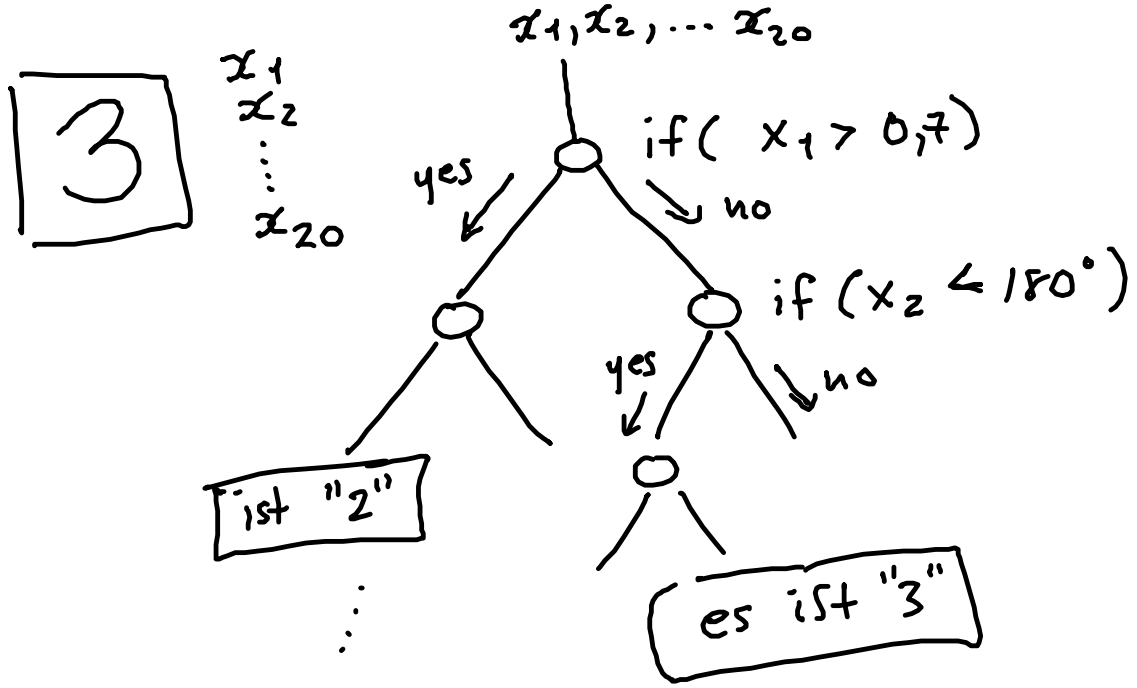
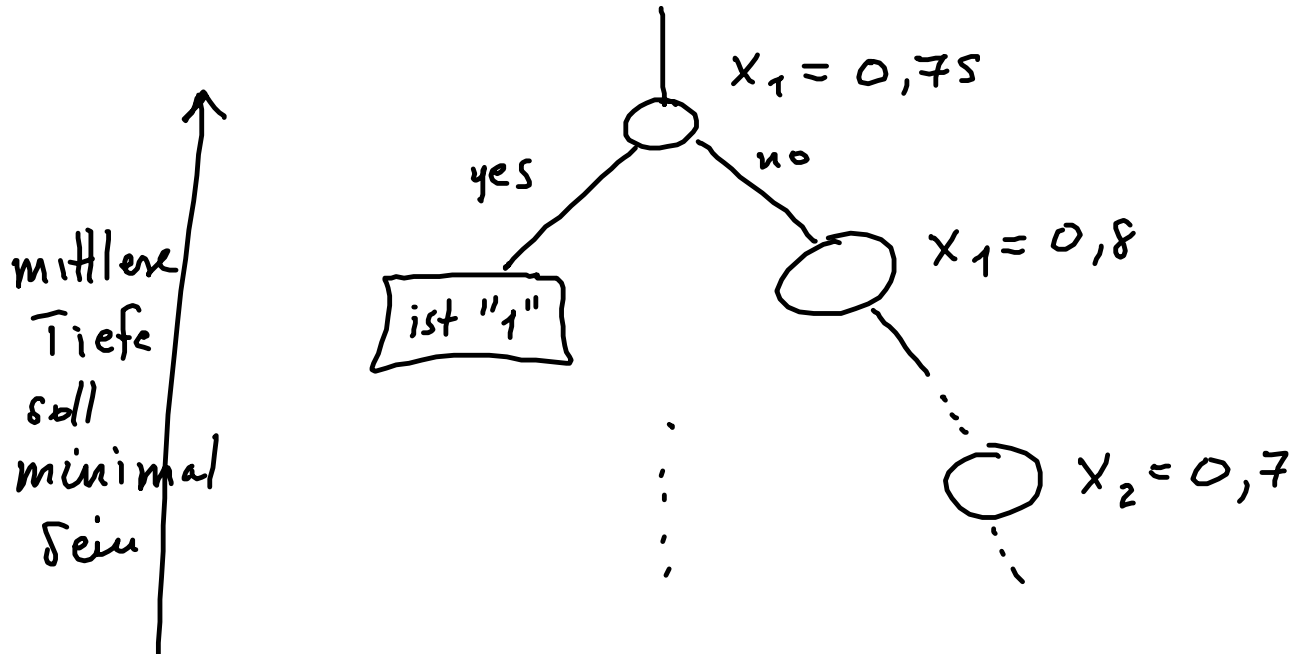
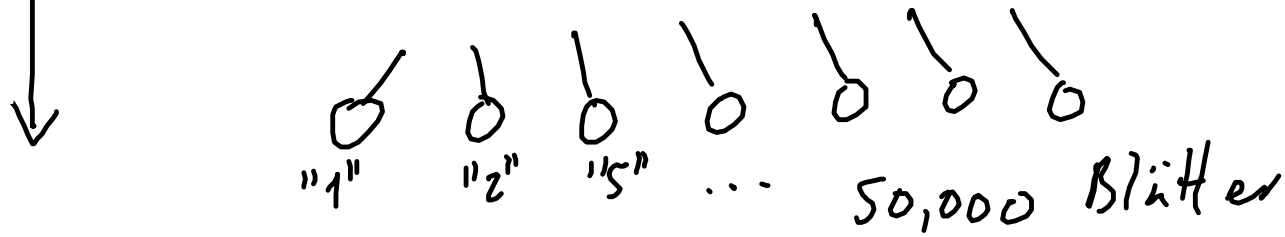
Backprop.

