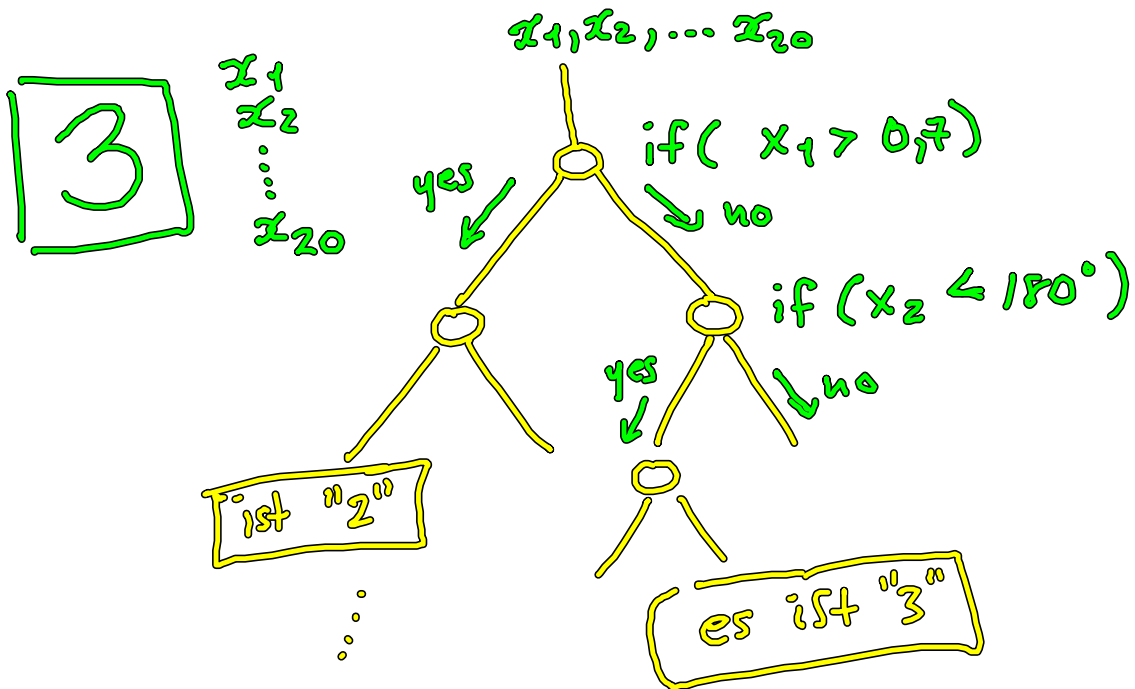
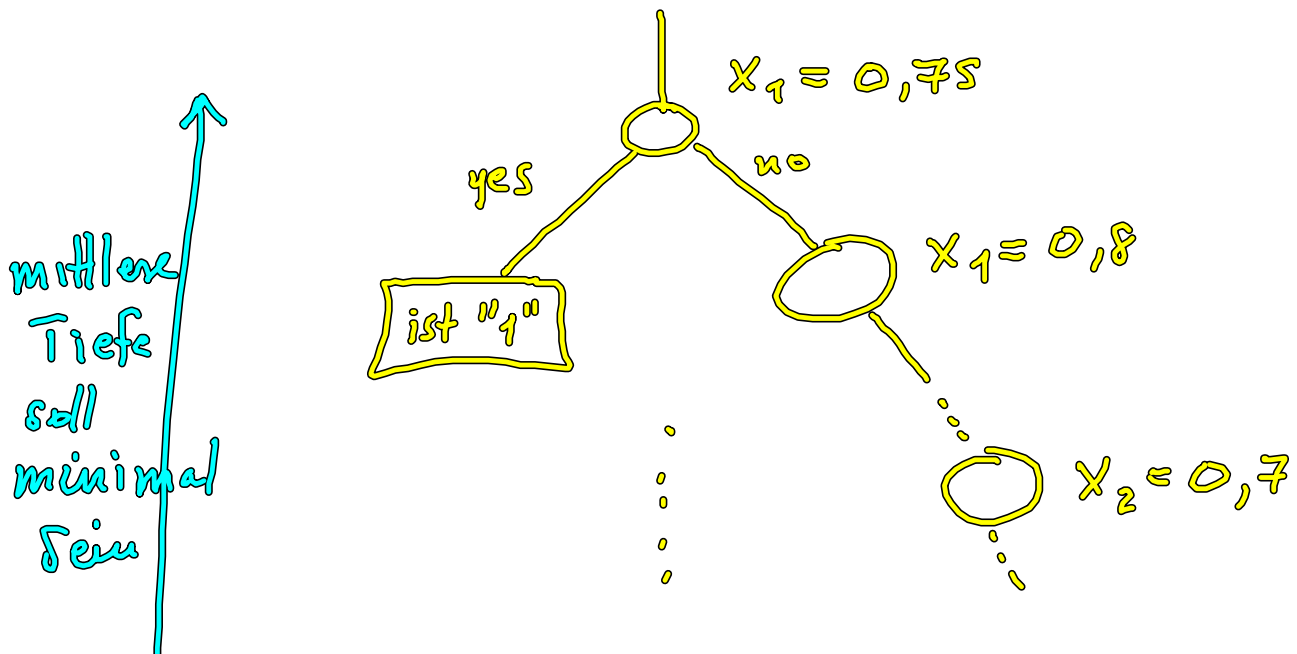


Backprop.

