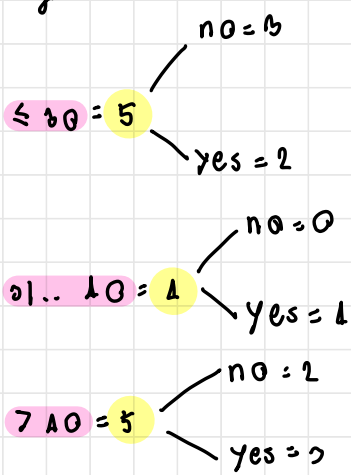
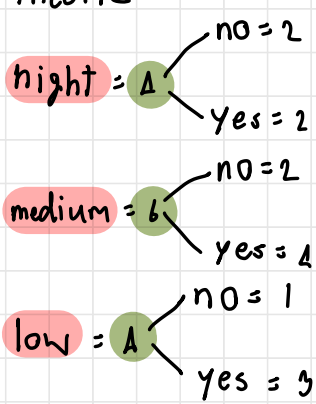
HW5

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

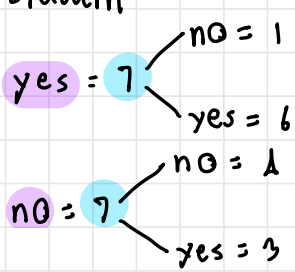
age



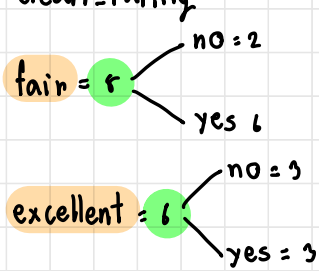
income



student



credit_rating



- Class

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

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

$$= 0.940$$

- Features

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

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

$$= \frac{5}{14}\left[-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right] + \frac{4}{14}\left[-\frac{4}{4}\log_2\left(\frac{4}{4}\right) - \frac{0}{4}\log_2\left(\frac{0}{4}\right)\right]$$

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

$$= 0.694$$

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

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

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

$$= 0.911$$

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

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

$$= 0.788$$

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

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

$$= 0.892$$

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

$Gain(age) = Info(D) - Info_{age}(D)$

$= 0.940 - 0.694$

$= 0.246$ root node

$Gain(income) = Info(D) - Info_{income}(D)$

$= 0.940 - 0.911$

$= 0.029$

$Gain(Student) = Info(D) - Info_{student}(D)$

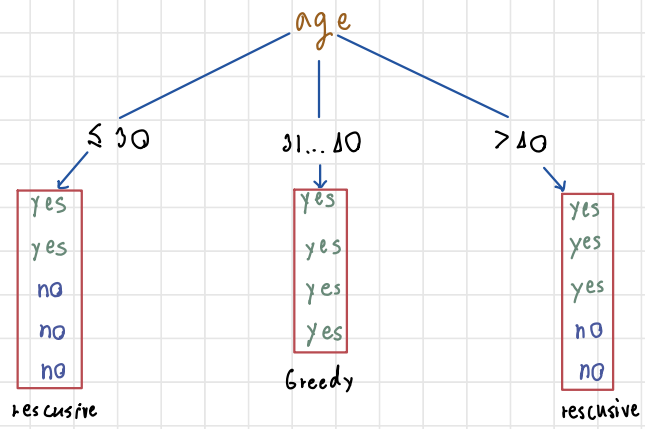
$= 0.940 - 0.788$

$= 0.152$

$Gain(Credit\_rating) = Info(D) - Info_{Credit\_rating}(D)$

$= 0.940 - 0.692$

$= 0.048$



• recusive age ≤ 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

• Class

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

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

$$= 0.971$$

• Features

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

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

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

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

$$= 0.4$$

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

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

$$= 0$$

$$Info_{credit\_rating}(D) = \frac{3}{5} I(1,2) + \frac{2}{5} I(1,1)$$

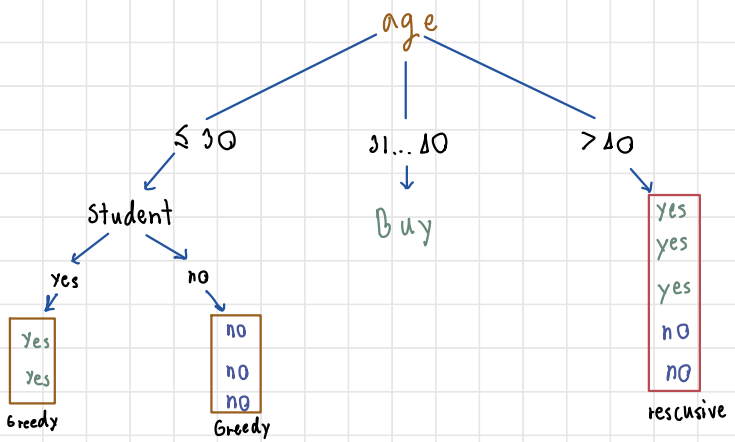
$$= \frac{3}{5}\left[-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$$

