

Seminarvortrag zum Thema  
**maschinelles Lernen I - Entscheidungsbäume**

von  
Lars-Peter Meyer

im Seminar  
Methoden wissensbasierter Systeme

bei  
Prof. Brewka  
im WS 2007/08

# Übersicht

- Überblick maschinelles Lernen
- Einführung in Entscheidungsbäume
- Lernen von Entscheidungsbäumen
- Informationsgehalt und ID3 sowie C4.5

# Überblick maschinelles Lernen

- Eine mögliche Definition von maschinellem Lernen [siehe Buch] :  
„Research in machine learning has been concerned with building computer programs able to construct new knowledge or to improve already possessed knowledge by using input information“
- Klassifikation maschinellen Lernens vorgeschlagen nach
  - Lernstrategie
  - Repräsentation von Wissen
  - Anwendungsbereich

# Klassifikation nach Lernstrategie

- Direkte Eingabe neuen Wissens und Auswendiglernen
- Lernen durch Anweisungen
- Lernen durch Deduktion
- Lernen durch Analogie
- Lernen aus Beispielen
- Lernen aus Beobachtungen und durch Entdeckungen

# Klassifikation nach Lernstrategie 1

- **Direkte Eingabe neuen Wissens und Auswendiglernen**
  - direktes Speichern von Daten und Fakten ohne Schlussfolgerungen
  - direkte Programmierung
  - so gut wie kein Aufwand beim Lernenden
- Lernen durch Anweisungen
- Lernen durch Deduktion
- Lernen durch Analogie
- Lernen aus Beispielen
- Lernen aus Beobachtungen und durch Entdeckungen

# Klassifikation nach Lernstrategie 2

- Direkte Eingabe neuen Wissens und Auswendiglernen
- **Lernen durch Anweisungen**
  - Lehrer gibt das neue Wissen in leicht verarbeitbarer Form und schrittweise an
  - hoher Aufwand beim Lehrer
- Lernen durch Deduktion
- Lernen durch Analogie
- Lernen aus Beispielen
- Lernen aus Beobachtungen und durch Entdeckungen

# Klassifikation nach Lernstrategie 3

- Direkte Eingabe neuen Wissens und Auswendiglernen
- Lernen durch Anweisungen
- **Lernen durch Deduktion**
  - Deduktion von neuem Wissen aus bereits vorhandenem
  - Logik-Programmierung
  - hat laut Buch in den letzten Jahren an Bedeutung gewonnen
- Lernen durch Analogie
- Lernen aus Beispielen
- Lernen aus Beobachtungen und durch Entdeckungen <sup>7</sup>

# Klassifikation nach Lernstrategie 4

- Direkte Eingabe neuen Wissens und Auswendiglernen
- Lernen durch Anweisungen
- Lernen durch Deduktion
- **Lernen durch Analogie**
  - Verallgemeinerung des Wissens und Übertragung auf neue Situationen bzw. Gebiete
  - Schwierigkeit des „Wiedererkennens“
- Lernen aus Beispielen
- Lernen aus Beobachtungen und durch Entdeckungen



# Klassifikation nach Lernstrategie 5

- Direkte Eingabe neuen Wissens und Auswendiglernen
- Lernen durch Anweisungen
- Lernen durch Deduktion
- Lernen durch Analogie
- **Lernen aus Beispielen**
  - Konzept soll möglichst alle Beispiele umfassen und negative Bsp. ausschließen
  - Unterscheidung nach
    - **Quelle**(Lehrer, Lernender, Umwelt) der Beispiele
    - **Art**(positiv&negativ, positiv)
    - **Ablauf** (inkrementell, gleichzeitig)
  - große Schlussfolgerungsfähigkeit notwendig
- Lernen aus Beobachtungen und durch Entdeckungen 9