ID3 - Induction Trees



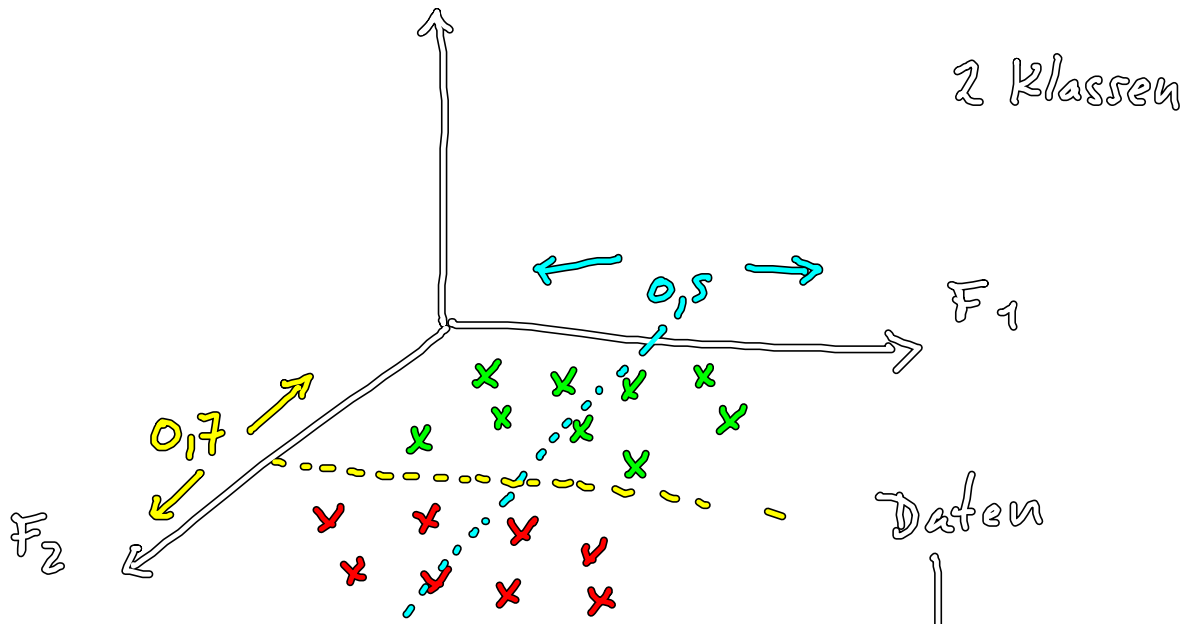
50,000 Zahlen

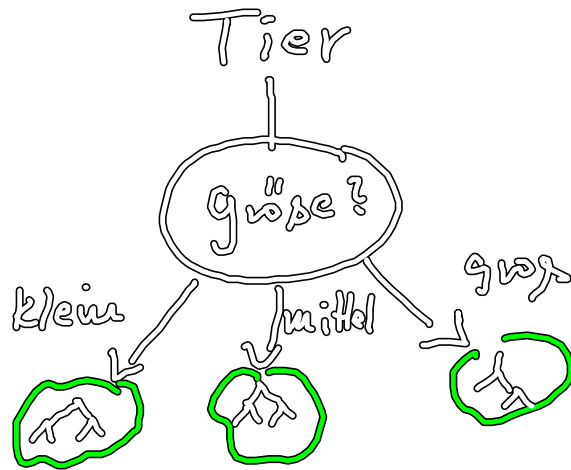
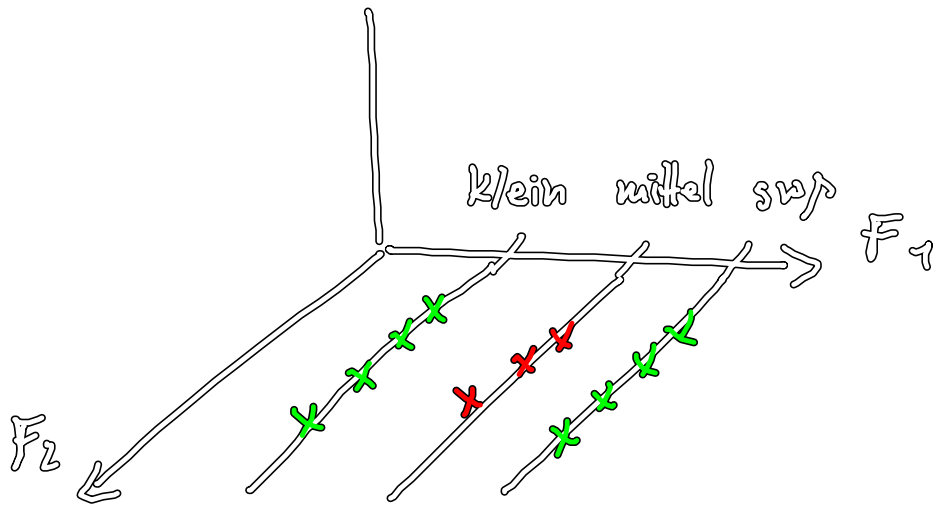
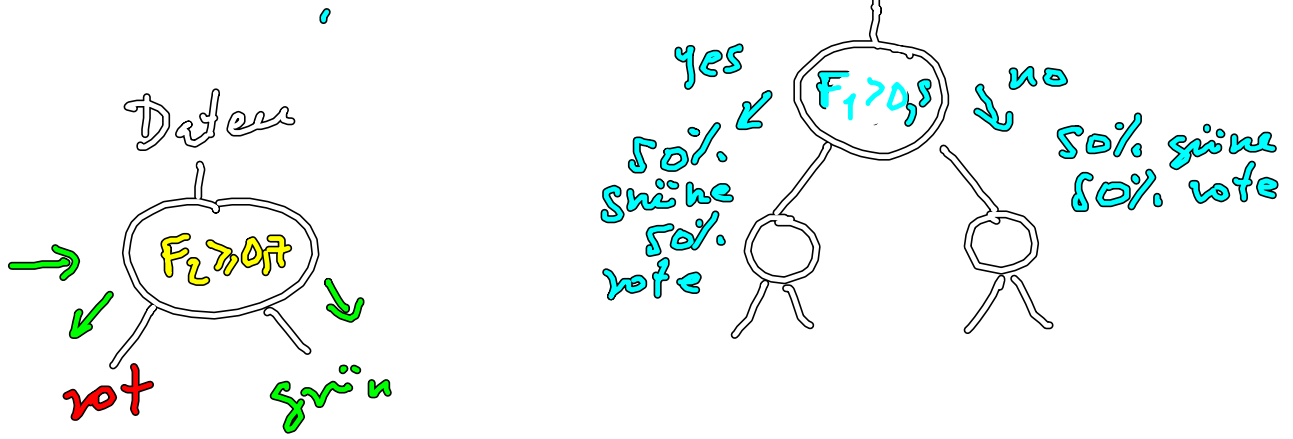




