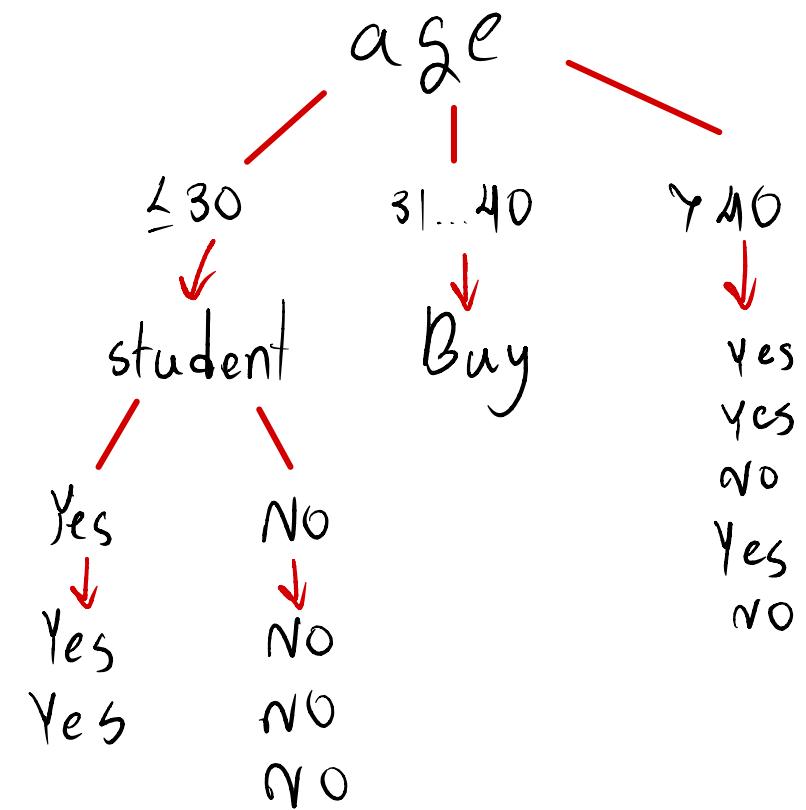
$$= \frac{2}{5} \left[-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right] + \frac{3}{5} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right]$$

$$\approx 0.951$$

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

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

$$\text{Gain(credit rating)} = 0.971 - 0.951 = 0.020$$



Age > 40

* Class

Yes NO

$$\text{Info}_{>40}(D) = I(3,2) = -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) = 0.971$$

* Feature

- Income

Medium Low

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

$$= \frac{3}{5} \left[-\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) \right] + \frac{2}{5} \left[-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] = 0.951$$

- student

Yes

NO

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

$$= \frac{3}{5} \left[-\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right) \right] + \frac{2}{5} \left[-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \right]$$

$$= 0.951$$

- credit rating

$$\text{Info}_{\text{cra..}}(D) = \frac{2}{5} \overset{\text{excellent}}{I(0,2)} + \frac{3}{5} \overset{\text{fair}}{I(3,0)}$$

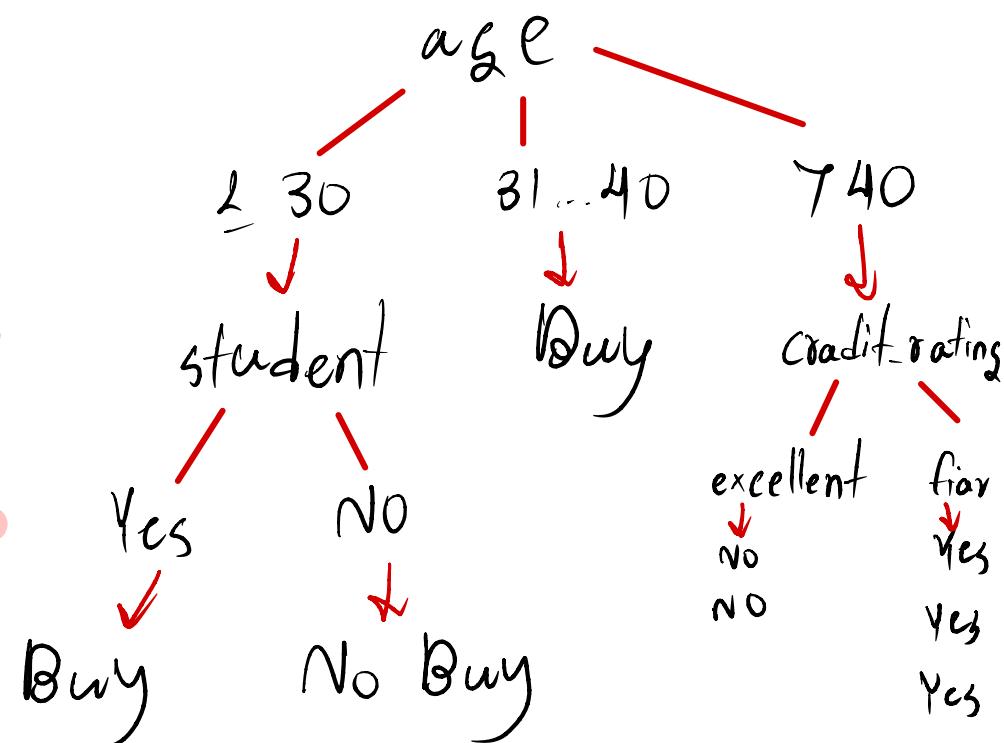
$$= \frac{2}{5} \left[-\frac{0}{2} \log_2 \left(\frac{0}{2} \right) - \frac{2}{2} \log_2 \left(\frac{2}{2} \right) \right] + \frac{3}{5} \left[-\frac{3}{3} \log_2 \left(\frac{3}{3} \right) - \frac{0}{3} \log_2 \left(\frac{0}{3} \right) \right]$$

—

$$\text{Gain(income)} = 0.971 - 0.951 = 0.020$$

$$\text{Gain (student)} = 0.971 - 0.951 = 0.020$$

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



Final

