

Min sample leaf & Max feature

by กลุ่มทดลองประเสริฐ & i hate monday

ສມາຜິກ

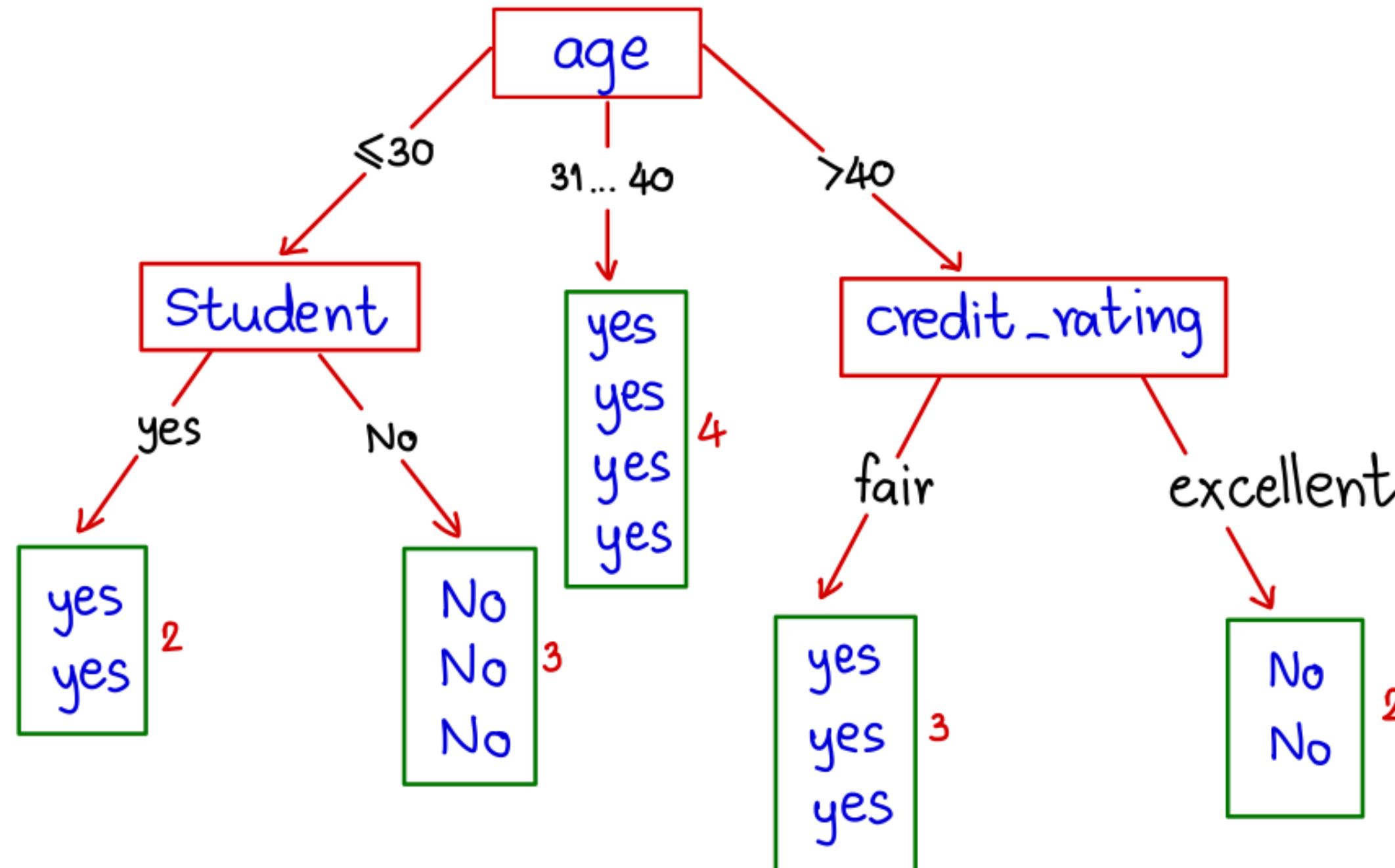
| | |
|--------------------------|-------------|
| ຂວัญສວຣຣຄ ກອງປະເສຣີ້ຈູ້ | 643020498-9 |
| ນຣີສຣາ ໂຍຮາຈັນທົກ | 643020504-0 |
| ນຣີການຕໍ່ ປ້າຍຍາຮັກເງົ່າ | 643020505-8 |
| ພິມພົກາ ຍອດສຣີ | 643020511-3 |
| ຮັຕະຫຼາກຮ ນາມວົງສ | 643020517-1 |
| ສຸ່ຈາດາ ອຸປພົງເງົ່າ | 643020524-4 |
| ຮົຕິພຣ ທົງພຸດຊາ | 643021267-3 |
| ສຸນິສາ ອຸດມບັນຮຸ | 643021278-8 |

TRAINING DATA SET

WHO BUYS COMPUTER?