"1" "2" "5" ... 50,000 Blätter

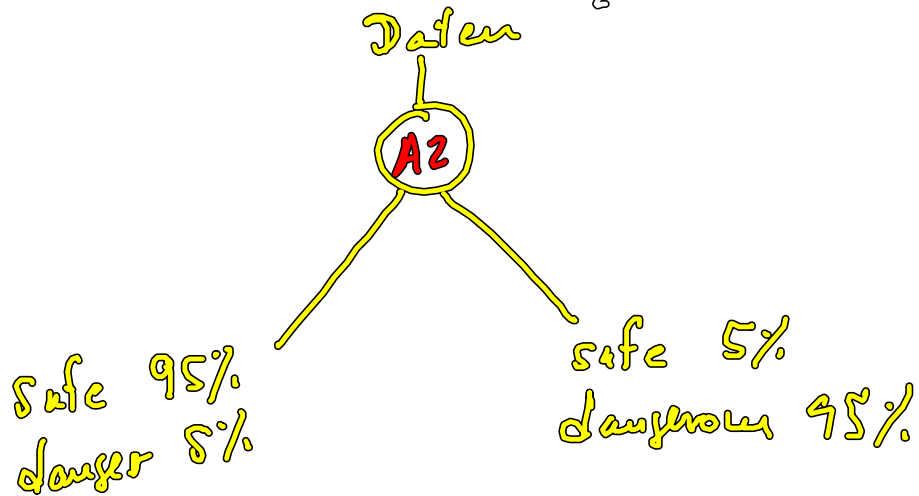
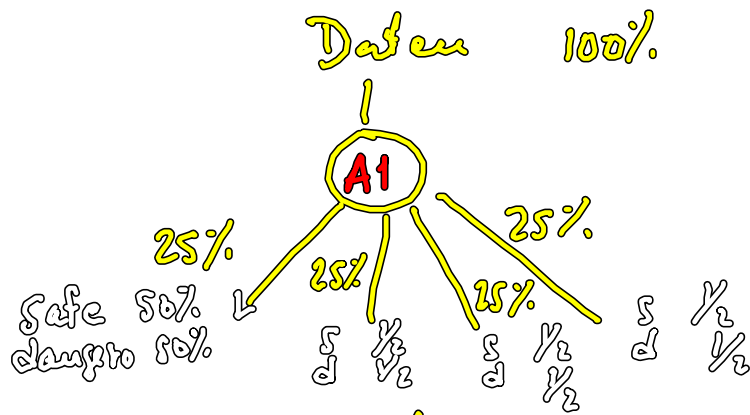
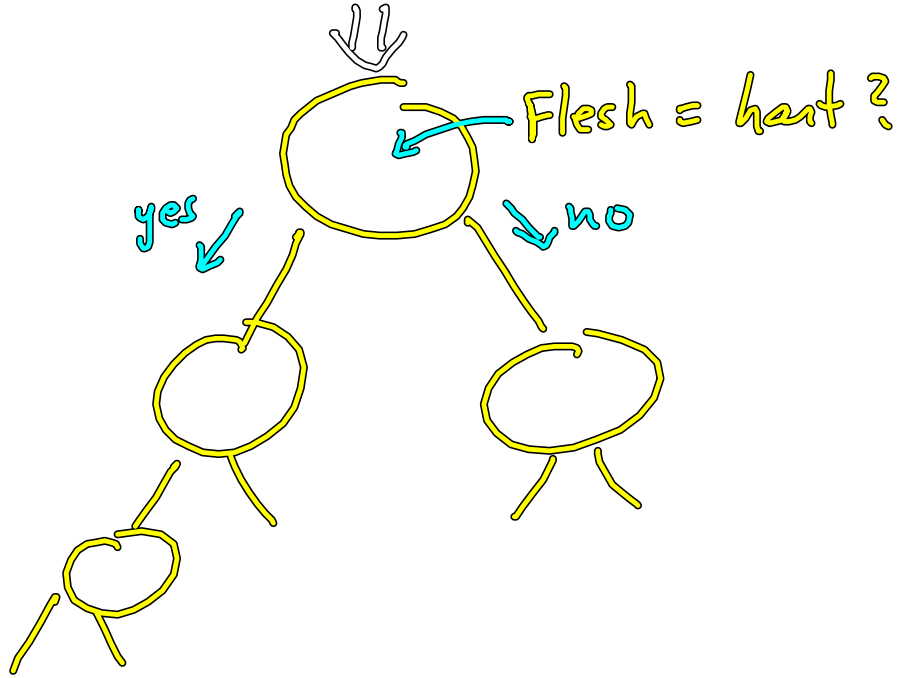
	X_1	X_2	X_3	...	X_{20}
"1"	0,75	0,8	0,1	...	
"5"	0,6	0,5			
		⋮			
		⋮			



