

Exercise 11

Entropy (buy computer)

	Yes	No
Buy	12	8

$$\begin{aligned}\mathcal{E}(\text{Buy}) &= \mathcal{E}(0,6,0,4) \\ &= -(0,6 \log_2 0,6) - (0,4 \log_2 0,4) \\ &= 0,97\end{aligned}$$

Entropy (Buy, Age)

	Yes	No
≤ 30	2	6
Age $31 \dots 40$	6	0
> 40	4	2

Data cleaning bei R11
(< 30 zu ≤ 30)

$$\begin{aligned}\mathcal{E}(\leq 30) &= \mathcal{E}(2,6) \\ &= 0,858\end{aligned}$$

$$\begin{aligned}\mathcal{E}(31 \dots 40) &= \mathcal{E}(6,0) \\ &= 0,521\end{aligned}$$

$$\begin{aligned}\mathcal{E}(> 40) &= \mathcal{E}(4,2) \\ &= 0,796\end{aligned}$$

$$\begin{aligned}\mathcal{E}(\text{buy}, \text{age}) &= P(8/20) \cdot 0,858 + P(6/20) \cdot 0,521 + P(4/20) \cdot 0,796 \\ &= 0,4 \cdot 0,858 + 0,3 \cdot 0,521 + 0,3 \cdot 0,796 \\ &= 0,736\end{aligned}$$

$$\begin{aligned}\text{gain}(\text{Buy}, \text{Age}) &= \mathcal{E}(\text{Buy}) - \mathcal{E}(\text{Buy}, \text{age}) \\ &= 0,234\end{aligned}$$

Entropy (Buy , Income)

Buy		Yes	No	
Income	High	3	2	5
	Med	5	3	8
	Low	4	3	7

$$\begin{aligned}
 E(\text{Buy}, \text{income}) &= P(5/20) \cdot E(3,2) + P(8/20) \cdot E(5,3) + P(7/20) \cdot E(4,3) \\
 &= 0,25 \cdot E(3,2) + 0,4 \cdot E(5,3) + 0,35 \cdot E(4,3) \\
 &= 0,25 \cdot 0,942 + 0,4 \cdot 0,953 + 0,35 \cdot 0,985 \\
 &= 0,235 + 0,381 + 0,344 \\
 &= 0,96
 \end{aligned}$$

$$\begin{aligned}
 E(\text{High}) &= E(3,2) \\
 &= E(0,6, 0,5) \\
 &= -(0,6 \log_2 0,6) - (0,5 \log_2 0,5) \\
 &= 0,442 + 0,5 \\
 &= 0,942
 \end{aligned}$$

$$\begin{aligned}
 E(\text{Med}) &= E(5,3) \\
 &= E(0,625, 0,375) \\
 &= 0,423 + 0,530 \\
 &= 0,953
 \end{aligned}$$

$$\begin{aligned}
 E(\text{low}) &= E(4,3) \\
 &= E(0,571, 0,428) \\
 &= 0,461 + 0,524 \\
 &= 0,985
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain}(\text{Buy}, \text{income}) &= E(\text{Buy}) - E(\text{Buy}, \text{income}) \\
 &= 0,97 - 0,96 \\
 &= \underline{\underline{0,01}}
 \end{aligned}$$

Entropy(Buy, Student)

	Yes	No	
Student	No	4	7
	Yes	8	1
		11	9

$$\begin{aligned}
 \mathcal{E}(\text{No}) &= (4, 7) \\
 &= (0,363, 0,636) \\
 &= 0,526 + 0,415 \\
 &= 0,941
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{E}(\text{Yes}) &= (8, 1) \\
 &= (0,888, 0,111) \\
 &= \cancel{0,752} 0,152 + 0,352 \\
 &= 0,504
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy(Buy, student)} &= P(11/20) \cdot 0,941 + \\
 &\quad P(9/20) \cdot 0,54 \\
 &= 0,55 \cdot 0,941 + 0,45 \cdot 0,54 \\
 &= 0,517 + 0,243 \\
 &= 0,76
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain(Buy, student)} &= \mathcal{E}(\text{buy}) - \mathcal{E}(\text{Buy, student}) \\
 &= 0,97 - 0,76 \\
 &= \underline{\underline{0,21}}
 \end{aligned}$$

Entropy(Buy, credit)

