

age	income	student	credit_rating	buys_computer
$\leq 30$	high	no	fair	no
$\leq 30$	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
$\leq 30$	medium	no	fair	no
$\leq 30$	low	yes	fair	yes
$> 40$	medium	yes	fair	yes
$\leq 30$	medium	yes	excellent	yes
31..40	med.	no	excellent	yes
31..40	high	yes	fair	yes
$> 40$	medium	no	excellent	no

Class P: buys\_computer = "yes"

Class N: buys\_computer = "no"

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

age	P <sub>i</sub>	n <sub>i</sub>	I(p <sub>i</sub> , n <sub>i</sub> )
$\leq 30$	2	3	0.991
31..40	4	0	0
$> 40$	3	2	0.991

$$\text{Info}_{\text{age}}(D) = \frac{5}{14} I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2) = 0.694$$

$$\text{Gain}(\text{age}) = \text{Info}(D) - \text{Info}_{\text{age}}(D) = 0.246$$

income	$p_i$	$n_i$	$I(p_i, n_i)$
high	2	2	1
medium	4	2	0.918
low	3	1	0.811

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

$$= 0.286 + 0.593 + 0.252$$

$$= 0.911$$

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

student	$p_i$	$n_i$	$I(p_i, n_i)$
yes	6	1	0.592
no	3	4	0.985

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

$$= 0.296 + 0.493$$

$$= 0.789$$

$$\text{Gain}(\text{student}) = 0.940 - 0.789$$

$$= 0.151$$

Credit rating	$p_i$	$n_i$	$I(p_i, n_i)$
fair	6	2	0.811
excellent	3	3	1

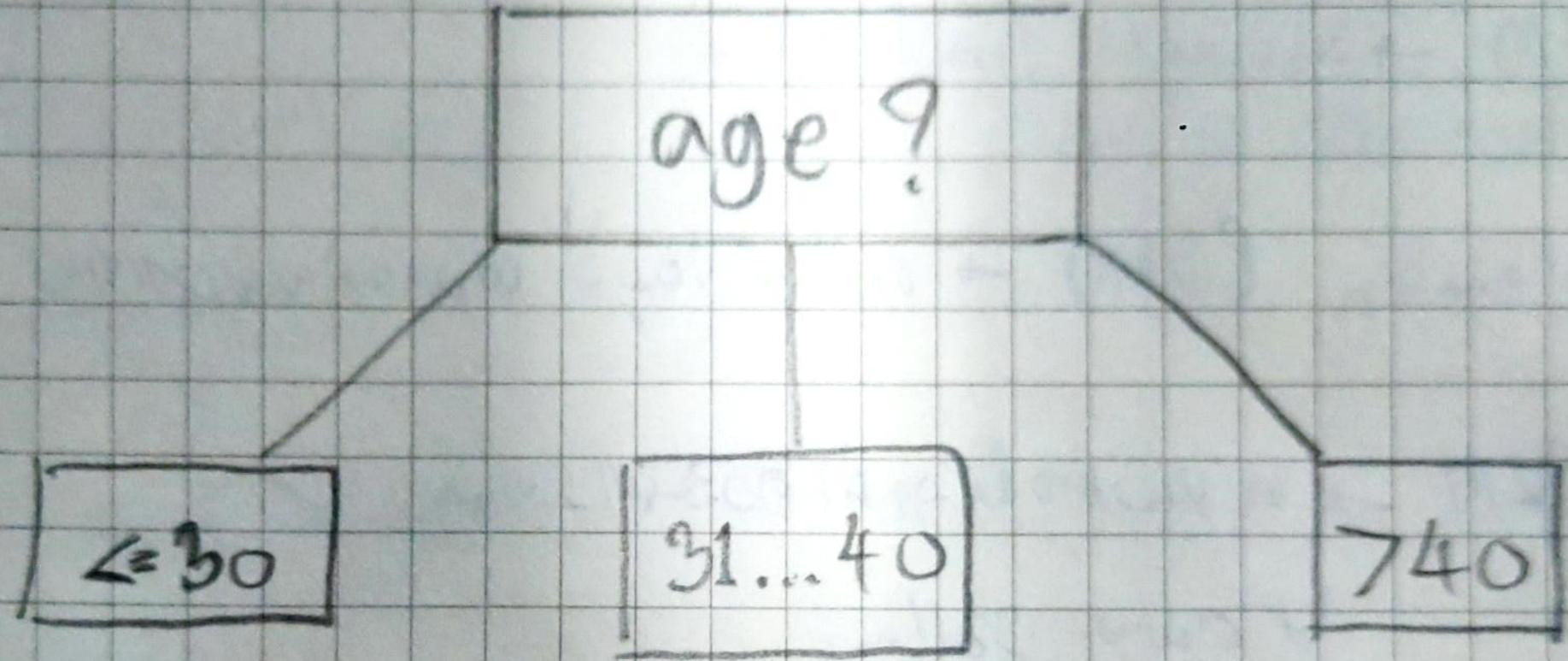
$$\text{Info}_{\text{credit}}(D) = \frac{8}{14} I(6, 2) + \frac{6}{14} I(3, 3)$$