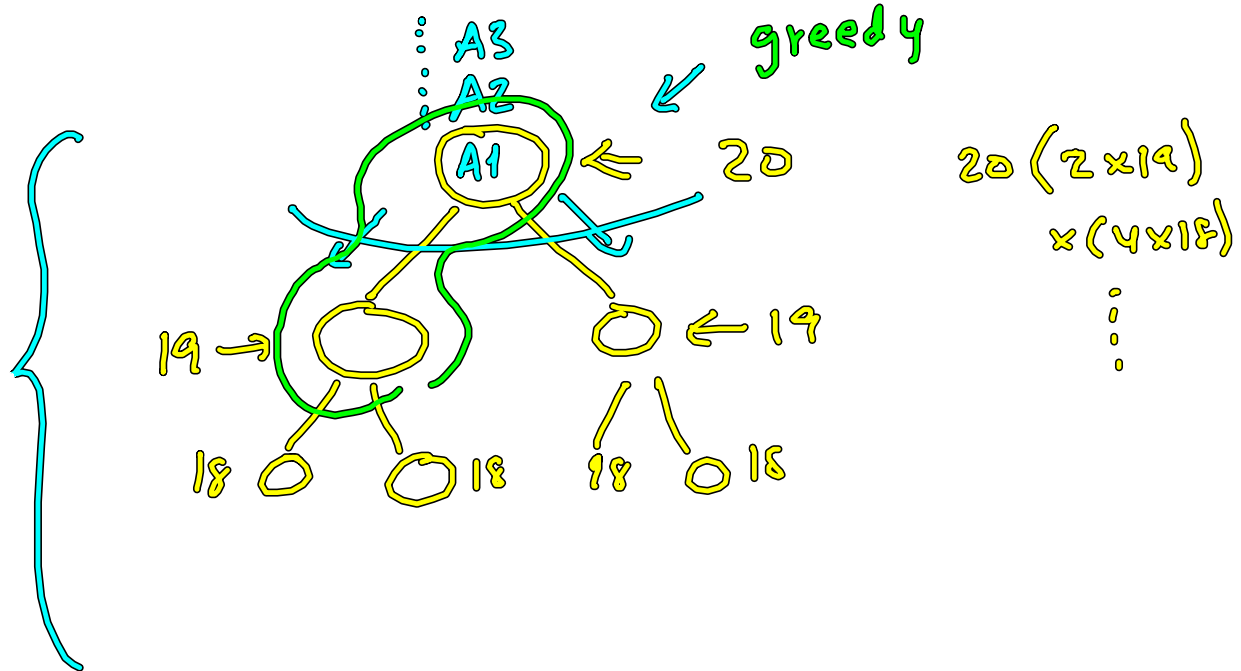


Unbekanntes Wesen

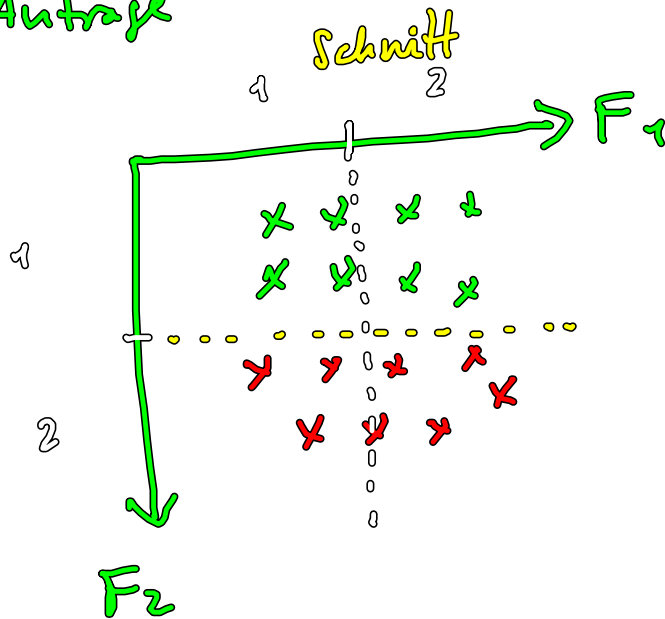
skin, color, size, flesh



20 Merkmale



Relevanz der Anfrage
Messung der "Güte" der Verteilung
nach Anfrage



X X X	⋮	xxv	100% grün
Y X Y	⋮	xxv	
X X X	⋮	---	0% grün
X X Y	⋮	xxxv	100% rot

{ 50% grün
50% rot

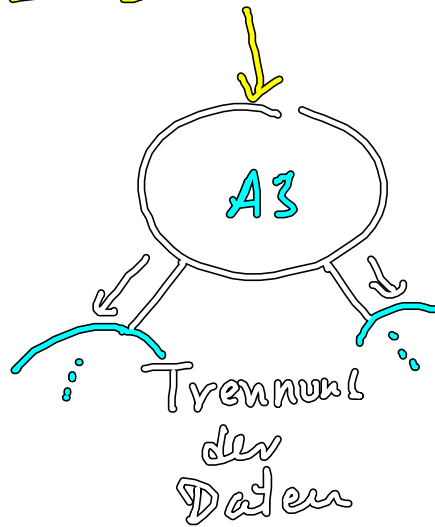
o o o
o o o

ID3

Entropie = $-\sum_{i=1}^n p(c_i/a_j) \log(p(c_i/a_j))$

Informationsgewinn?

ID3



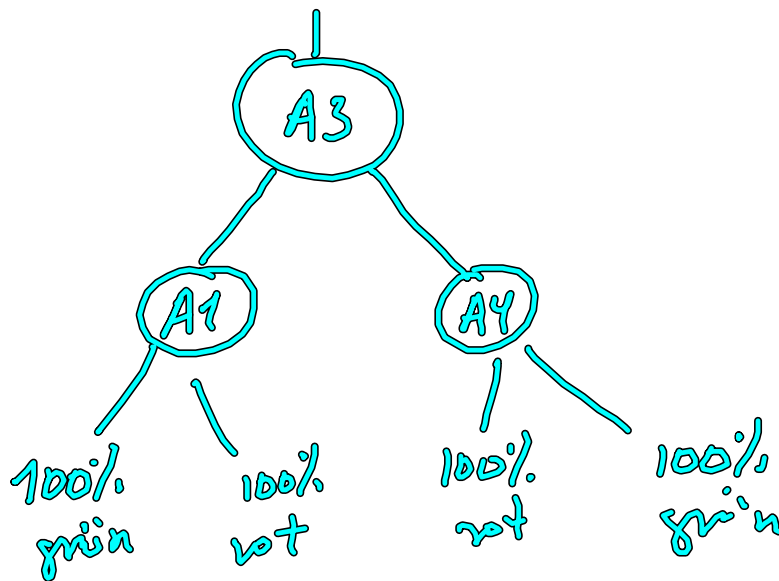