# Klassifikation nach Lernstrategie 6

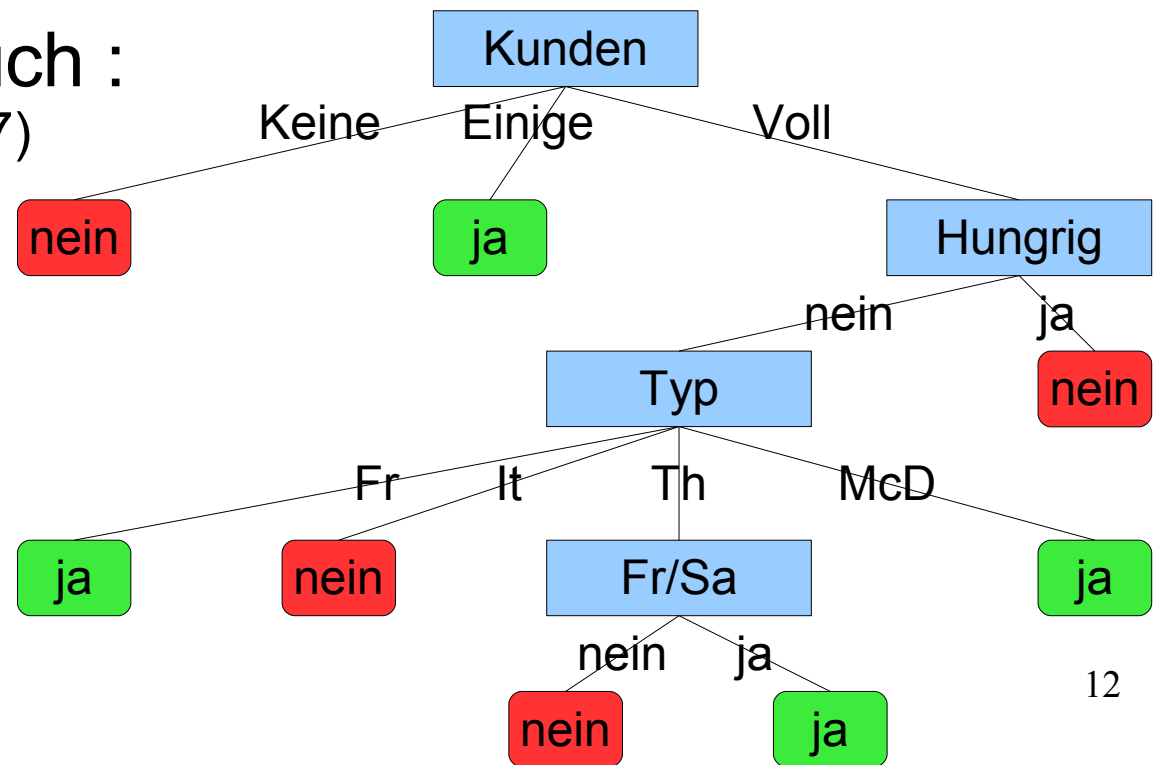
- Direkte Eingabe neuen Wissens und Auswendiglernen
- Lernen durch Anweisungen
- Lernen durch Deduktion
- Lernen durch Analogie
- Lernen aus Beispielen
- **Lernen aus Beobachtungen und durch Entdeckungen**
  - anspruchsvollste Form des Lernens, keine Steuerung durch einen Lehrer
  - Lernen durch
    - passive Beobachtung
    - aktive Experimente

# Klassifikation nach Repräsentation des Wissens

- Parameter in algebraischen Ausdrücken
- Entscheidungsbäume
- Formale Grammatiken
- Regeln
- Ausdrücke basierend auf formaler Logik
- Begriffshierarchien
- .....

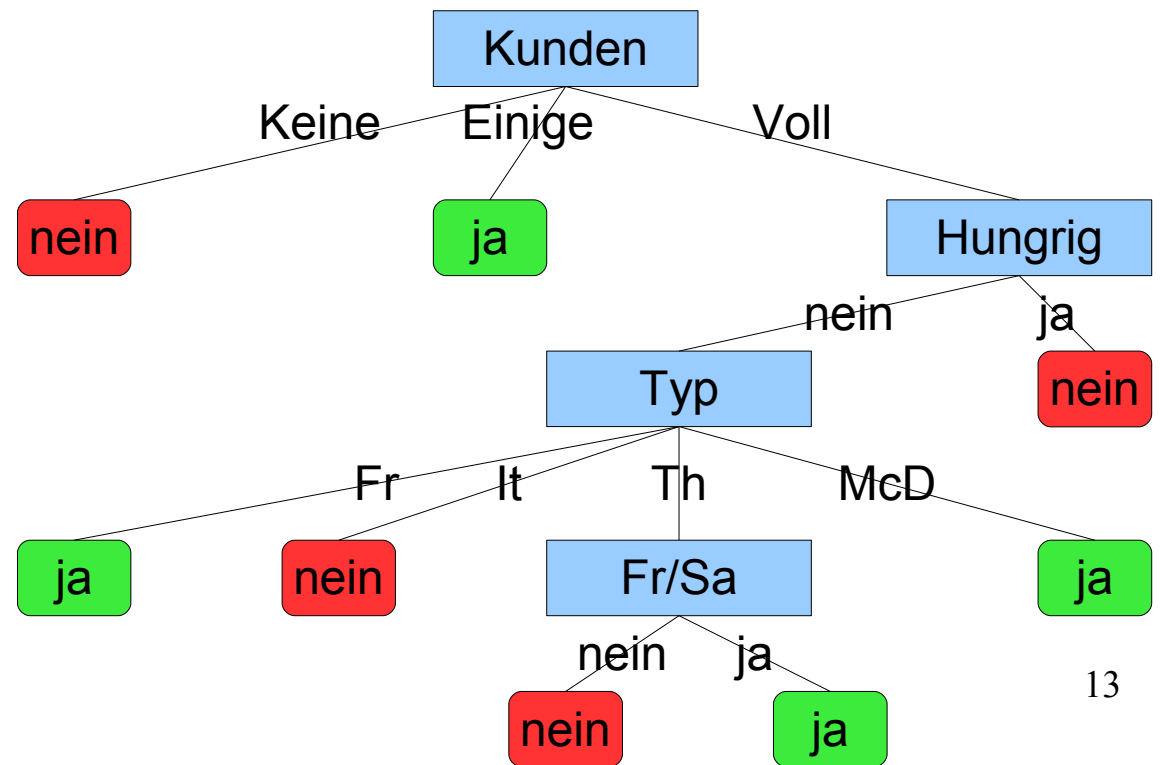
# Entscheidungsbäume

- gerne Einschränkung auf Klassifikation mit nur zwei Klassen
- Regeln ableitbar
- Erzeugbar aus Beispielen
- Bsp. Restaurantbesuch :  
(siehe Brewka, WBS im SS07)