ID3 – Induction Trees

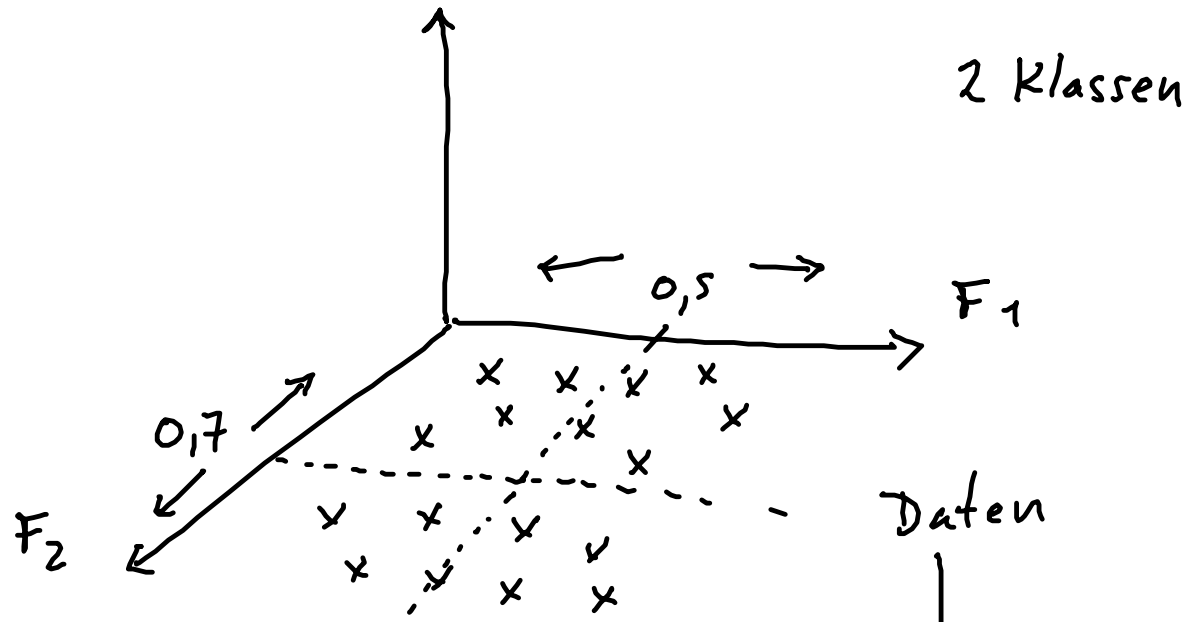


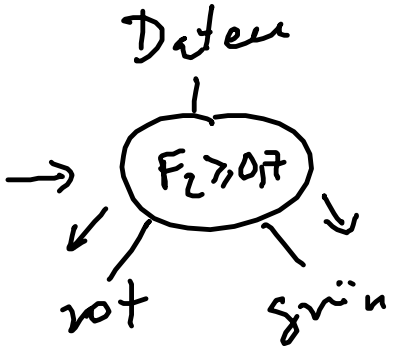
50,000 Zahlen

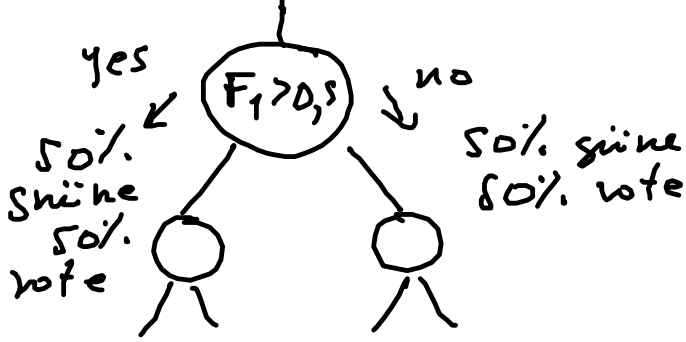


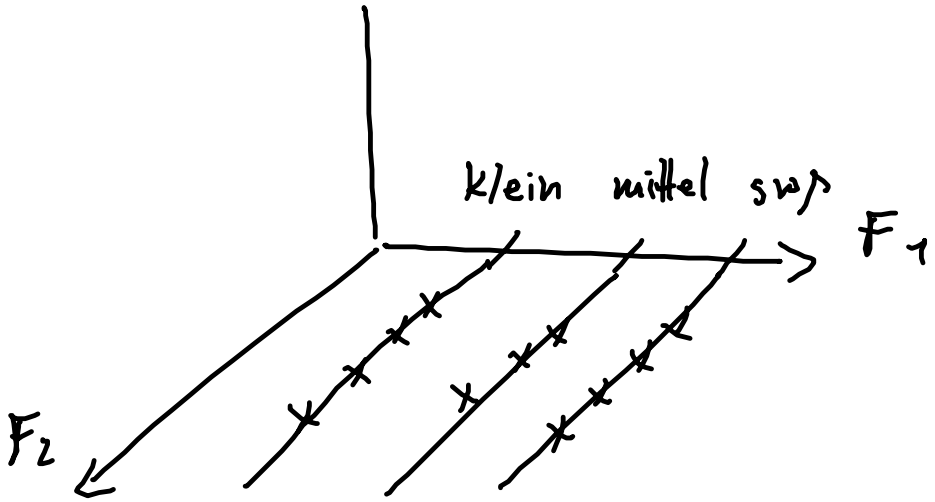


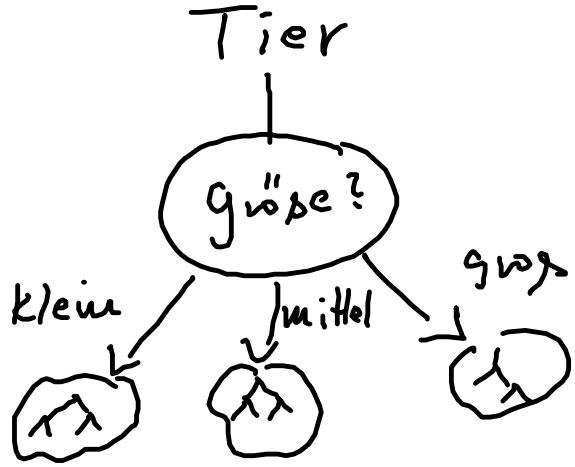
	X_1	X_2	X_3	.....	X_{20}
"1"	0,75	0,8	0,1	.....	
"5"	0,6	0,5			
⋮					