$$= 0.463 + 0.429$$

$$= 0.892$$

$$\text{Gain}(\text{credit}) = 0.940 - 0.892 =$$

$$= 0.048$$



Age <= Gain compared Age w/ Root node (accuracy = 72% to 91%)

Age <= 30

age	income	student	credit-rating	buy-computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
<=30	medium	yes	excellent	yes

$$\text{Info}_{\text{age} \leq 30} (D) = I(2, 3) = -\frac{2}{5} \log_2 \left(\frac{2}{5}\right) - \frac{3}{5} \log_2 \left(\frac{3}{5}\right) = 0.971$$

Info age <= 30 w.r.t income

$$\begin{aligned} \text{Info}_{\text{age} \leq 30, \text{income}} (D) &= \frac{2}{5} I(0, 2) + \frac{2}{5} I(1, 1) + \frac{1}{5} I(1, 0) \\ &= \frac{2}{5} (0) + \frac{2}{5} (-1) + \frac{1}{5} (0) \\ &= 0.4 \end{aligned}$$

Info student

$$\begin{aligned} \text{Info}_{\text{age} \leq 30, \text{student}} (D) &= \frac{2}{5} I(2, 0) + \frac{3}{5} I(0, 3) \\ &= \frac{2}{5} (0) + \frac{3}{5} (0) \\ &= 0 \end{aligned}$$

Info Credit-rating

$$\begin{aligned} \text{Info}_{\text{age} \leq 30, \text{credit-rating}} (D) &= \frac{2}{5} I(1, 2) + \frac{2}{5} I(1, 1) \\ &= \frac{2}{5} (0.918) + \frac{2}{5} (1) \\ &= 0.551 + 0.4 \\ &= 0.951 \end{aligned}$$

$$\text{Gain}_{\text{age} \leq 30, \text{income}} (D) = 0.971 - 0.4 = 0.571$$

$$\text{Gain}_{\text{age} \leq 30, \text{student}} (D) = 0.971 - 0 = 0.971$$

$$\text{Gain}_{\text{age} \leq 30, \text{credit-rating}} (D) = 0.971 - 0.951 = 0.02$$

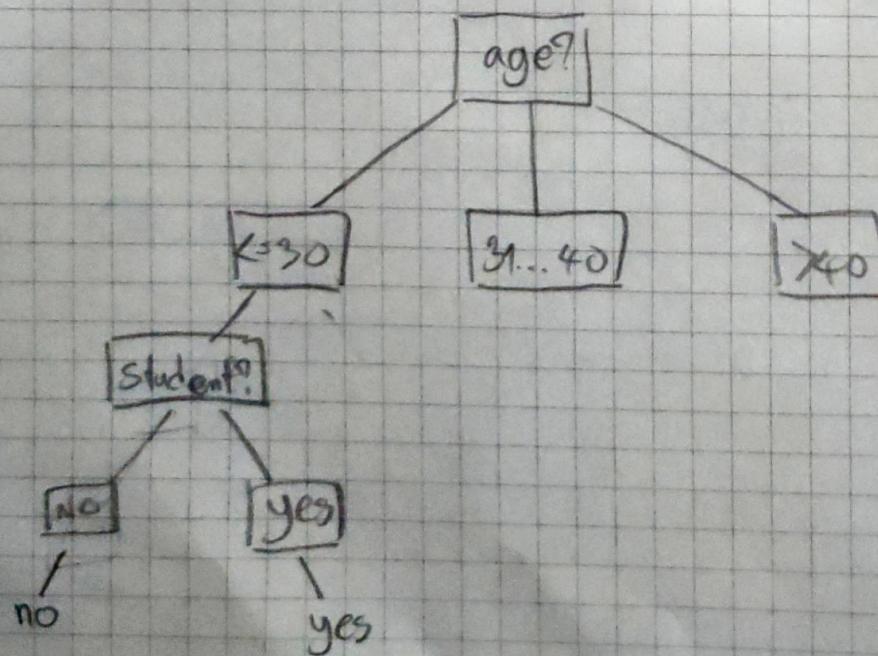
ஒவ்வொரு student க்கு Decision node யின் age  $\leq 30$   
 மற்றும் கூடுதல் Student க்கு மாற்றும் age  $< 30$  என  
 and "student" buys computer  
 Student = "yes"