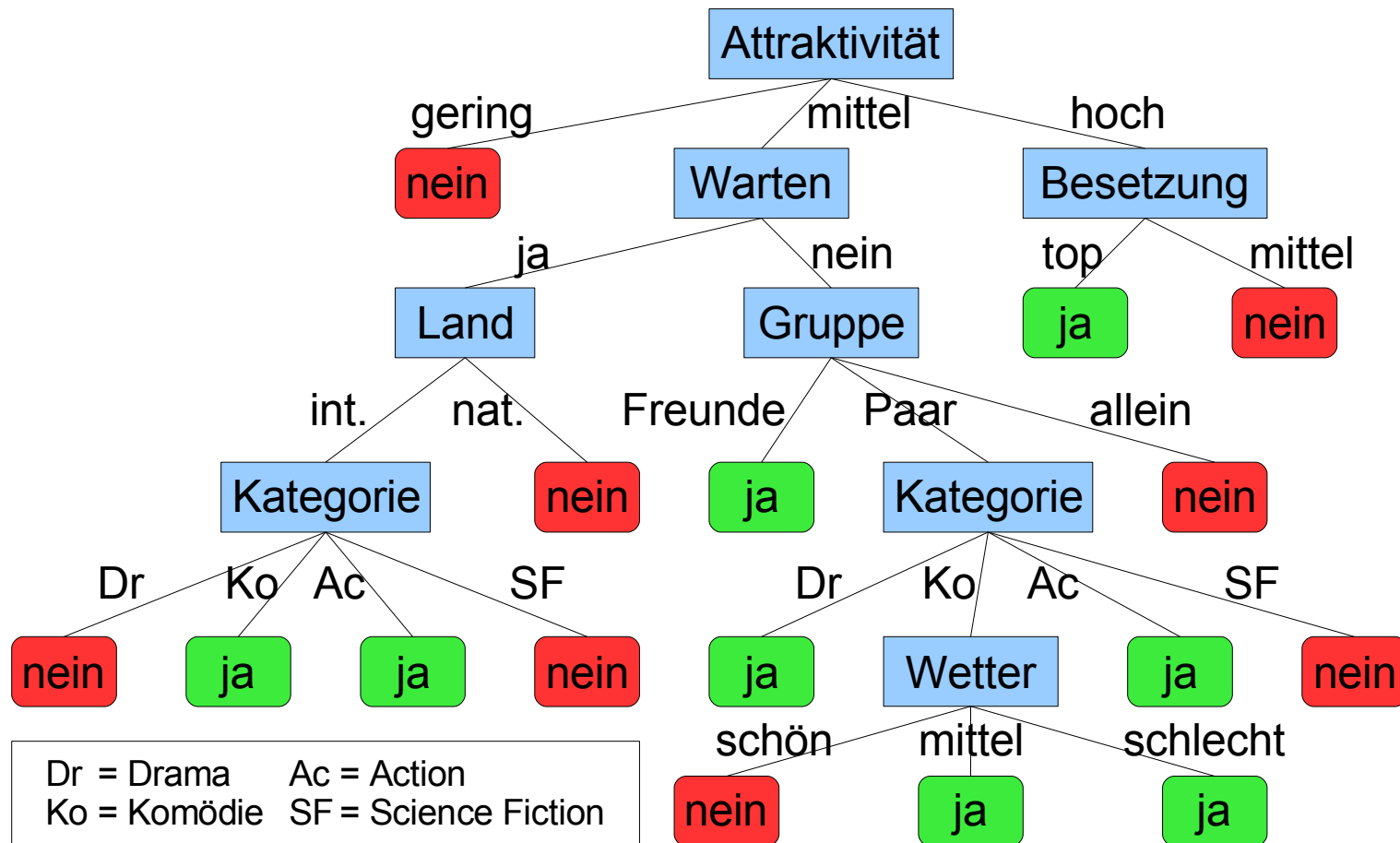


# Regeln zu Restaurantbesuch

- if Kunden=Keine then nein
- if Kunden=Einige then ja
- if Kunden=Voll & Hungrig=ja then nein
- ...