A1: Entropie 1

A2: Entropie 2

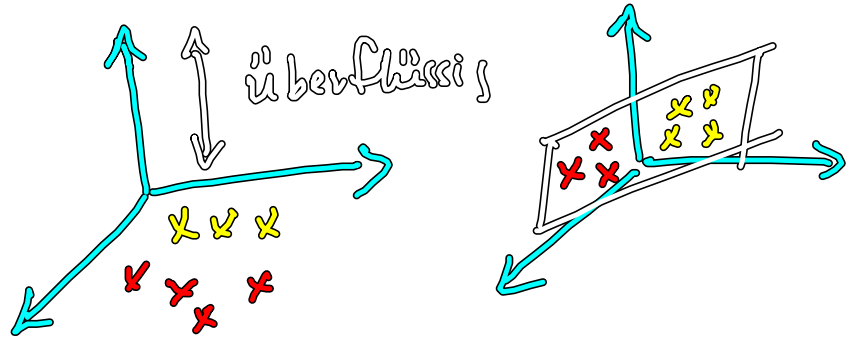
A3: Entropie 3 Gewinne

⋮
A20: ⋮

Wähle minimale Entropie



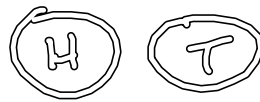
Merkmals-
reduktion



Entropie

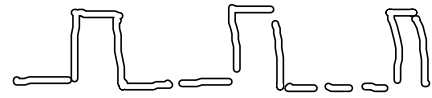
Experiment

Münze



$\frac{1}{2}$ $\frac{1}{2}$

Signal



Entropie

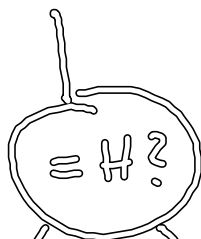
$$\text{Entropie} = - \sum P(c_i) \log_2(P(c_i))$$

$$= - \left(\underbrace{\frac{1}{2} \log_2 \frac{1}{2}}_H + \frac{1}{2} \log_2 \frac{1}{2} \right)$$

