

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

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

age	<=30	31-40	>40
yes	2	4	3
no	3	0	2

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

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

$$\text{Gain}(\text{age}) = 0.940 - 0.694 = 0.246$$

income	low	medium	high
yes	3	4	2
no	1	2	2

$$\text{Info}_{\text{income}}(D) = \frac{4}{14} I(3,1) + \frac{6}{14} I(4,2) + \frac{4}{14} I(2,2) = 0.911$$

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

student	yes	no
yes	6	3
no	1	4

$$\text{Info}_{\text{student}}(D) = \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4) = 0.7845$$

$$\text{gain}(\text{student}) = 0.940 - 0.784 = 0.151$$

credit_rating	fair	excellent
yes	6	3
no	2	3

$$\text{Info}_{\text{credit}}(D) = \frac{8}{14} I(6,2) + \frac{6}{14} I(3,3) = 0.842$$

$$\text{gain}(\text{credit}) = 0.940 - 0.842 = 0.098$$

Root node

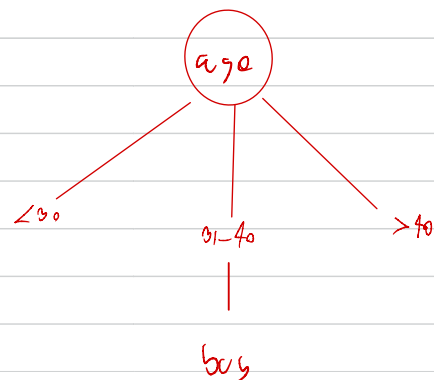
$$\text{gain}(\text{age}) = 0.246$$

$$\text{gain}(\text{income}) = 0.029$$

$$\text{gain}(\text{student}) = 0.151$$

$$\text{gain}(\text{credit}) = 0.098$$

$\therefore \text{age} > \text{student} > \text{credit rating} > \text{income}$



Wirsung student > credit > income

age	student	buys_computer
<=30	no	no
<=30	no	no
31...40	no	yes
>40	no	yes
>40	yes	yes
>40	yes	no
31...40	yes	yes
<=30	no	no
<=30	yes	yes
>40	yes	yes
<=30	yes	yes
31...40	no	yes
31...40	yes	yes
>40	no	no

$$Info(D) = 5(5,1) = 1$$

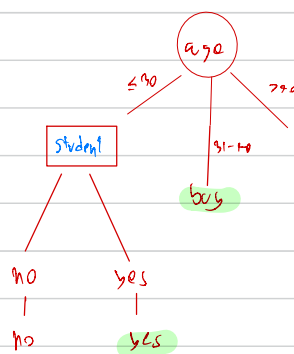
$$Info_{\leq 30}(D) = \frac{2}{5} (2,0) + \frac{3}{5} (0,3) = 0$$

$$Info_{>40}(D) = \frac{2}{5} (2,1) + \frac{2}{5} (1,1) = 0.951$$

$$gain(\leq 30) = 1 - 0 = 1$$

$$gain(>40) = 1 - 0.951 = 0.049$$

7.1.1



Wirsung credit > income

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

$$Info(D) = J(9,2) = -\frac{2}{9} \log_2 \frac{2}{9} - \frac{7}{9} \log_2 \frac{7}{9} = 0.970$$

$$Info_{credit}(D) = \frac{3}{9} J(3,0) + \frac{2}{9} J(0,2) = 0$$

$$Info_{income}(D) = \frac{2}{9} J(2,1) + \frac{7}{9} J(1,1) = 0.951$$

$$gain(credit) = 0.970 - 0 = 0.970$$

$$gain(income) = 0.970 - 0.951 = 0.019$$

7.1.2

Resulting tree:

