

# CAL3

---

Carsten Gips (HSBI)

Unless otherwise noted, this work is licensed under CC BY-SA 4.0.

## CAL3: Erweiterung von CAL2 für nicht-disjunkte Klassen

1) Anfangsschritt:  $\alpha^{(0)} = *$  (totales Unwissen)

2)  $n$ -ter Lernschritt: Objekt  $v$  mit Klasse  $k$

- Rückweisung (Endknoten mit  $*$ ): Ersetze  $*$  durch Vereinigungsklasse  $/k1/$
- Endknoten mit Vereinigungsklasse:
  - Zähler für  $k$  erhöhen, bzw.
  - $k$  mit Anzahl 1 in Vereinigungsklasse einfügen

Falls nun die Summe aller Klassen am Endknoten größer/gleich  $S_1$  (Statistikschwelle):

- Für **genau eine** Klasse gilt:  $P(k|\tilde{x}) \geq S_2$ :  
=> Abschluss: Ersetze Vereinigungsklasse durch  $k$  (für immer!)
- Für **alle** Klassen gilt:  $P(k|\tilde{x}) < S_2$ :  
=> Differenzierung: Ersetze Vereinigungsklasse durch neuen Test:  $\kappa \leftarrow x_{t+1}(*, \dots, *, /k1/, *, \dots, *)$

$x_{t+1}$ : nächstes Attribut, auf dem aktuellen Pfad  $\tilde{x}$  noch nicht verwendet

Symbol  $k$  mit Anzahl 1 an Position  $i$  wenn  $x_{t+1}(v) = i$

## Beispiel mit CAL3

$x_1$	$x_2$	$k$
0	0	A
0	1	B
0	1	A
1	0	B
1	1	A

- $S_1 = 4, S_2 = 0.7$

## Beispiel mit CAL3

$x_1$	$x_2$	$k$
0	0	A
0	1	B
0	1	A
1	0	B
1	1	A

- $S_1 = 4, S_2 = 0.7$

Tafelbeispiel CAL3

**Ergebnis:**  $x_1(A, x_2(B, A))$

Trainingsfehler:  $1/5 = 0.2 < 1 - S_2 = 1 - 0.7 = 0.3$

- **Parameter:**

- $S_1$ : Statistikschwelle, problemabhängig wählen
- $S_2$ :  $0.5 < S_2 \leq 1.0$
- Klassifikationsfehler kleiner als  $1 - S_2$ 
  - kleiner Fehler  $\Rightarrow$  großer Baum
  - großer Fehler  $\Rightarrow$  kleiner Baum

- **Abbruch:**

- Alle Trainingsobjekte richtig klassifiziert  
 $\Rightarrow$  Kein Fehler in einem kompletten Durchlauf
- Alle Endknoten mit eindeutigen Klassensymbolen belegt
- Differenzierung nötig, aber alle Merkmale verbraucht
- Lernschrittzahl überschritten

- CAL3: Erweiterung von CAL2 für überlappende Klassen
  - Parameter  $S_1$  (Anzahl Objekte bis Entscheidung),  $S_2$  (Dominanz?)
  - Trainingsfehler wg. überlappender Klassen!

# LICENSE



Unless otherwise noted, this work is licensed under CC BY-SA 4.0.