$$= - \left(\frac{1}{2} (-1) + \frac{1}{2} (-1) \right)$$

$$= 1$$

Experiment



yes

no



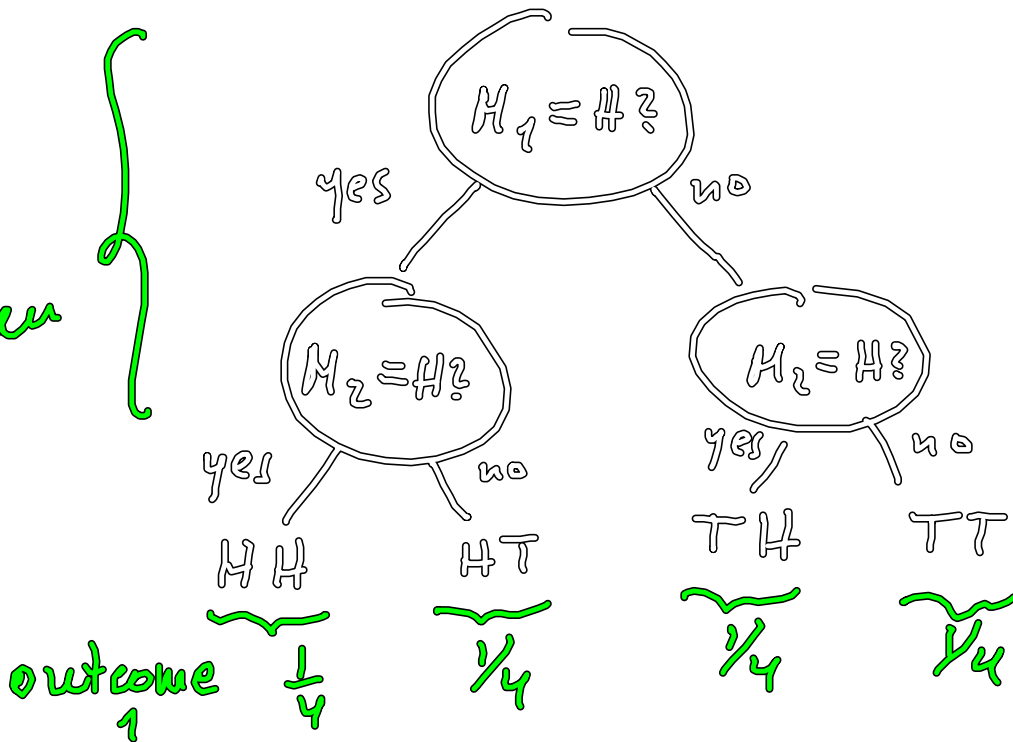
H

T

Experiment 2

①

②

 $\frac{1}{2}$
 $\frac{1}{2}$ H
T $\frac{1}{2}$
 $\frac{1}{2}$ H
T2
Fragen

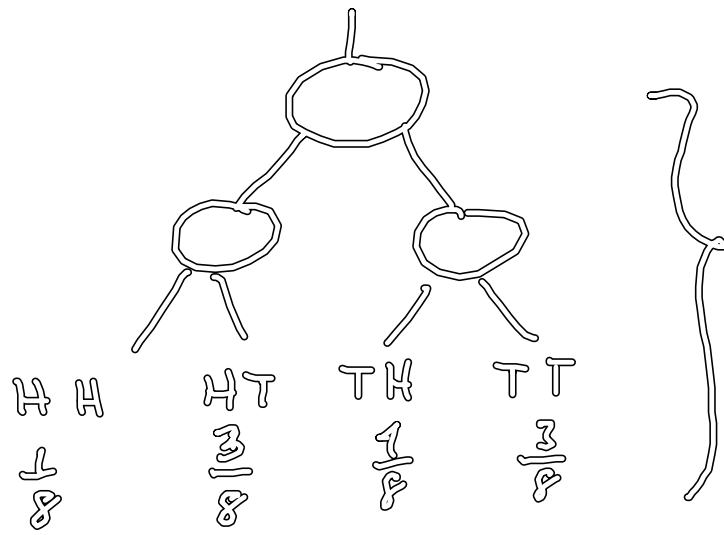
$$E = - \left(\frac{1}{4} \log \frac{1}{4} + \frac{1}{4} \log \frac{1}{4} + \dots + \frac{1}{4} \log \frac{1}{4} \right)$$

$$= -4 \left(\frac{1}{4} (-2) \right) = 2$$

M₁M₂ $\frac{1}{2}$
 $\frac{1}{2}$ H
T $\frac{1}{4}$
 $\frac{3}{4}$ H
T

M1

M2



$$E = - \left(\frac{1}{8} \log \frac{1}{8} + \frac{3}{8} \log \frac{3}{8} + \frac{1}{8} \log \frac{1}{8} + \frac{3}{8} \log \frac{3}{8} \right)$$