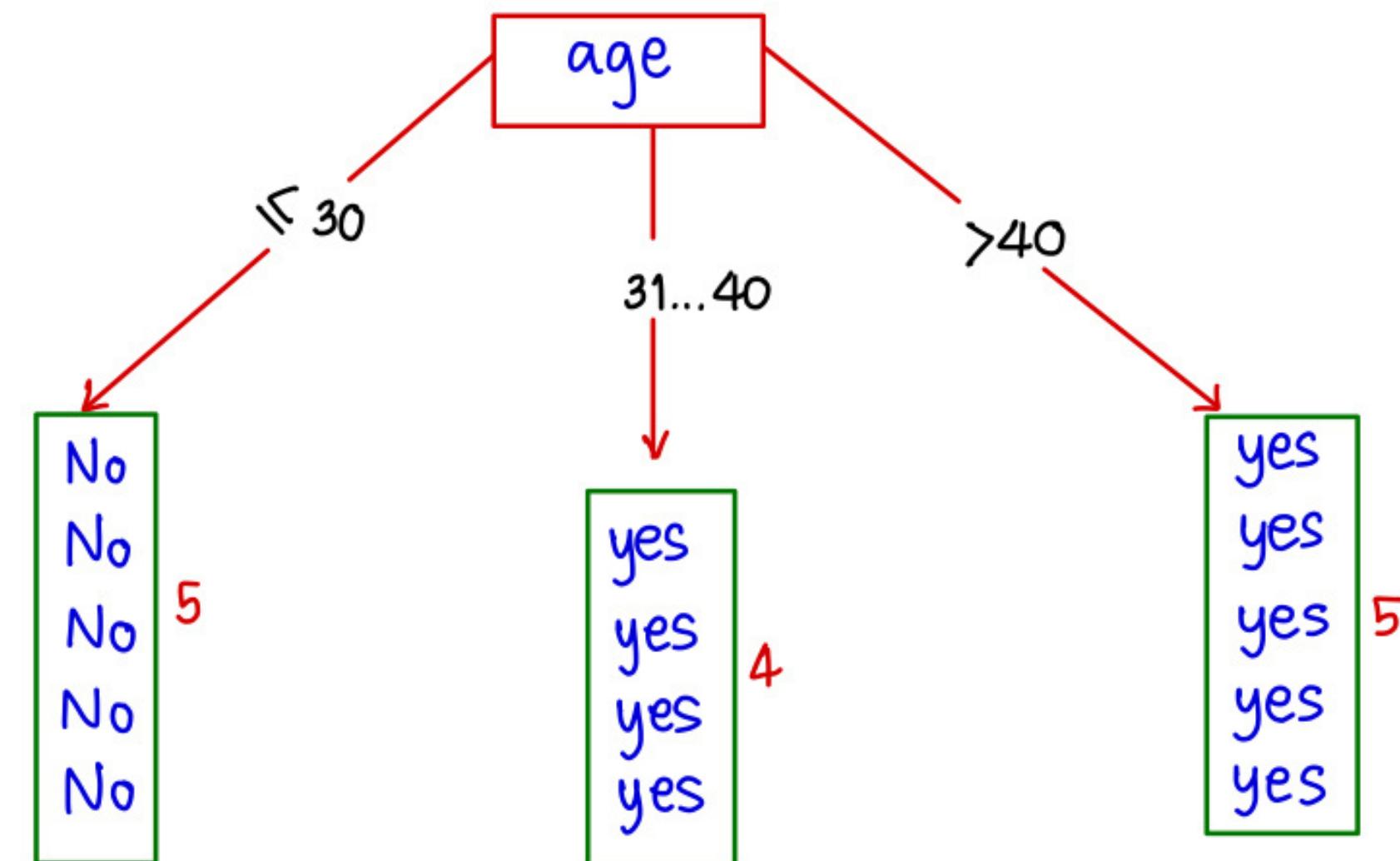
| age | income | student | credit_rating | buys_computer |
|---------|--------|---------|---------------|---------------|
| <=30 | high | no | fair | no |
| <=30 | high | no | excellent | no |
| 31...40 | high | no | fair | yes |
| >40 | medium | no | fair | yes |
| >40 | low | yes | fair | yes |
| >40 | 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 |

DECISION TREE



MIN SAMPLE LEAF = 4

MIN SAMPLE LEAF = 4



MIN SAMPLE LEAF = 4

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

$= 0.940$

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

$= 0.694$

| 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 |

MIN SAMPLE LEAF = 4

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

| 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 |

MIN SAMPLE LEAF = 4

$$\begin{aligned}
 \text{Info}_{\text{student}}(D) &= \frac{4}{14} I(6,1) + \frac{7}{14} I(3,4) \\
 &= \frac{4}{14} \left[-\frac{6}{4} \log_2 \left(\frac{6}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right] + \\
 &\quad \frac{7}{14} \left[-\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{4}{4} \log_2 \left(\frac{4}{4} \right) \right] \\
 &= 0.488
 \end{aligned}$$