age	student	buys computer
$\leq 30$	yes	yes
$> 30$	yes	yes

Student = "no"

age	student	buys computer
$\leq 30$	no	no
$> 30$	no	no
$> 30$	no	no

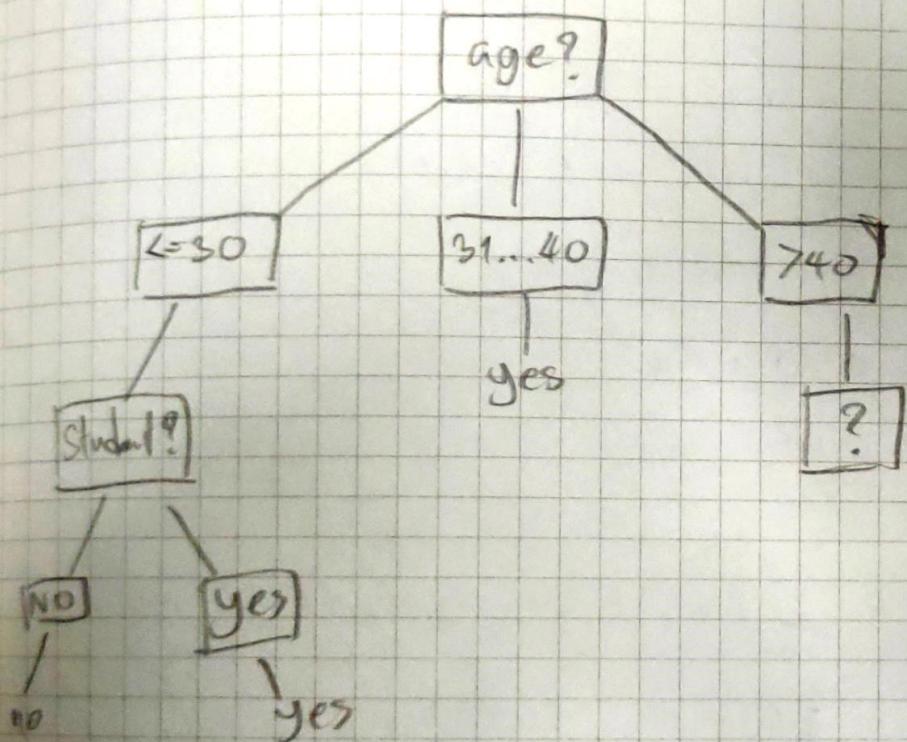
ஒவ்வொரு student க்கு yes நிலையில் buys-computer க்கு yes  
 மற்றும் கூடுதல் student க்கு no நிலையில் buys-computer க்கு no என  
 பின்னர் Decision node யின் Student



age : 31...40

age	income	Student	Credit-rating	buys-computer
31...40	high	no	fair	yes
31...40	low	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes

min. 1st new age : 31...40 & buys-computer 6da yes raus  
für dt. Decision tree Induktions



now age > 40

age	income	student	Credit-rating	buys comp.
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
>40	medium	yes	fair	yes
>40	medium	no	excellent	no

$$\text{Info}_{\text{age} > 40}(D) = I(3,2) = -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) = 0.911$$

income

$$\begin{aligned} \text{Info}_{\text{age} > 40, \text{income}}(D) &= \frac{3}{5} I(2,1) + \frac{2}{5} I(1,1) \\ &= \frac{3}{5} (0.918) + \frac{2}{5} (0) \\ &= 0.551 + 0.4 \\ &= 0.951 \end{aligned}$$

student

$$\begin{aligned} \text{Info}_{\text{age} > 40, \text{student}}(D) &= \frac{2}{5} I(1,1) + \frac{3}{5} I(2,1) \\ &= \frac{2}{5} (1) + \frac{3}{5} (0.918) \\ &= 0.4 + 0.551 \\ &= 0.951 \end{aligned}$$