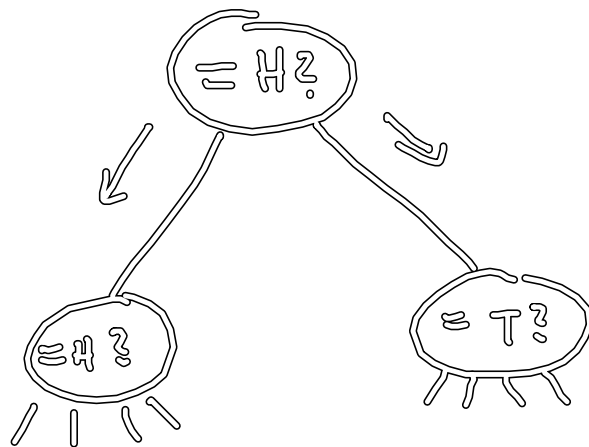
$$= - \frac{1}{4} (-3) - \frac{3}{4} \log \frac{3}{8}$$

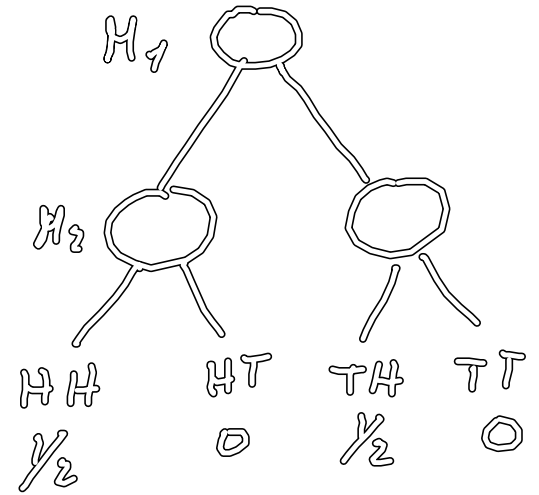
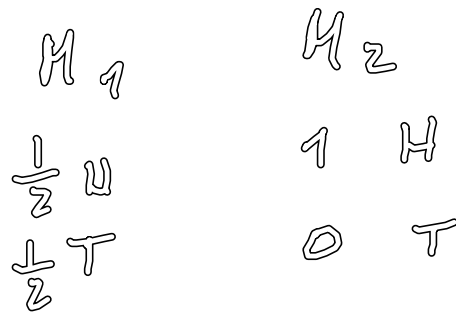
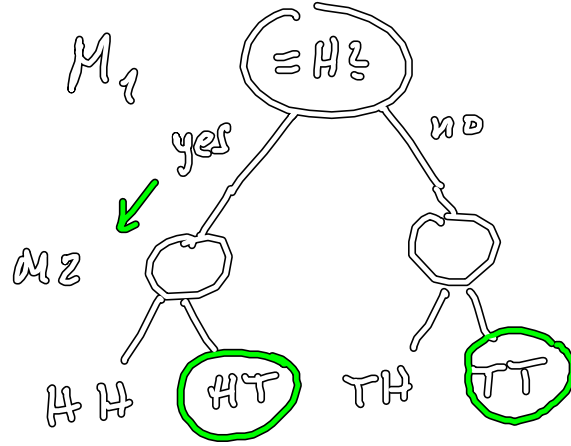
$$= 0,75 + 0,75 \left(\log \frac{8}{3} \right)$$

$$= \boxed{1,8}$$

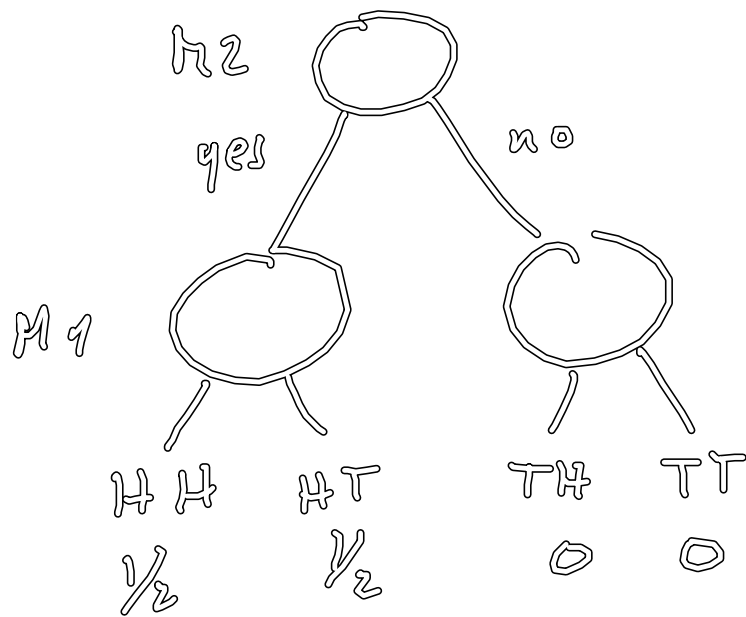
M2

M1





$$\begin{aligned}
 \text{Entropie} &= - \left(\frac{1}{2} \log \frac{1}{2} + 0 \log 0 \right. \\
 &\quad \left. + \frac{1}{2} \log \frac{1}{2} + 0 \log 0 \right) \\
 &= - \left(-\frac{1}{2} + \dots + \left(-\frac{1}{2}\right) + 0 \right) \\
 &= +1 \\
 \text{Konvention} \quad 0 \log 0 &= 0
 \end{aligned}$$