	Yes	No
credit	fair	7 3
	Excellent	5 5

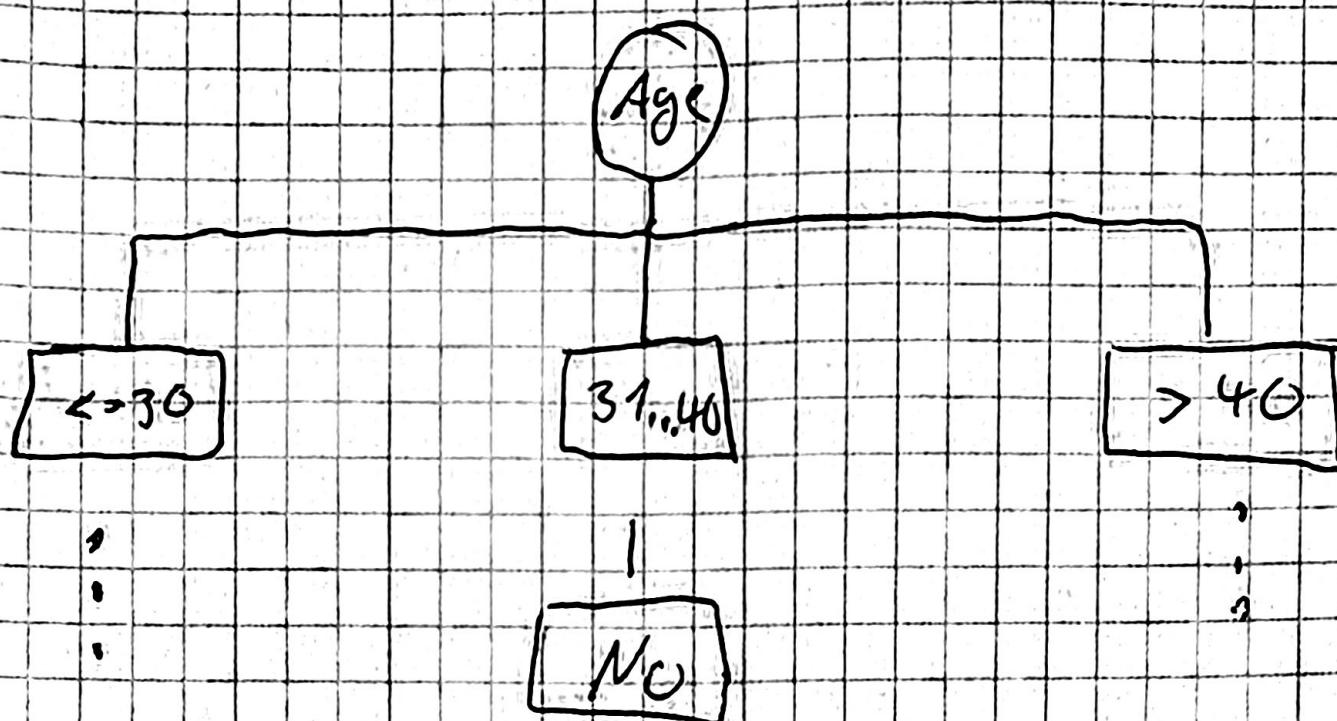
$$\begin{aligned}
 \mathcal{E}(\text{fair}) &= (7, 3) \\
 &= (0,7, 0,3) \\
 &= 0,360 + 0,521 \\
 &= 0,881
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{E}(\text{excellent}) &= (5, 5) \\
 &= (0,5, 0,5) \\
 &= 0,5 + 0,5 \\
 &= 1
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{E}(\text{Buy, credit}) &= P(10/20) \cdot 0,881 + P(10/20) \cdot \cancel{0,88}^1 \\
 &= 0,5 \cdot 0,881 + 0,5 \cdot 1 \\
 &= 0,4405 + 0,5 \\
 &= 0,9405
 \end{aligned}$$

$$\begin{aligned}
 \text{Gain}(\text{Buy, credit}) &= \mathcal{E}(\text{Buy}) - \mathcal{E}(\text{Buy, credit}) \\
 &= 0,97 - 0,94 \\
 &= \underline{\underline{0,03}}
 \end{aligned}$$

Es wird das mit dem Höchsten gain gewählt



WebGraphviz is [Graphviz](#) in the Browser

Enter your graphviz data into the Text Area:

(Your Graphviz data is private and never harvested)

[Sample 1](#) [Sample 2](#) [Sample 3](#) [Sample 4](#) [Sample 5](#)

```
digraph G {  
    "Age" -> "<=30"  
    "Age" -> "31...40"  
    "Age" -> ">40"  
    "31...40" -> "No"  
    "<=30" -> "..."  
    ">40" -> "...."  
}
```

Generate Graph!