| 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 |

MIN SAMPLE LEAF = 4

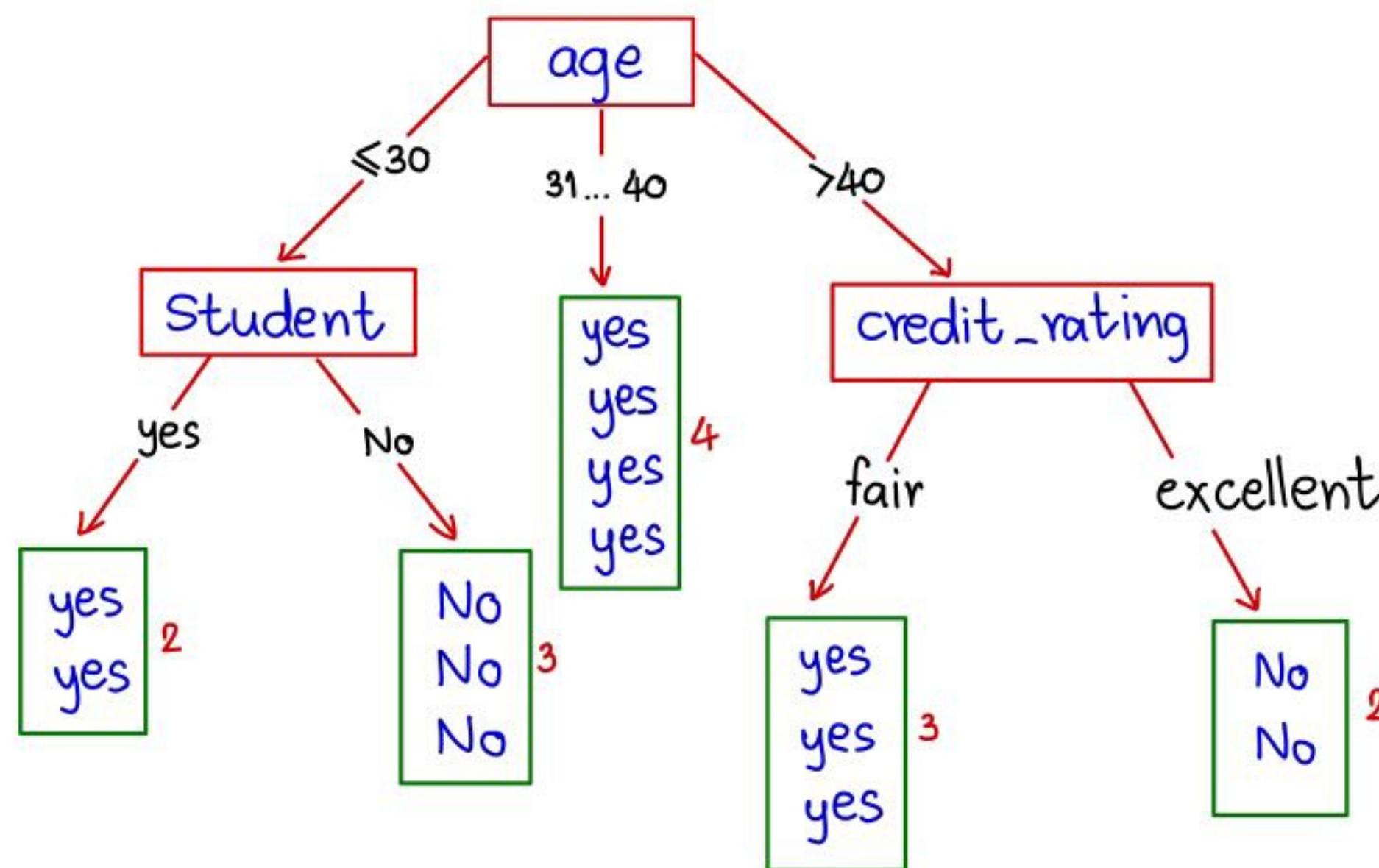
Info_{credit_rating}

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

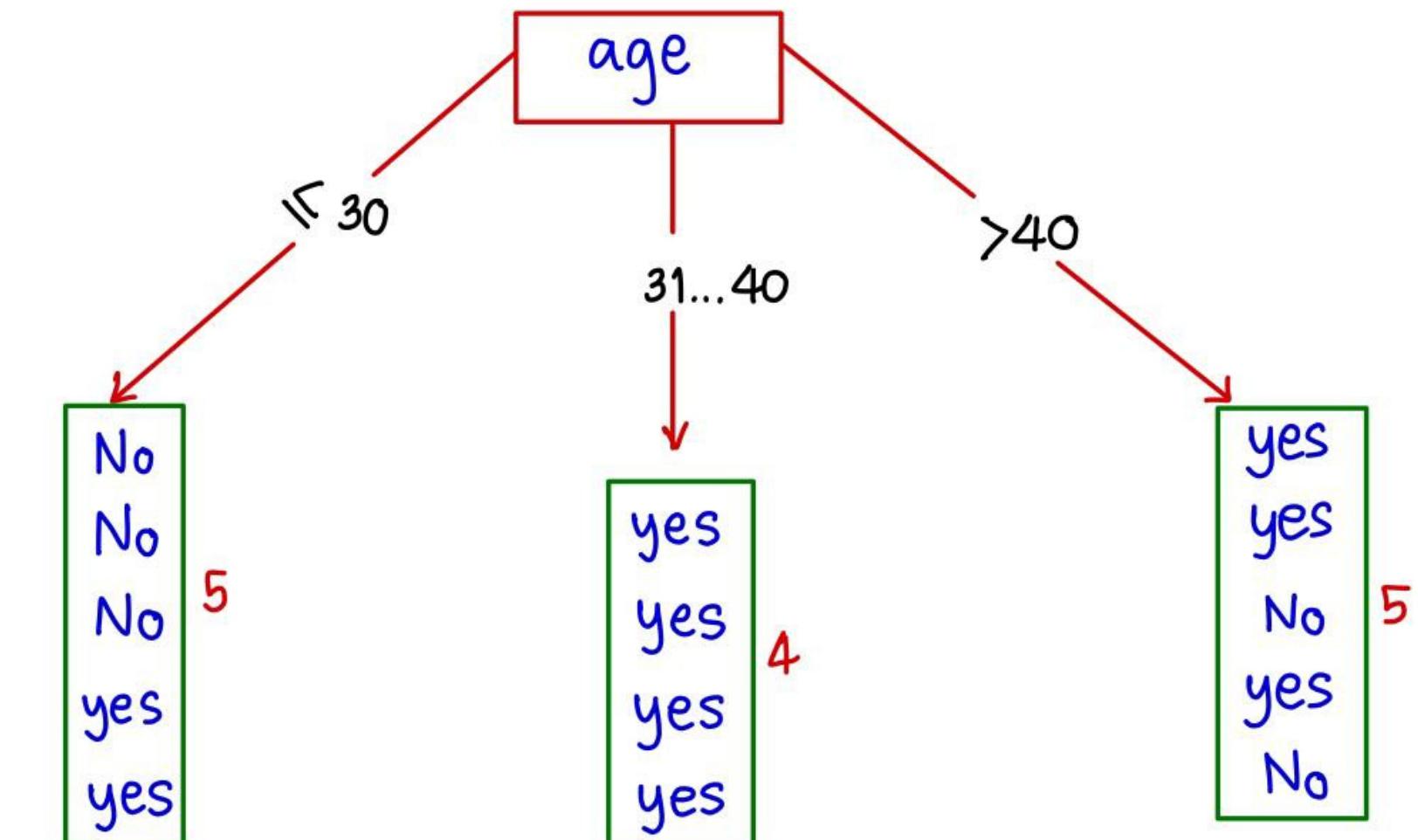
| | | | | |
|--------------------|---|-----------------|---|-----------|
| Gain age | = | $0.940 - 0.694$ | = | $0.246 *$ |
| Gain income | = | $0.940 - 0.911$ | = | 0.029 |
| Gain student | = | $0.940 - 0.788$ | = | 0.152 |
| Gain credit_rating | = | $0.940 - 0.892$ | = | 0.048 |

| 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 |

MIN SAMPLE LEAF



NOT SET UP



MIN SAMPLE LEAF = 4

MAX FEATURE

MAX FEATURE = 2

สูบ 2 feature จากทั้งหมด 4 feature(age, income, student, credit_rating)

เลือก credit_rating และ income

Max_feature = 2

① สูบ 2 ฟีเจอร์ จาก age, income, student, credit_rating
จะเหลือ income และ credit_rating

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

income

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

| 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 |

MAX FEATURE = 2

สุ่ม 2 feature จากทั้งหมด 4 feature(age, income, student, credit_rating)
ได้ credit_rating และ income

credit_rating

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

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

$$= 0.892$$

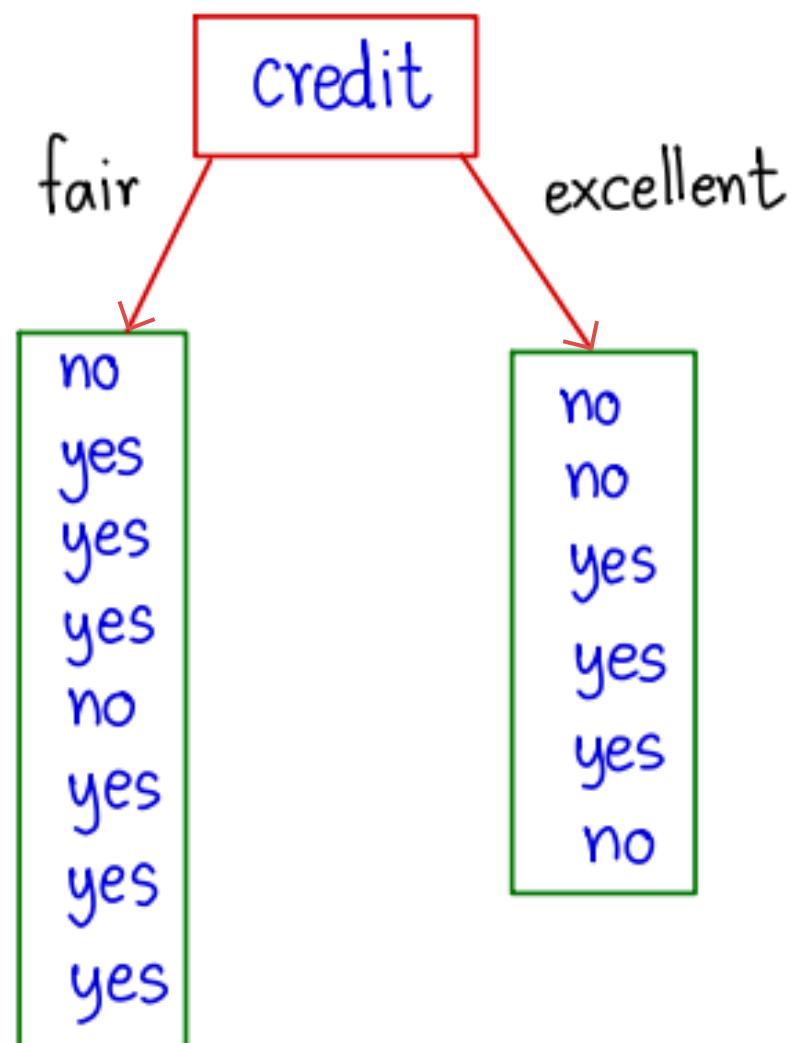
| age | income | student | credit_rating | buys_computer |
|---------|--------|---------|---------------|---------------|
| <=30 | high | no | fair | no |
| <=30 | high | no | excellent | no |
| 31...40 | high | no | fair | yes |
| >40 | medium | no | fair | yes |
| >40 | low | yes | fair | yes |
| >40 | 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 |