ID3 - Beispiel

Site

$$P(\text{safe} | \text{large}) = 5/7$$

$$P(\text{dangerous} | \text{large}) = 2/7$$

$$P(\text{large}) = 7/16$$

$$P(\text{safe} | \text{small}) = 5/9$$

$$P(\text{dangerous} | \text{small}) = 4/9$$

$$P(\text{small}) = 9/16$$

Bewertung der möglichen Anfragen

$$- \sum_{j=1}^m p(a_j) \sum_{\dots} p(c_i | a_j) \log(p(c_i | a_j))$$

Size

$a_1 = \text{large}$

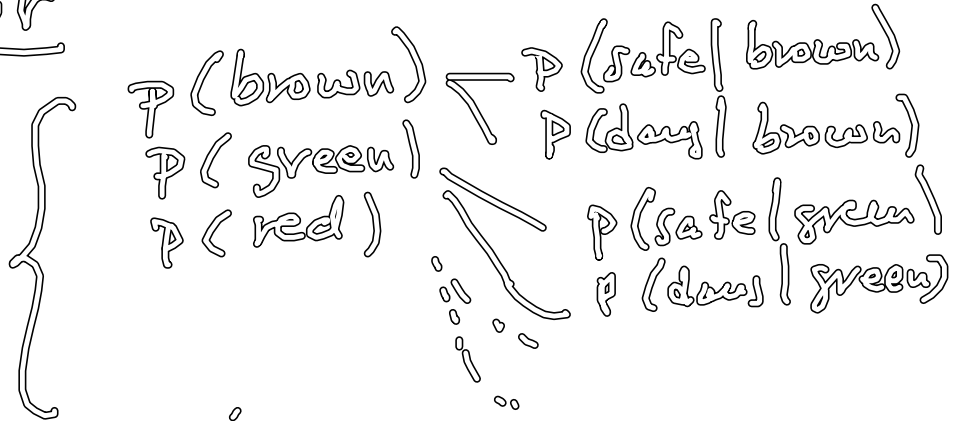
$a_2 = \text{small}$

$p(c_1 | a_1) = p(\text{safe} | \text{large})$

\vdots

$$\left. \begin{aligned} \text{Bewertung} &= - \left(\frac{7}{16} \left(\frac{5}{7} \log \frac{5}{7} + \frac{2}{7} \log \frac{2}{7} \right) \right. \\ &\quad \left. + \frac{9}{16} \left(\frac{5}{9} \log \frac{5}{9} + \frac{4}{9} \log \frac{4}{9} \right) \right) \\ &= 0,935 \end{aligned} \right\}$$

Color



= ... Zahl 2

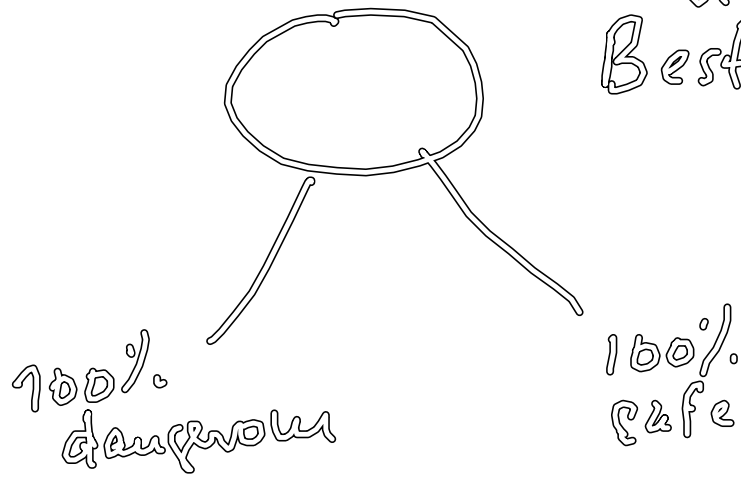
Skin

Zahl 3

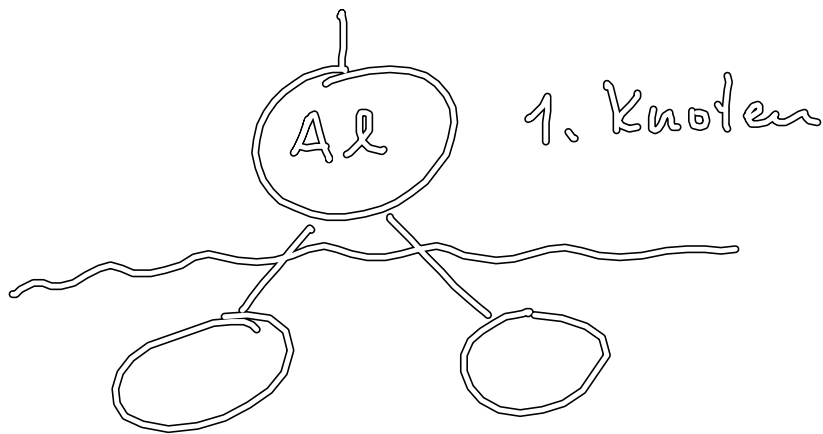
flesh

Zahl 4

„Überhaupt
Beste Trennung



$$1 \cdot \log(1) + 0 - \log 0 = 0$$



ohne
Attribut
2

Entscheidungsbäume

Nach-
folger

- ID3 (Entropie) → \mathcal{R}
- CHAID (Chi-Quadrat Test)
- C4.5 (stetige Attribute)