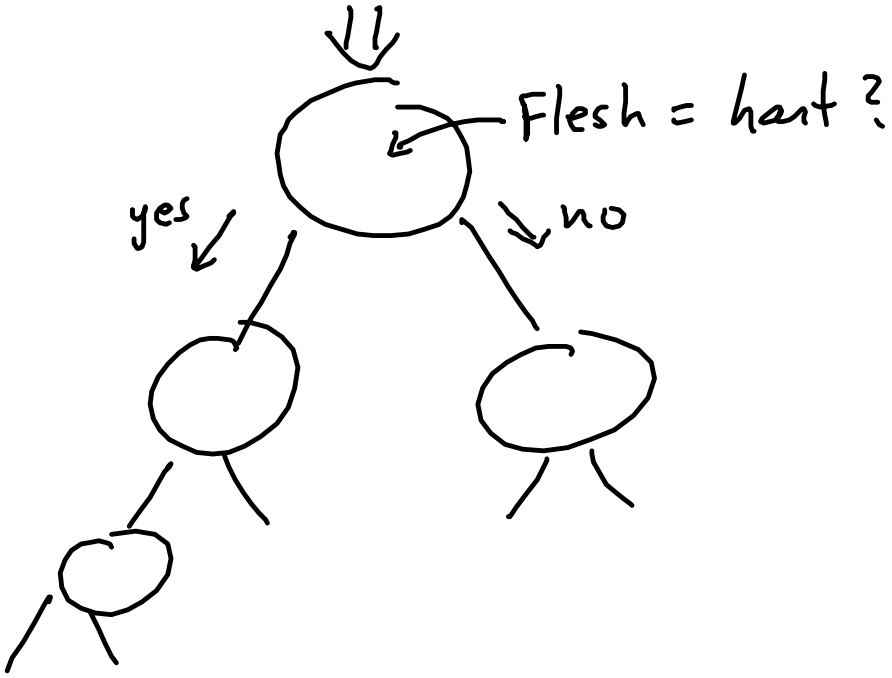


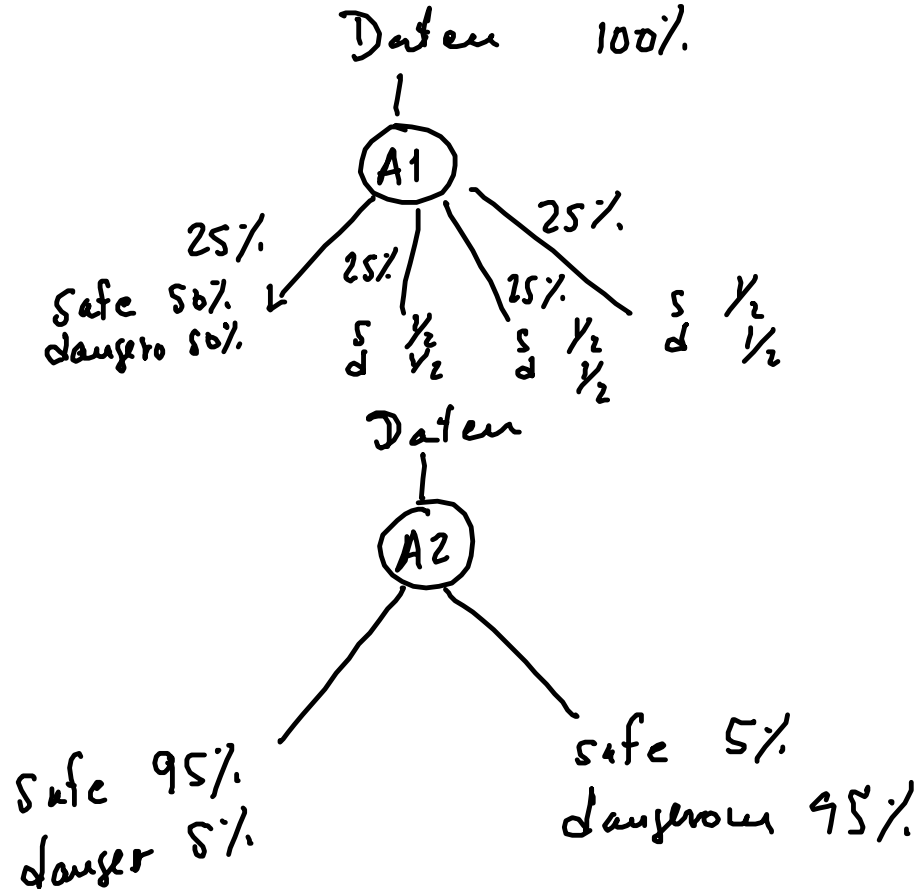




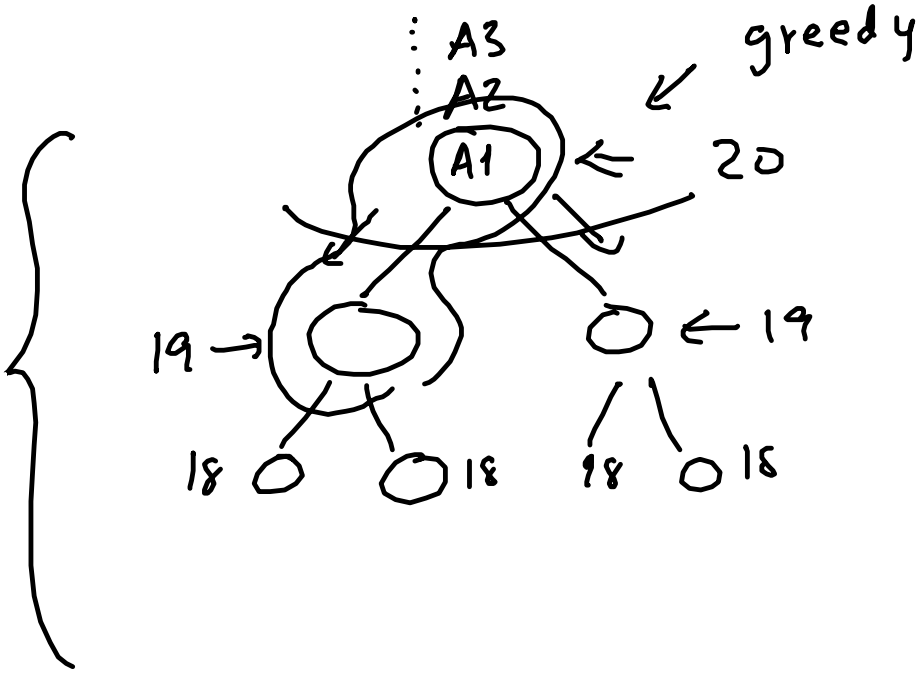
Unbekanntes Wesen

skin, color, size, flesh


Flesh = hart ?
yes
no

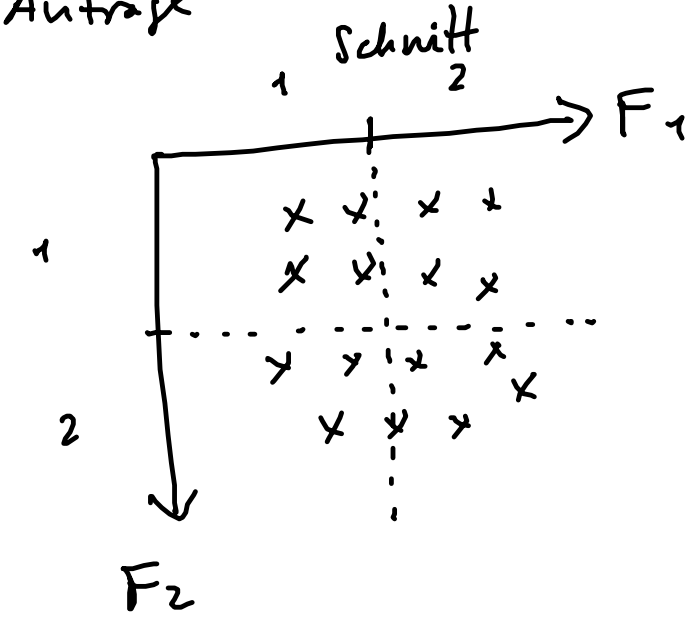

Daten 100%
A1
25%
25%
25%
25%
Safe 50%