MAX FEATURE = 2

สุ่ม 2 feature จากทั้งหมด 4 feature(age, income, student, credit_rating)
ได้ credit_rating และ income

$$\text{Gain (income)} = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.940 - 0.911 = 0.029$$

$$\begin{aligned}\text{Gain (credit_rating)} &= \text{Info}(D) - \text{Info}_{\text{credit_rating}}(D) = 0.940 - 0.892 \\ &= 0.048 *\end{aligned}$$



MAX FEATURE = 2

`credit_rating = fair`

สูบ 2 feature จากทั้งหมด 3 feature (age, income, student)
ได้ income และ age

② ใจดี credit_rating มาก

2.1 credit_rating = fair

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

สูบ 2 ใจดี จากราก age, income, student และตัว income ตัว age

| 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 |

MAX FEATURE = 2

`credit_rating = fair`

สูม 2 feature จากทั้งหมด 3 feature(age, income, student)
ได้ income และ age

info_{income} (D)

$$\begin{aligned}
 &= \frac{3}{8} I(2,1) + \frac{3}{8} I(2,1) + \frac{2}{8} I(2,0) \\
 &= \frac{3}{8} \left[-\frac{2}{3} \log_2 \left(\frac{2}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) \right] + \frac{3}{8} \left[-\frac{2}{3} \log_2 \left(\frac{2}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) \right] \\
 &\quad + \frac{2}{8} \left[-\frac{2}{2} \log_2 \left(\frac{2}{2}\right) - \frac{0}{2} \log_2 \left(\frac{0}{2}\right) \right] \\
 &= 0.396
 \end{aligned}$$

| 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 |

MAX FEATURE = 2

`credit_rating = fair`

สูบ 2 feature จากทั้งหมด 3 feature(age, income, student)
ได้ income และ age

$$\begin{aligned} \text{info}_{\text{age}}(D) &= \frac{3}{8} I(1,2) + \frac{2}{8} I(2,0) + \frac{3}{8} I(3,0) \\ &= \frac{3}{8} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right) \right] + \frac{2}{8} \left[-\frac{2}{8} \log_2 \left(\frac{2}{2}\right) - \frac{0}{2} \log_2 \left(\frac{0}{2}\right) \right] \\ &\quad + \frac{3}{8} \left[-\frac{3}{3} \log_2 \left(\frac{3}{3}\right) - \frac{0}{3} \log_2 \left(\frac{0}{3}\right) \right] \\ &= 0.344 \end{aligned}$$