$$= 0.951$$

$$Gain(income) = Info(D) - Info_{income}(D)$$
$$= 0.971 - 0.4$$
$$= 0.571$$

$$Gain(Student) = Info(D) - Info_{student}(D)$$
$$= 0.971 - 0$$
$$= 0.971$$ → root node

$$Gain(Credit\_rating) = Info(D) - Info_{Credit\_rating}(D)$$
$$= 0.971 - 0.951$$
$$= 0.02$$



• recusive age > 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

• Class

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

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

$$= 0.0971$$

• Features

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

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

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

$$= 0.951$$

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

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

$$= 0.951$$

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

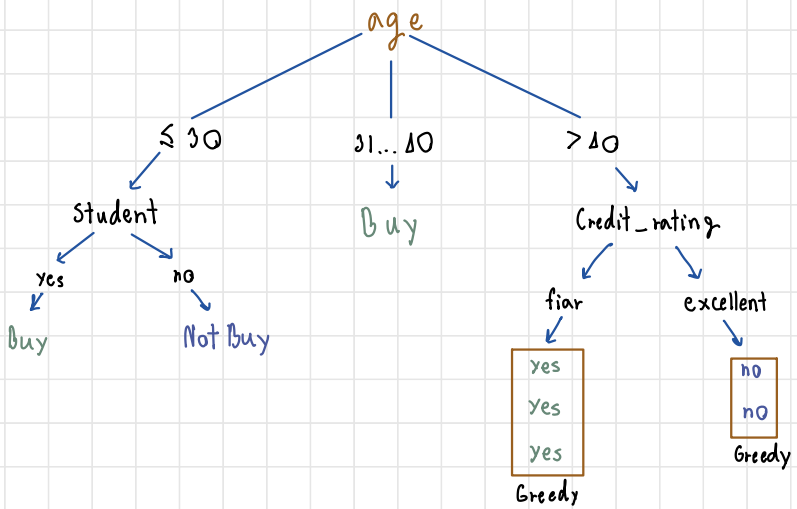
$$= \frac{3}{5}\left[-\frac{3}{3}\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right)\right] + \frac{2}{5}\left[-\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right)\right]$$

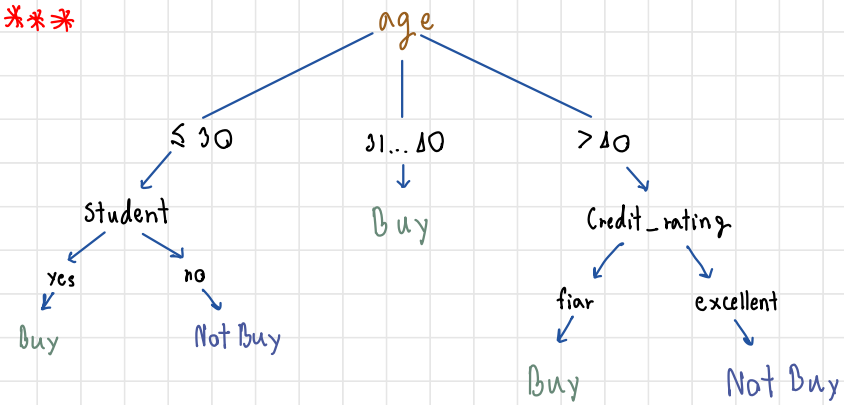
$$= 0$$

$$Gain(income) = Info(D) - Info_{income}(D)$$
$$= 0.971 - 0.951$$
$$= 0.02$$

$$Gain(Student) = Info(D) - Info_{student}(D)$$
$$= 0.971 - 0.951$$
$$= 0.02$$

$$Gain(Credit\_rating) = Info(D) - Info_{credit\_rating}(D)$$
$$= 0.971 - 0$$
$$= 0.971$$ root node




age
≤ 30
31...40
>40
Student
Buy
Credit_rating
yes
no
fiar
excellent
Buy
Not Buy
Buy
Not Buy