danger 50%
Daten
A2
Safe 95%
danger 5%
safe 5%
dangerous 95%

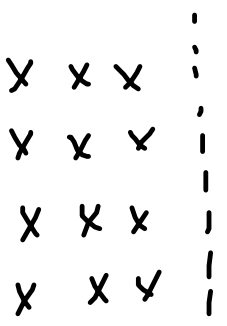
20 Merkmale



Relevanz der Anfrage

Messung der "Güte" der Verteilung nach Anfrage





xxv 100% pün
xx1 0% rot

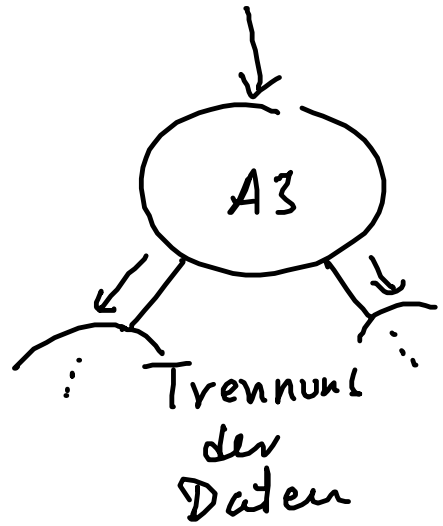
xxxv 0% grün
xxx 100% rot

