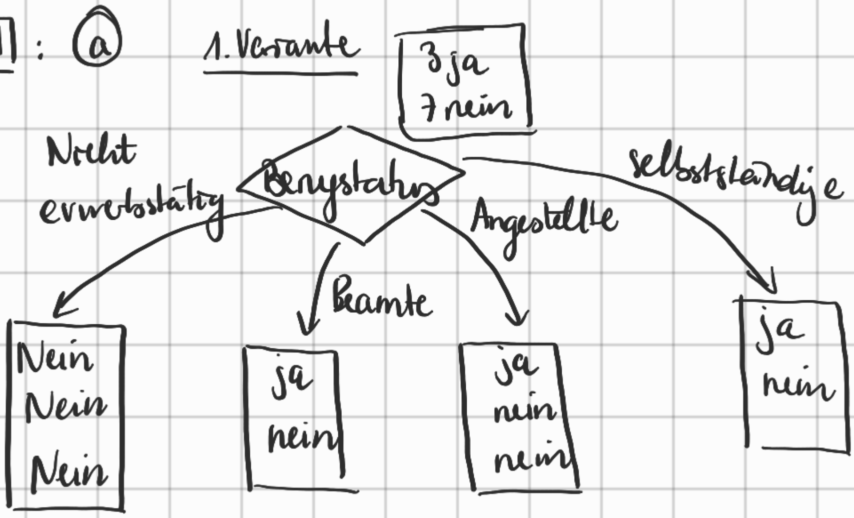
A1: a) 1. Variante



$$I_{parent} = \text{Info}([3,7]) = -\frac{3}{10} \log_2\left(\frac{3}{10}\right) - \frac{7}{10} \log_2\left(\frac{7}{10}\right)$$

$$= 0{,}8813 \text{ bits}$$

Berechnung durchschnittlicher $I_{children}$:

$$\text{info}([0,3]) = \text{entropy}([0,3])$$

$$= -\frac{0}{3} \cdot \log_2\left(\frac{0}{3}\right) - \frac{3}{3} \log_2\left(\frac{3}{3}\right)$$

$$= -0 - 1 \cdot \log_2(1) = 0 \text{ bit}$$

$$\text{info}([1,1]) = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1 \text{ bit}$$

$$\text{info}([1,2]) = -\frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) = 0{,}644 \text{ bit}$$

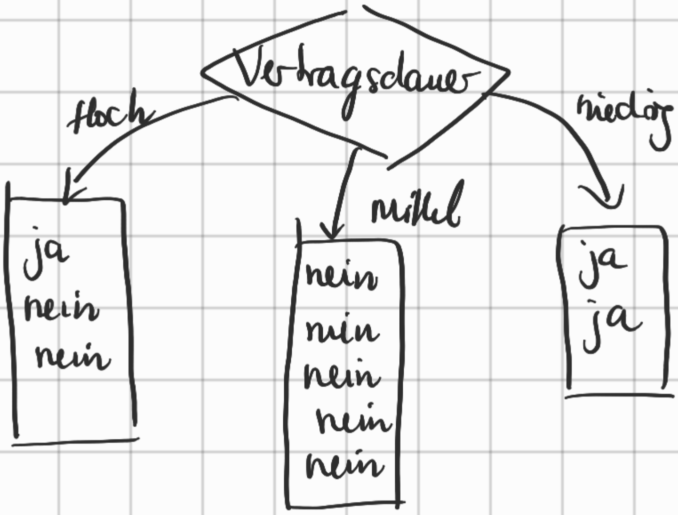
$$\text{info}([1,1]) = 1 \text{ bit}$$

$\Rightarrow$ im Mittel: $\left(\frac{3}{10}\right) * 0 + \frac{2}{10} \cdot 1 + \frac{2}{10} \cdot 1 + \frac{3}{10} \cdot 0{,}644 = 0{,}5932$ bit

Somit ist: gain (Berufstatus)

$$= 0{,}8813 \text{ bit} - 0{,}5932 \text{ bit} = 0{,}288 \text{ bit}$$

2. Variante



Analog der obigen Berechnung haben wir

$info([1,2]) = 0{,}644$ bit

$info([0,5]) = 0$ bit

$info([2,0]) = 0$ bit

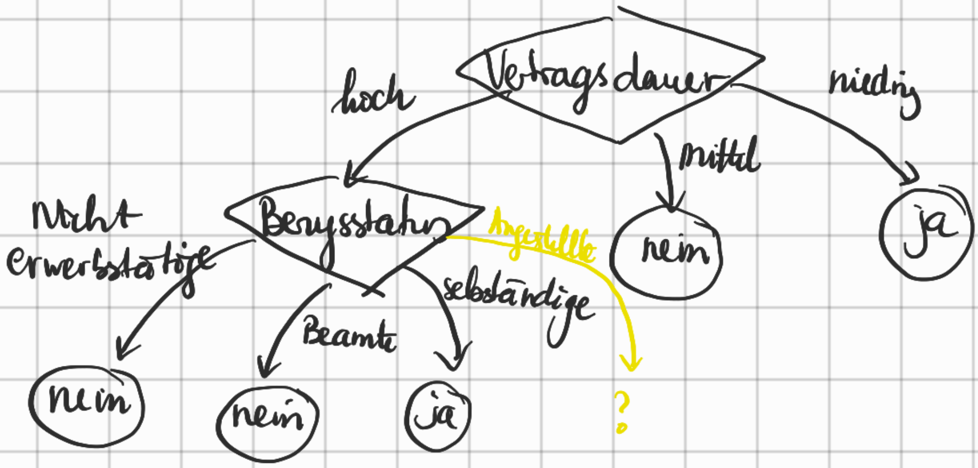
$\Rightarrow$ im Mittel: $I_{children} = \frac{3}{10}(0{,}644) = 0{,}1932$ bits

Somit gain(Vertragsdauer) $= 0{,}8813 - 0{,}1932 = 0{,}6881$ bit

$\rightarrow$ Reihenfolge der Attribute: Vertragsdauer dann Berufsstatus

da gain(Vertragsdauer) > gain(Berufsstatus)

$\Rightarrow$ entstandener Baum:



(b)

Kundennummer	Kündigung
11	ja
12	nein

← Kündigungsgezählt.