| 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 |

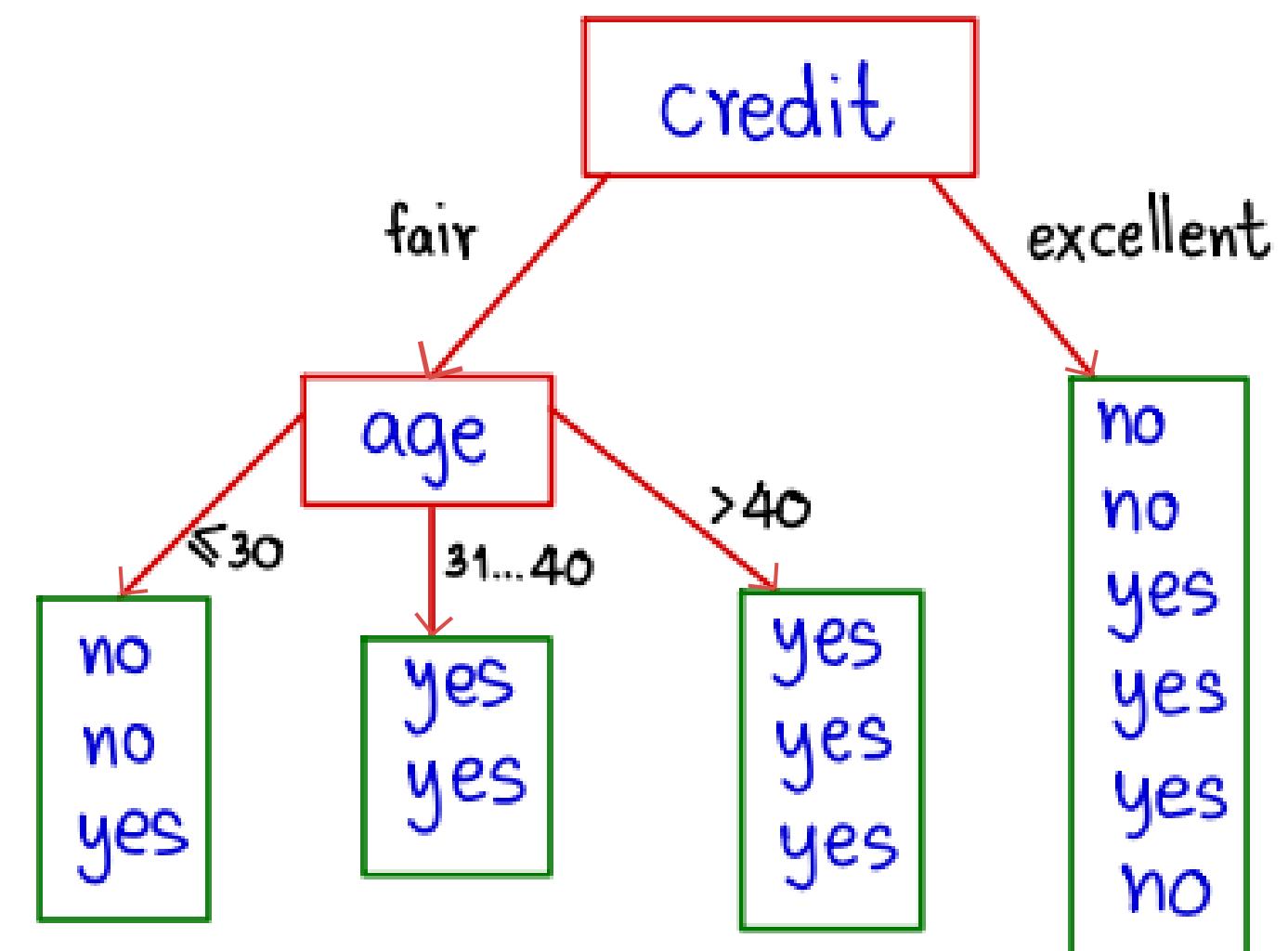
MAX FEATURE = 2

credit_rating = fair

สูง 2 feature จากทั้งหมด 3 feature(age, income, student)
ได้ income และ age

$$\text{Gain (income)} = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.811 - 0.396 = 0.415$$

$$\text{Gain (age)} = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.811 - 0.344 = 0.467 *$$



MAX FEATURE = 2

credit_rating = fair
Age <= 30

สูง 2 feature จากทั้งหมด 2 feature (student, income)
ได้ student และ income

กรณี age <= 30

กรณี student , income แบบ

credit = fair , age <= 30

Info (D) = I(1,2)

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

| 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 |

MAX FEATURE = 2

credit_rating = fair
Age <= 30

สูบ 2 feature จากทั้งหมด 2 feature (student, income)
เลือก student และ income

$$\begin{aligned} \text{Info}_{\text{student}}(D) &= \frac{2}{3} I(0,2) + \frac{1}{3} I(1,0) \\ &= \frac{2}{3} \left[-\frac{0}{2} \log_2 \left(\frac{0}{2} \right) - \frac{1}{2} \log_2 \left(\frac{2}{2} \right) \right] + \frac{1}{3} \left[-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) - \frac{0}{1} \log_2 \left(\frac{0}{1} \right) \right] \\ &= 0 \end{aligned}$$

| 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 |

MAX FEATURE = 2

`credit_rating = fair`
`Age <= 30`

សូម 2 feature ទៅក្នុងលាង 2 feature (student, income)
 តាម student និង income

$$\begin{aligned}
 \text{Info}_{\text{income}}(D) &= \frac{1}{3} I(0,1) + \frac{1}{3} I(0,1) + \frac{1}{3} I(1,0) \\
 &= \frac{1}{3} \left[-\frac{0}{1} \log_2 \left(\frac{0}{1} \right) - \frac{1}{1} \log_2 \left(\frac{1}{1} \right) \right] + \frac{1}{3} \left[-\frac{0}{1} \log_2 \left(\frac{0}{1} \right) - \frac{1}{1} \log_2 \left(\frac{1}{1} \right) \right]^0 \\
 &\quad + \frac{1}{3} \left[-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) - \frac{0}{1} \log_2 \left(\frac{0}{1} \right) \right]^0 \\
 &= 0
 \end{aligned}$$

| 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 |

MAX FEATURE = 2

credit_rating = fair
Age <= 30

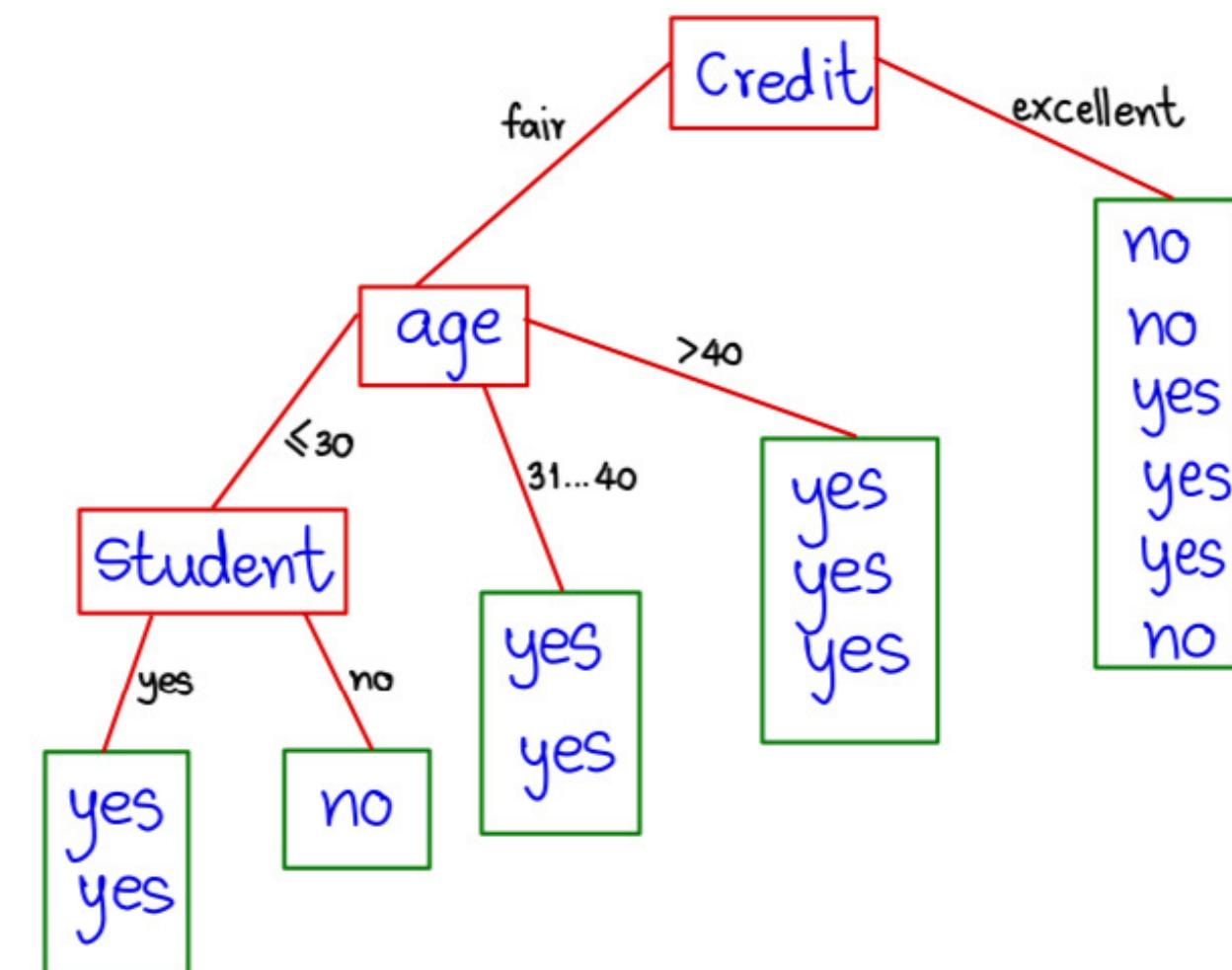
สุ่ม 2 feature จากทั้งหมด 2 feature(student, income)