{ 50% grün ...
  50% rot ...

ID3 **Entropie** $= -\sum_{i=1}^{n} p(c_i|a_j) \log(p(c_i|a_j))$

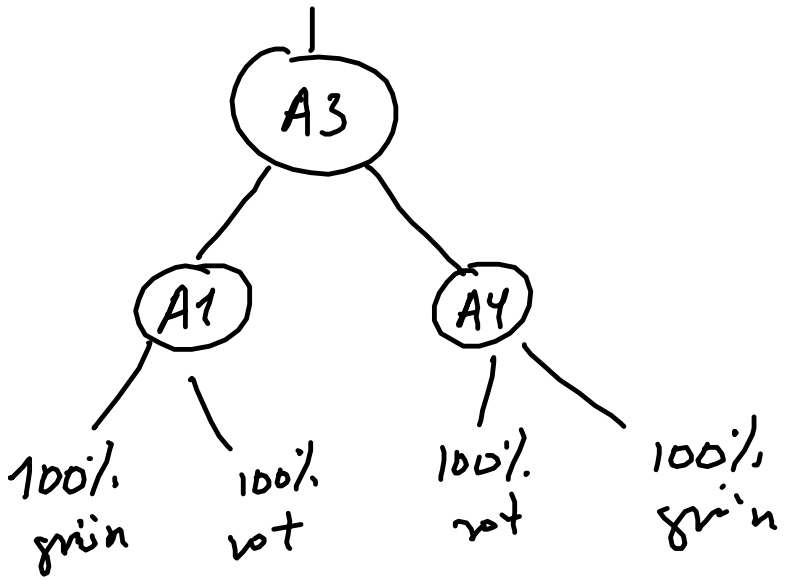
Informationsgewinn?

ID3

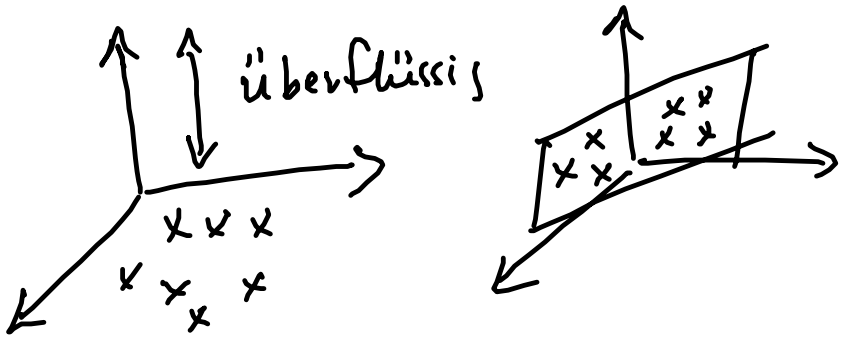


A1: Entropie$_1$
A2: Entropie$_2$
A3: Entropie$_3$ Gewinner
⋮ ⋮
A20:

wähle minimale Entropie



} Merkmals-reduktion



Entropie

Experiment

Münze

H T

1/2 1/2

Signal

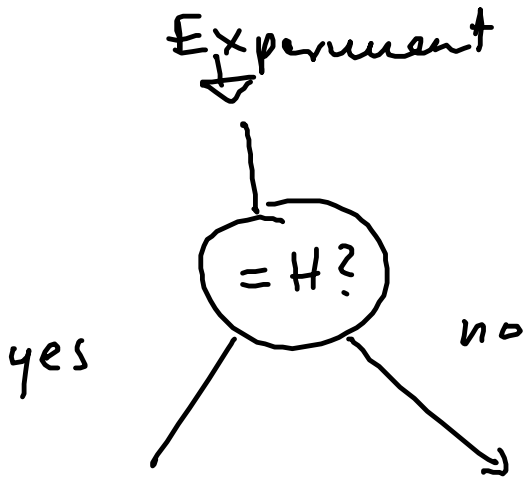
Entropie

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

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

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

$$= 1$$

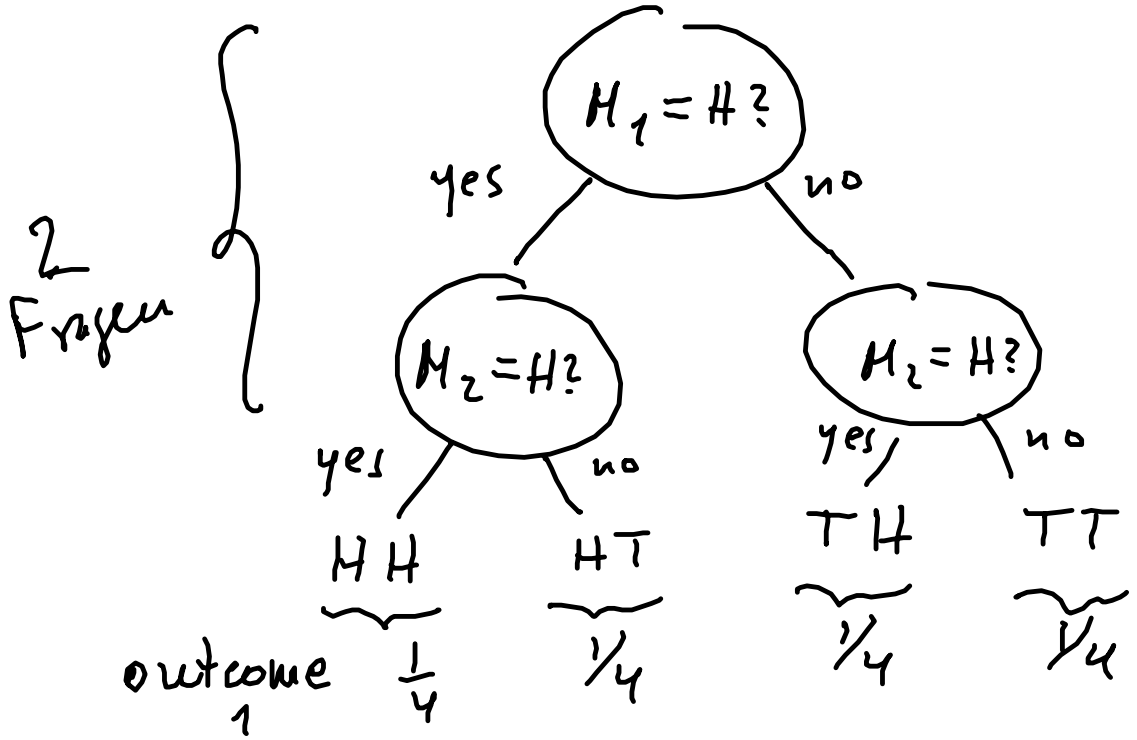


H T

Experiment 2

1 2

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



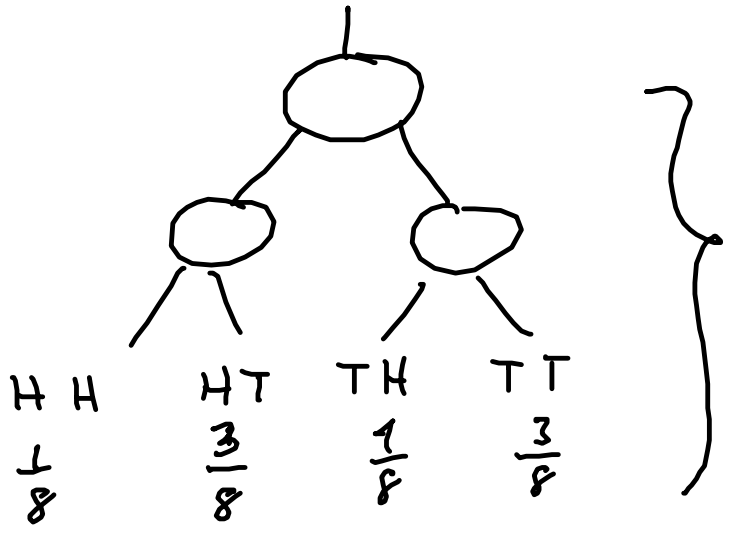
$$E = -\left(\frac{1}{4}\log\frac{1}{4} + \frac{1}{4}\log\frac{1}{4} + \cdots \frac{1}{4}\log\frac{1}{4}\right)$$