Credit-rating

$$\begin{aligned} \text{Info}_{\text{age} > 40, \text{credit-rating}}(D) &= \frac{3}{5} (3,0) + \frac{2}{5} (0,2) \\ &= \frac{3}{5} (0) + \frac{2}{5} (0) \\ &= 0 \end{aligned}$$

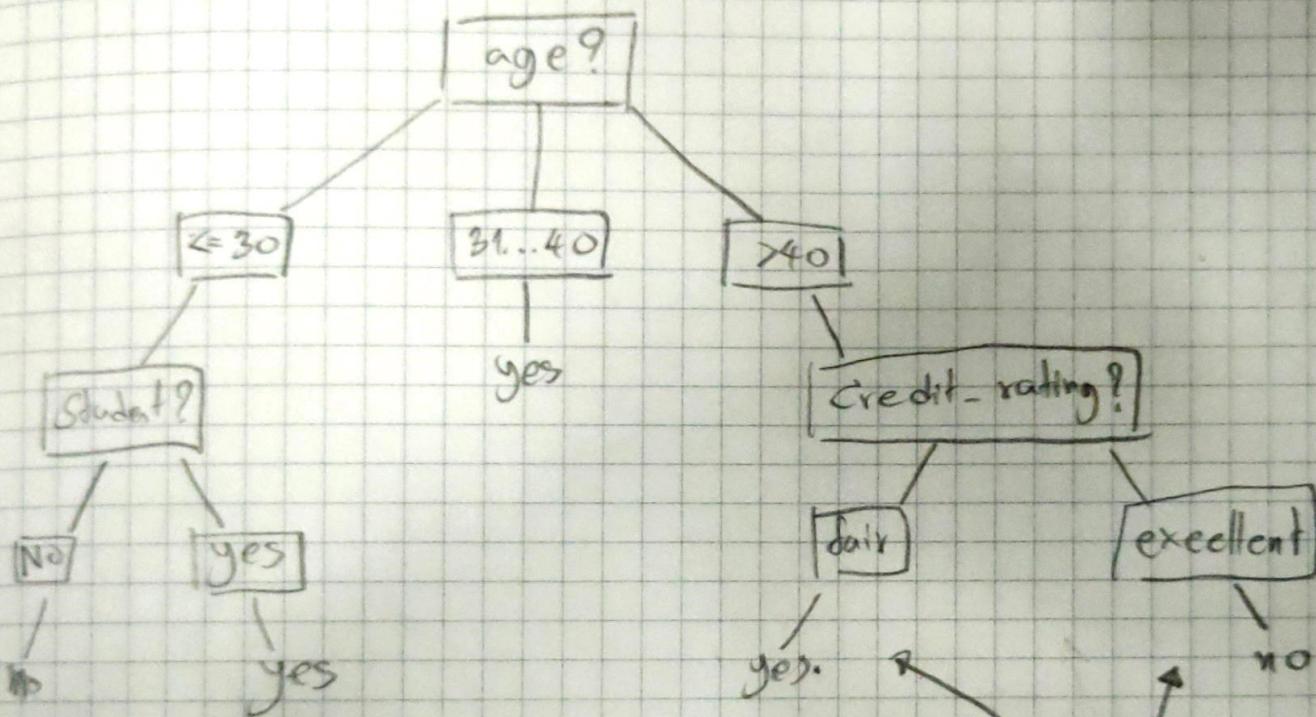
d) Gain

$$\text{Gain}_{\text{age} > 40, \text{income}}(D) = 0.971 - 0.951 = 0.02$$

$$\text{Gain}_{\text{age} > 40, \text{student}}(D) = 0.971 - 0.951 = 0.02$$

$$\text{Gain}_{\text{age} > 40, \text{credit\_rating}}(D) = 0.971 - 0 = 0.971 \quad \checkmark$$

Wurzel credit-rating kategorien zu Node and vorne age > 40



Wurzel wenn age > 40 und Credit-rating wertenswert

credit-rating = "fair"

age	credit-rating	buys.computer
≥ 40	fair	yes
≥ 40	fair	yes
≥ 40	fair	yes

credit-rating = "excellent"

age	credit-rating	buys.computer
≥ 40	excellent	no
≥ 40	excellent	no

Wurzel mit Credit-rating wertenswert kaufen buys-computer möglich