ได้ student และ income

$$\text{Gain}(\text{student}) = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.918 - 0 = 0.918 *$$

$$\text{Gain}(\text{income}) = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.918 - 0 = 0.918$$

สามารถเลือกได้ 2 อย่างเดียว



MAX FEATURE = 2

credit_rating = excellent

สูม 2 feature จากทั้งหมด 3 feature(age, income, student)
ได้ student และ income

credit_rating = excellent

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

สูม 2 ค่าผลรวม จาก age, income, student ของ income และ student

| 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 |

MAX FEATURE = 2

credit_rating = excellent

สุ่ม 2 feature จากทั้งหมด 3 feature(age, income, student)

ได้ student และ income

$$\begin{aligned}
 \text{Info}_{\text{student}}(D) &= \frac{3}{6} I(1,2) + \frac{3}{6} I(1,2) \\
 &= \frac{3}{6} \left[-\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) \right] + \frac{3}{6} \left[-\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) \right] \\
 &= 0.918
 \end{aligned}$$

| 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 |

MAX FEATURE = 2

credit_rating = excellent

สูม 2 feature จากทั้งหมด 3 feature(age, income, student)
ได้ student และ income

$$\begin{aligned}
 \text{Info}_{\text{income}}(D) &= \frac{1}{6} I(0,1) + \frac{3}{6} I(2,1) + \frac{2}{6} I(1,1) \\
 &= \frac{1}{6} \left[-\frac{0}{1} \log_2 \left(\frac{0}{1} \right) - \frac{1}{1} \log_2 \left(\frac{1}{1} \right) \right] + \frac{3}{6} \left[-\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) \right] \\
 &\quad + \frac{2}{6} \left[-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right] \\
 &= 0.459
 \end{aligned}$$

| 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 |

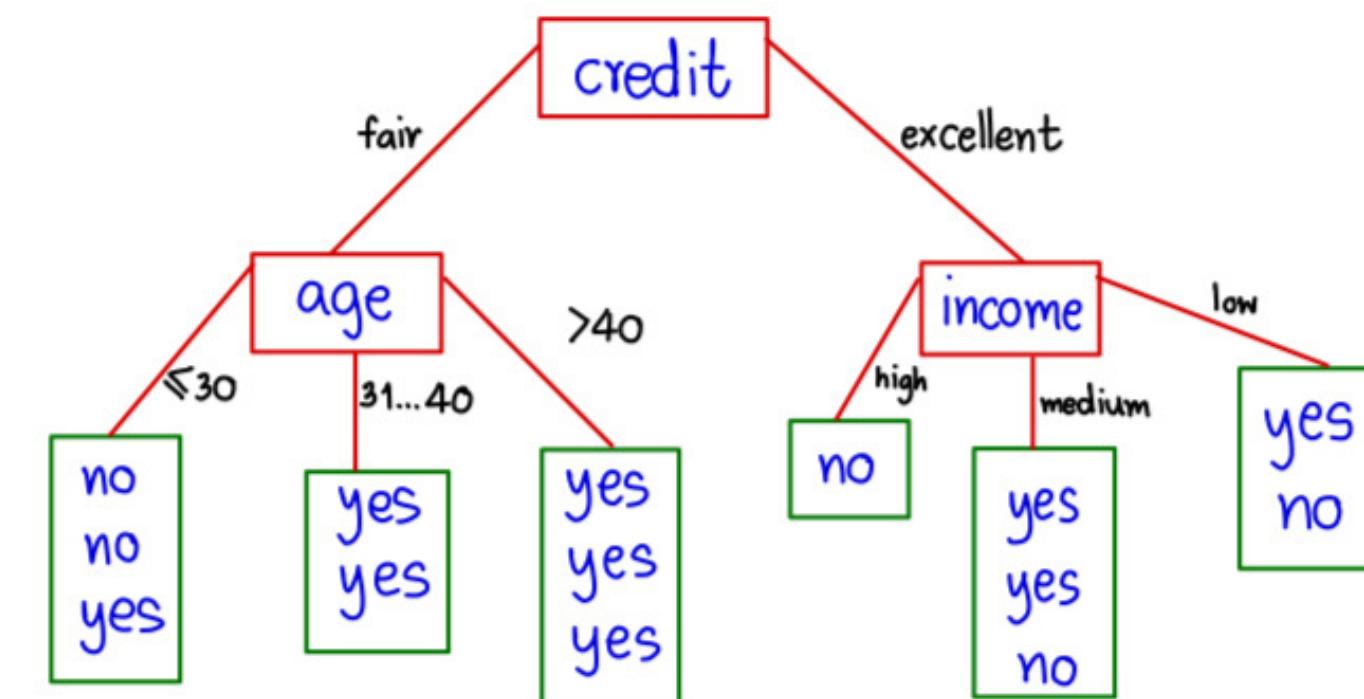
MAX FEATURE = 2

credit_rating = excellent

สูง 2 feature จากทั้งหมด 3 feature(age, income, student)
ได้ student และ income

$$\text{Gain (student)} = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 1 - 0.918 = 0.082$$

$$\text{Gain (income)} = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 1 - 0.459 = 0.241 *$$