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

	H_1		H_2
$\frac{1}{2}$	H	$\frac{1}{4}$	H
$\frac{1}{2}$	T	$\frac{3}{4}$	T

M1

M2



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

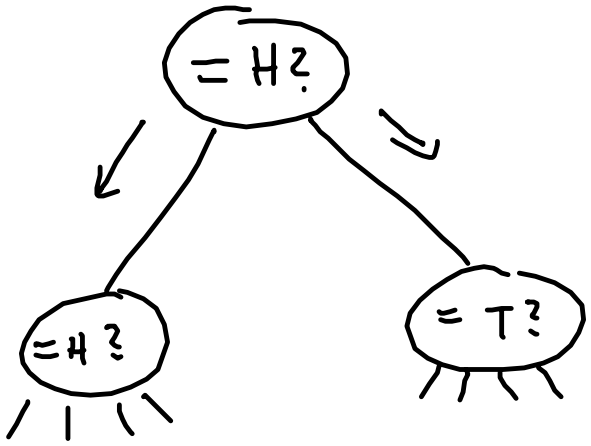
$$= -\frac{1}{4}(-3) - \frac{3}{4}\log\frac{3}{8}$$

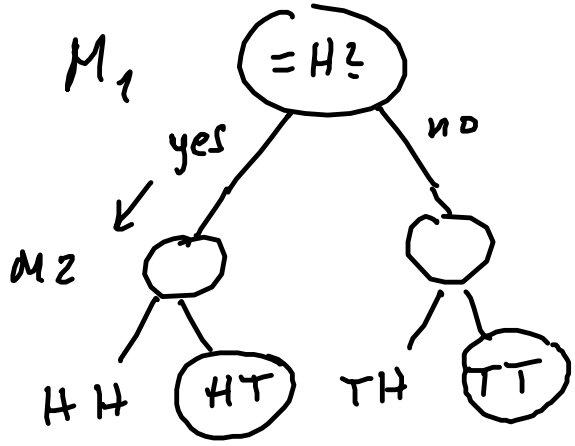
$$= 0{,}75 + 0{,}75\left(\log\frac{3}{8}\right)$$

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

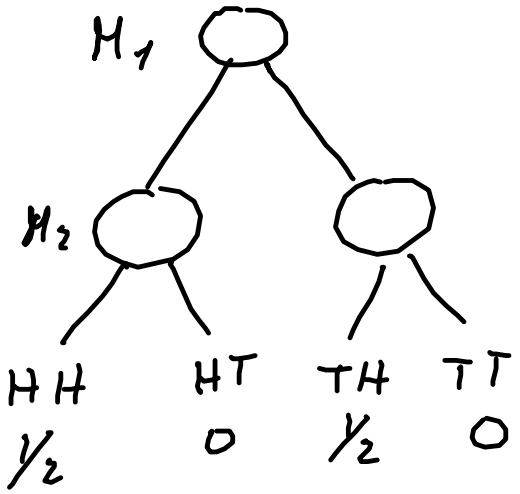
M2

M1





M_1	M_2
$\frac{1}{2}$ H	1 H
$\frac{1}{2}$ T	0 T

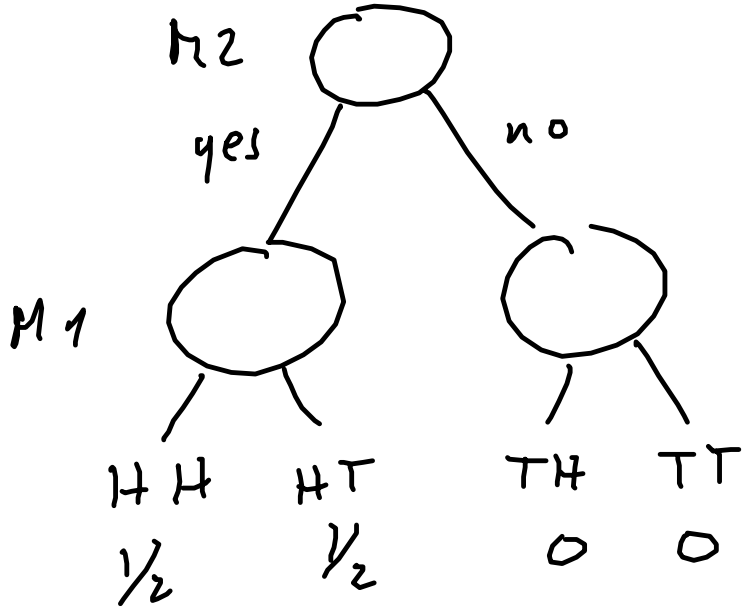


$$\text{Entropie} = -\left(\tfrac{1}{2}\log\tfrac{1}{2} + 0\log 0 + \tfrac{1}{2}\log\tfrac{1}{2} + 0\log 0 \right)$$