(c) Entscheidungsregeln

- If (Vertragsdauer == Hoch) && (Berufsstatus == Nicht-Erwerbstätige
  ∨ Berufsstatus == Beamte)

  then Klasse ist Nein

- If (Vertragsdauer == Hoch) && (Berufsstatus == Selbstständige)
  then Klasse ist ja

- If (Vertragsdauer == mittel) then Klasse ist nein

- If (Vertragsdauer == niedrig) then Klasse ist ja

(d) Information-Gain := wird verwendet, um die besten Merkmale (Attribute) zu bestimmen, die eine Maximum an Informationen über eine Klasse liefern.

Gain-Ratio := wird zur Normierung des Information gain eines Attributes gegenüber der Entropie dieses Attributs verwendet, um zu vermeiden, dass das Attributi, das mehrere Splits haben, immer gewählt wird.

Gini-Index :=

e) $G_{orginal} = 1 - \left(\frac{3}{10}\right)^2 - \left(\frac{7}{10}\right)^2 = 1{,}4$

Berechnung gain (Vertragsdauer) mit Gini index:

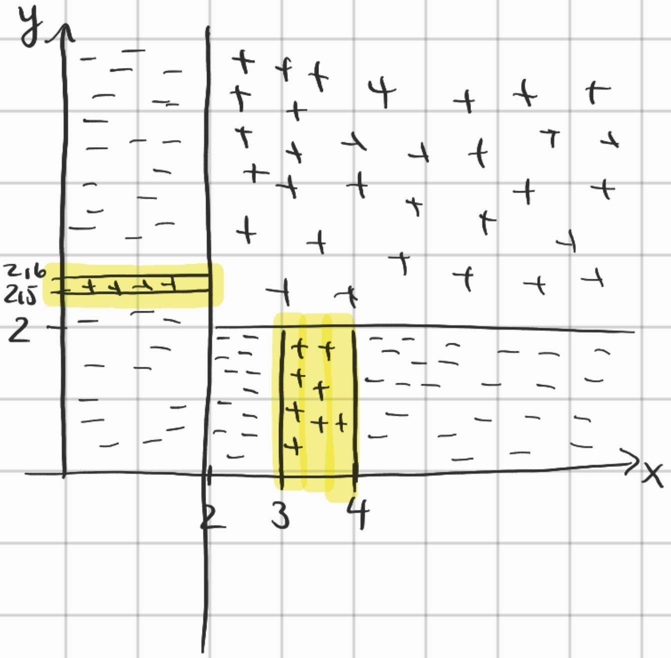
$$\left.\begin{aligned} G_{ver.=hoch} &= 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2 = 1{,}3 \\ G_{ver.=mittel} &= 1 - \left(\frac{0}{5}\right)^2 - \left(\frac{5}{5}\right)^2 = 0 \\ G_{ver=niedrig} &= 1 - \left(\frac{2}{2}\right)^2 - \left(\frac{0}{2}\right)^2 = 0 \end{aligned}\right\} \Rightarrow G_{Childen} = \frac{3}{10} * 1{,}3 = 0{,}4$$

$$\Rightarrow \Delta_{Vertragsdauer} = G_{orginal} - G_{children}$$

$$= 1{,}4 - 0{,}4 = 1$$

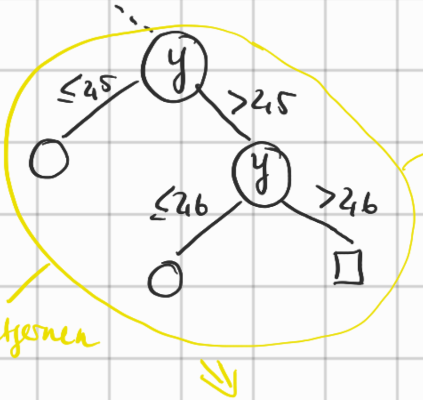
A2:

a) Legende: ■ := − ; O := + } partitions bezeichner



b) Overfitting: Partition

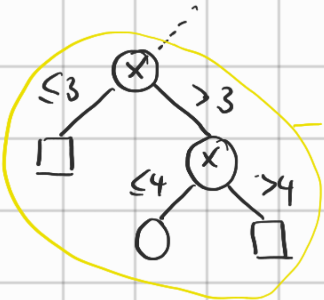
① partition



zuentfernen

Vermutlich sind die Ausreißer
⇒ zu viel an die Daten angepasst.

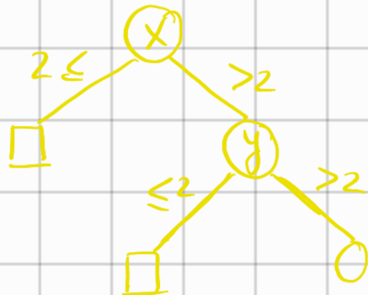
② partition



zuentfernen

gelb markierte partition

⇒ Mögliche entstandener Baum:



ⓒ Generalisierungsfähigkeit

:= ist die Fähigkeit eines Modells, sich an neue bisher unbekannte Daten anzupassen und zwar aus derselben Verteilung wie jene, die zum Erstellen des Modells verwendet wurde.