MAX FEATURE = 2

credit_rating = excellent
Income = medium

สุ่ม 2 feature จากทั้งหมด 2 feature(age, student)
ได้ age และ student

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

$$= 0.918$$

| 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 |

MAX FEATURE = 2

credit_rating = excellent
Income = medium

สูม 2 feature จากทั้งหมด 2 feature(age, student)
ได้ age และ student

Info_{student}

$$\begin{aligned}
 I(D) &= \frac{1}{3} I(1,0) + \frac{2}{3} I(1,1) \\
 &= \frac{1}{3} \left[-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) - \frac{0}{1} \log_2 \left(\frac{0}{1} \right) \right] + \frac{2}{3} \left[-\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) \right] \\
 &= 0.6
 \end{aligned}$$

| 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 |

MAX FEATURE = 2

credit_rating = excellent
Income = medium

สุ่ม 2 feature จากทั้งหมด 2 feature(age, student)
ได้ age และ student

$$\begin{aligned}
 \text{Info}_{\text{age}}(D) &= \frac{1}{3} I(\leq 30) + \frac{1}{3} I(31 \dots 40) + \frac{1}{3} I(> 40) \\
 &= \frac{1}{3} \left[-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) - \frac{0}{1} \log_2 \left(\frac{0}{1} \right) \right] + \frac{1}{3} \left[-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) \frac{0}{1} \log_2 \left(\frac{0}{1} \right) \right] \\
 &\quad + \frac{1}{3} \left[-\frac{0}{1} \log_2 \left(\frac{0}{1} \right) - \frac{1}{1} \log_2 \left(\frac{1}{1} \right) \right] \\
 &= 0
 \end{aligned}$$

| 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 |

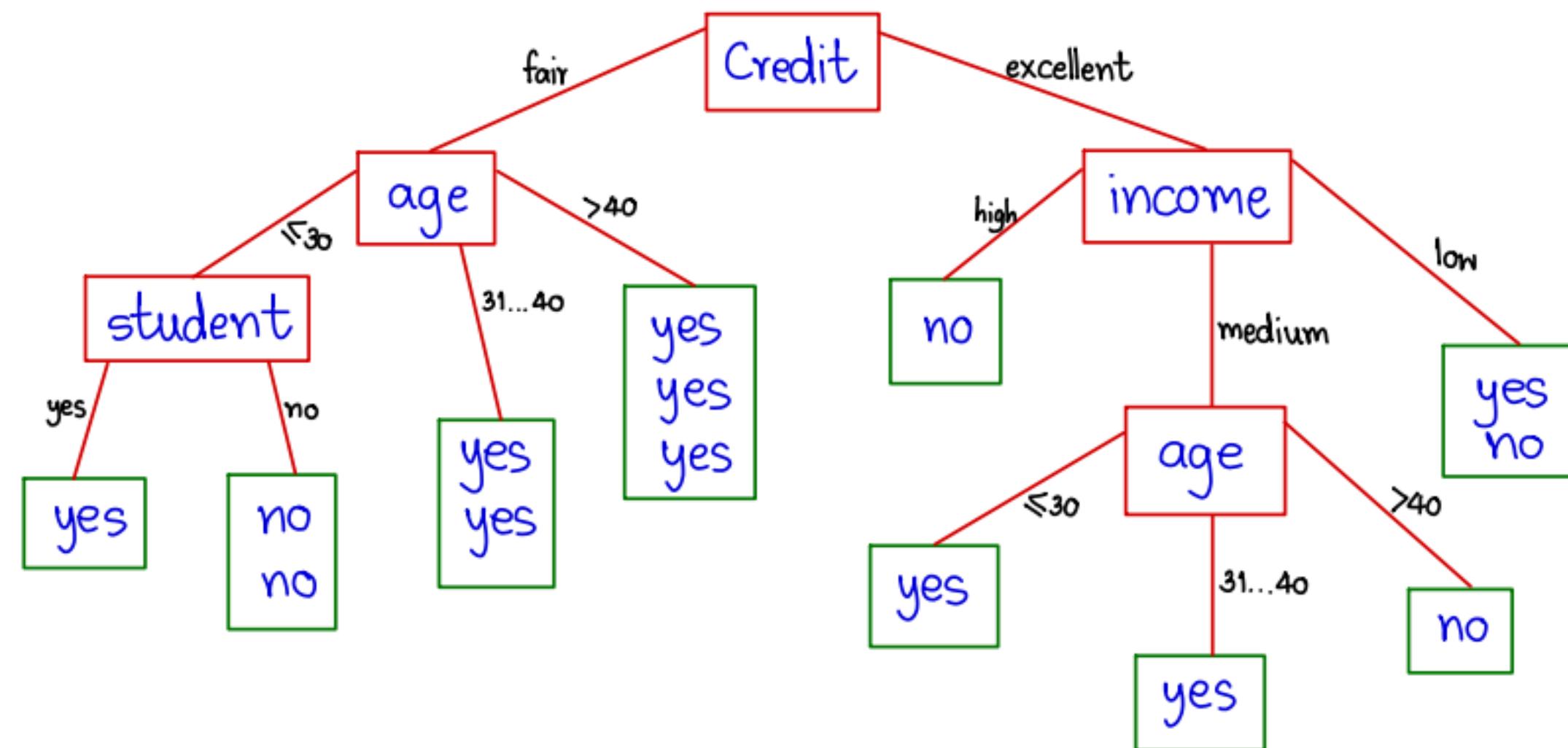
MAX FEATURE = 2

credit_rating = excellent
Income = medium

สุ่ม 2 feature จากทั้งหมด 2 feature(age, student)
ได้ age และ student

$$\text{Gain (student)} = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 0.918 - 0.6 = 0.318$$

$$\text{Gain (age)} = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.918 - 0 = 0.918 *$$



MAX FEATURE = 2

credit_rating = excellent
Income = low

สูบ 2 feature จากทั้งหมด 2 feature(student , age)
ได้ student และ age

กรณี Income = low

3.3 กรณี student , age แบบ

Credit = excellent , Income = low

$$\text{Info}(D) = I(1,1) = \left[-\frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] - \left[\frac{1}{2} \log_2\left(\frac{1}{2}\right) \right] = 1$$

| 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 |

MAX FEATURE = 2

credit_rating = excellent
Income = low

សំរាប់ 2 feature ទាំងអស់ 2 feature (student , age)
តើ student និង age

3.3 ទីនេះ student , age ឬណែនាំ

Credit = excellent , Income = low

$$\begin{aligned}
 \text{Info}_{\text{age}}(D) &= \frac{1}{2} I(1,0) + \frac{1}{2} I(0,1) \\
 &= \frac{1}{2} \left[-\frac{1}{1} \log_2 \left(\frac{1}{1} \right) - \frac{0}{1} \log_2 \left(\frac{0}{1} \right) \right] + \frac{1}{2} \left[-\frac{0}{1} \log_2 \left(\frac{0}{1} \right) - \frac{1}{1} \log_2 \left(\frac{1}{1} \right) \right] \\
 &= 0
 \end{aligned}$$