$$= -\left(-\tfrac{1}{2} + 0 + \left(-\tfrac{1}{2}\right) + 0 \right)$$

$$= +1$$

Konvention $0 \log 0 = 0$



ID3 – Beispiel

Size

$$P(\text{safe} \mid \text{large}) = 5/7$$
$$P(\text{dangerous} \mid \text{large}) = 2/7$$
$$P(\text{large}) = 7/16$$

$$P(\text{safe} \mid \text{small}) = 5/9$$
$$P(\text{dangerous} \mid \text{small}) = 4/9$$
$$P(\text{small}) = 9/16$$

Bewertung der möglichen Anfragen

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

Size

a_1 = large
a_2 = small

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

$$\text{Bewertung} = -\left(\frac{7}{16}\left(\frac{5}{7}\log\frac{5}{7} + \frac{2}{7}\log\frac{2}{7}\right) + \frac{9}{16}\left(\frac{5}{9}\log\frac{5}{9} + \frac{4}{9}\log\frac{4}{9}\right)\right)$$

$$= 0{,}935$$

Color

- $p(\text{brown})$ — $p(\text{safe}|\text{brown})$, $p(\text{dang}|\text{brown})$
- $p(\text{green})$ — $p(\text{safe}|\text{green})$, $p(\text{dang}|\text{green})$
- $p(\text{red})$
- ...

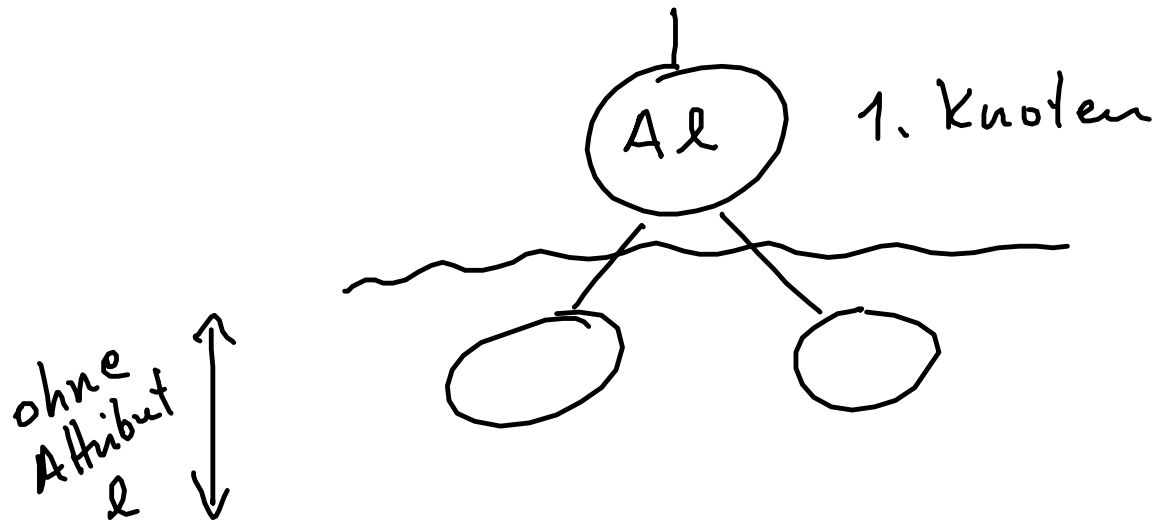
$= \dots$ Zahl 2

Skin Zahl 3

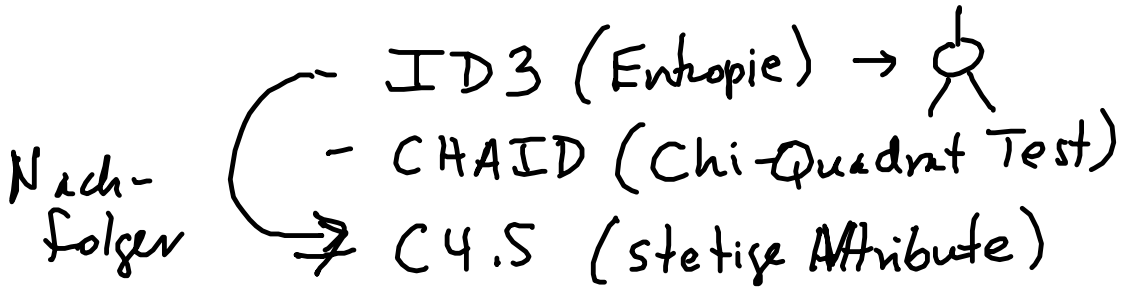
Flesh Zahl 4

Überhaupt Beste Trennung

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



Entscheidungsbäume


- ID3 (Entropie)
- CHAID (Chi-Quadrat Test)
Nach-
folger
C4.5 (stetige Attribute)