| 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 |

MAX FEATURE = 2

credit_rating = excellent
Income = low

สูบ 2 feature จากทั้งหมด 2 feature (student , age)
ได้ student และ age

3.3 ไฝ่ student , age แบบ

Credit = excellent , Income = low

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

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

$$= 1$$

| 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 |

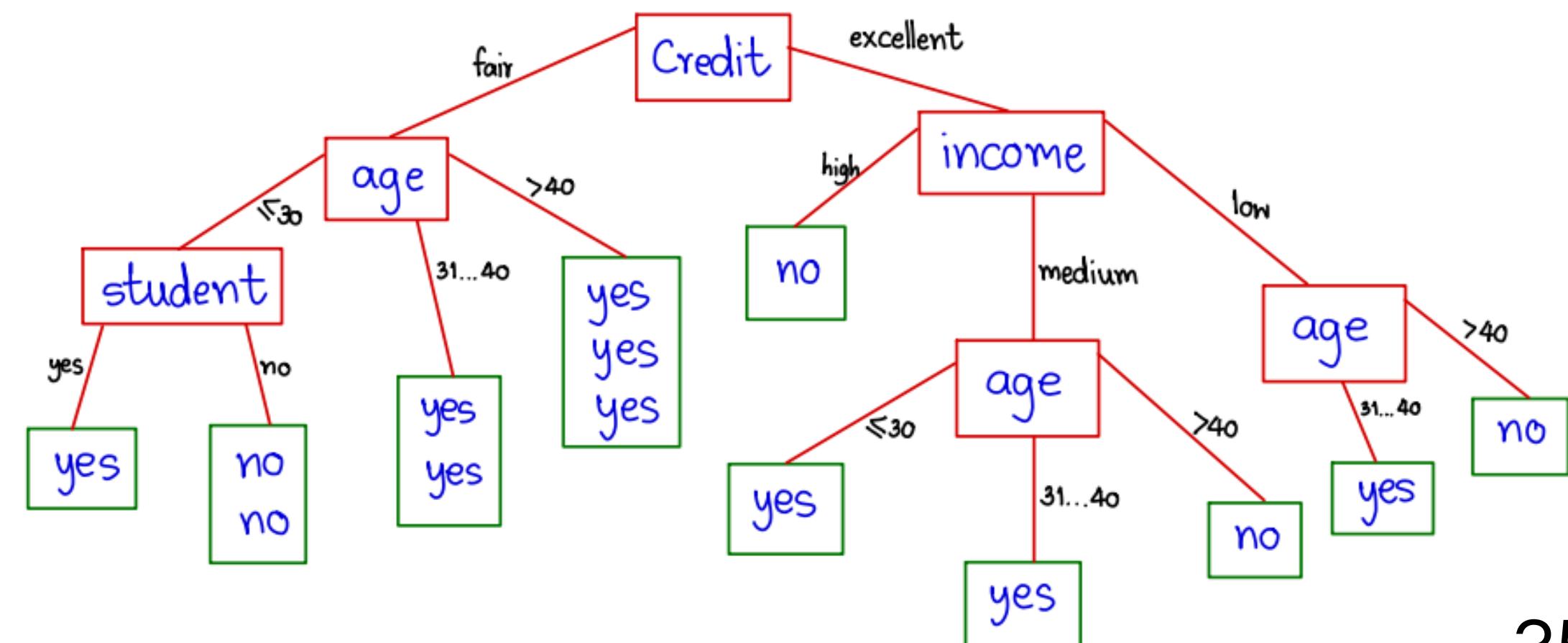
MAX FEATURE = 2

credit_rating = excellent
Income = low

สุ่ม 2 feature จากทั้งหมด 2 feature(student , age)
ได้ student และ age

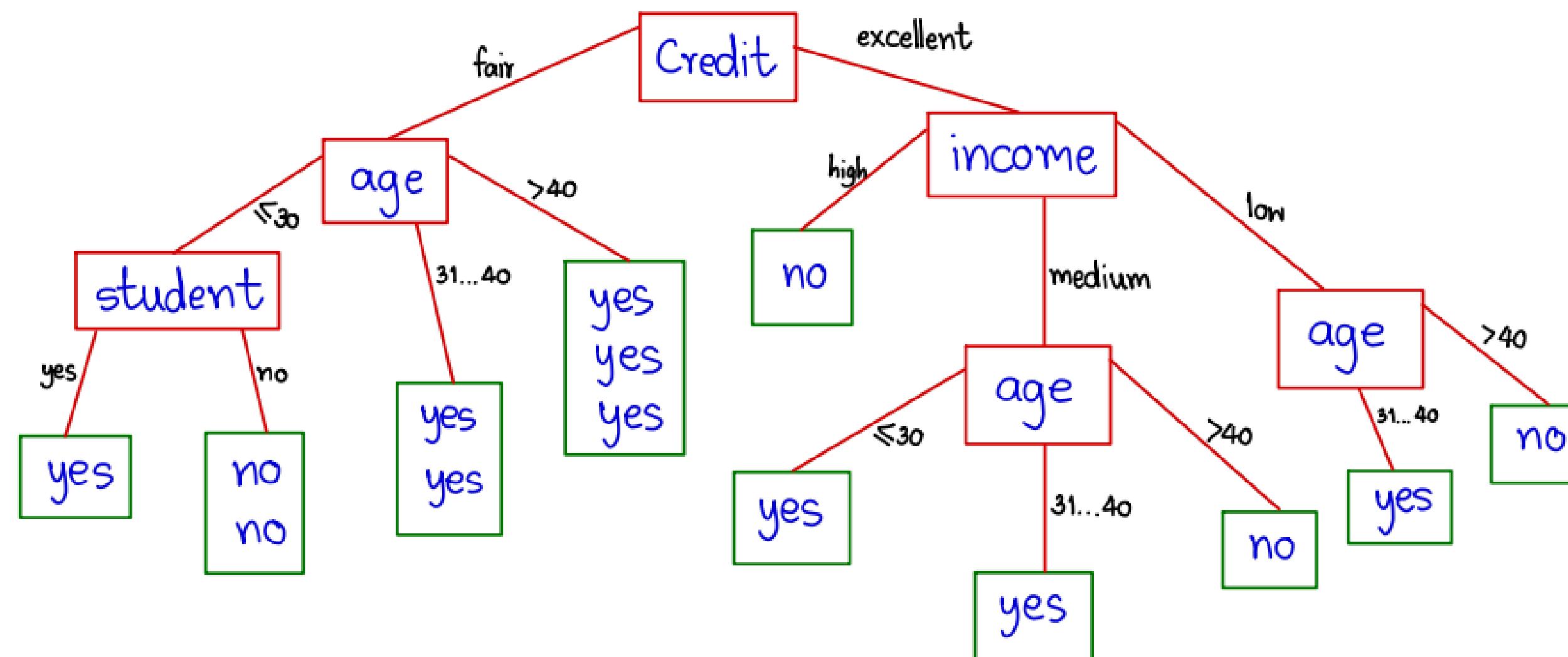
$$\text{Gain (age)} = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 1 - 0 = 1 *$$

$$\text{Gain (student)} = \text{Info}(D) - \text{Info}_{\text{student}}(D) = 1 - 1 = 0$$



MAX FEATURE = 2

credit_rating = excellent
Income = low



Thank You