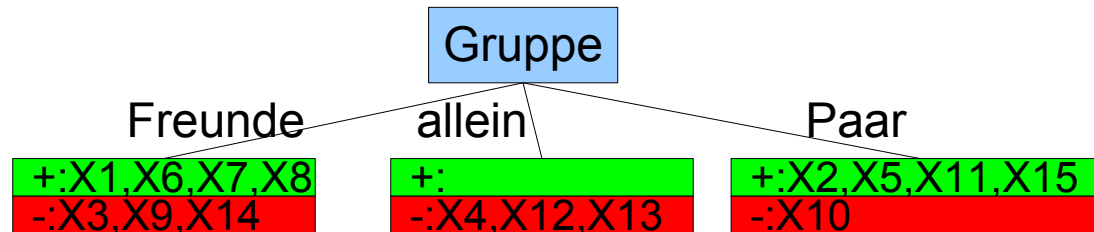
# Beispiele zum Kinobesuch - Baum



# Beispiele zum Kinobesuch - Tabelle

Bsp.	Attr.	€	Loge	Wetter	Warten	Bes	Kat.	Land	Res.	Gruppe	Erg.
X1	+	€€	ja	-	ja	+	AC	int.	ja	Freunde	ja
X2	0	€	ja	0	nein	0	KO	int.	nein	Paar	ja
X3	0	€	nein	0	ja	0	DR	int.	nein	Freunde	nein
X4	-	€	ja	0	ja	0	SF	int.	nein	allein	nein
X5	0	€	ja	0	nein	0	DR	int.	nein	Paar	ja
X6	+	€€	ja	+	nein	+	SF	int.	ja	Freunde	ja
X7	0	€	ja	-	nein	0	KO	nat.	nein	Freunde	ja
X8	0	€	nein	-	ja	0	AC	int.	nein	Freunde	ja
X9	-	€	ja	+	nein	0	KO	nat.	nein	Freunde	nein
X10	0	€	ja	+	nein	0	KO	int.	nein	Paar	nein
X11	+	€	ja	0	ja	+	DR	int.	nein	Paar	ja
X12	0	€	nein	-	ja	0	AC	nat.	nein	allein	nein
X13	+	€€	ja	0	ja	0	SF	int.	nein	allein	nein
X14	0	€	ja	+	ja	+	DR	int.	ja	Freunde	nein
X15	0	€	ja	-	nein	0	AC	int.	nein	Paar	ja

# Mögliche Auswahl-Attribute für den Kinobesuch - Gruppe

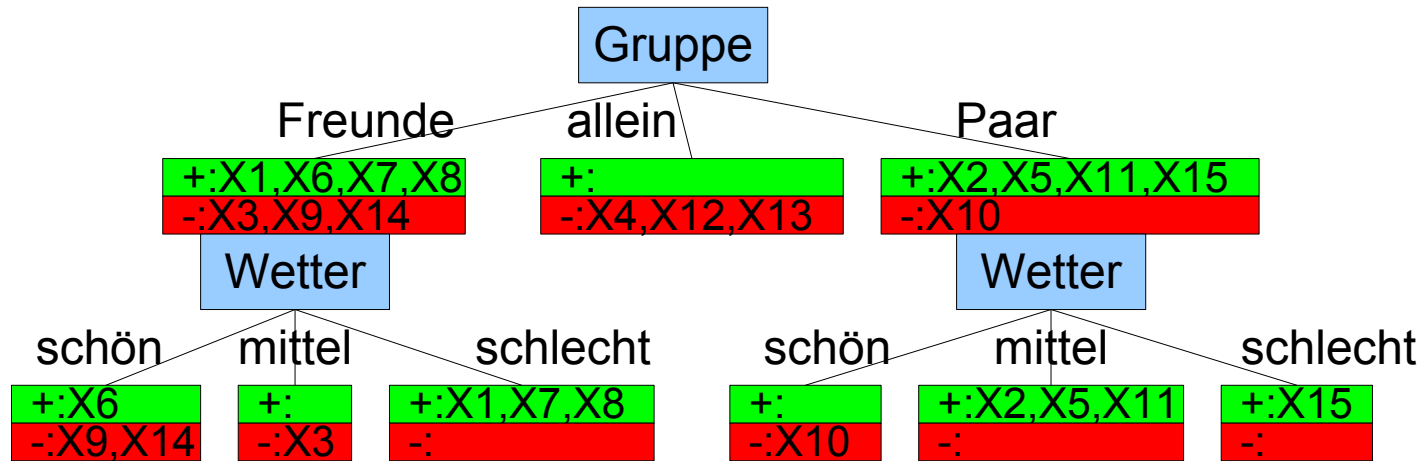


Bsp.	Attr.	€	Loge	Wetter	Warten	Bes	Kat.	Land	Res.	Gruppe	Erg.
X1	+	€€	ja	-	ja	+	AC	int.	ja	Freunde	ja
X3	O	€	nein	O	ja	O	DR	int.	nein	Freunde	nein
X6	+	€€	ja	+	nein	+	SF	int.	ja	Freunde	ja
X7	O	€	ja	-	nein	O	KO	nat.	nein	Freunde	ja
X8	O	€	nein	-	ja	O	AC	int.	nein	Freunde	ja
X9	-	€	ja	+	nein	O	KO	nat.	nein	Freunde	nein
X14	O	€	ja	+	ja	+	DR	int.	ja	Freunde	nein

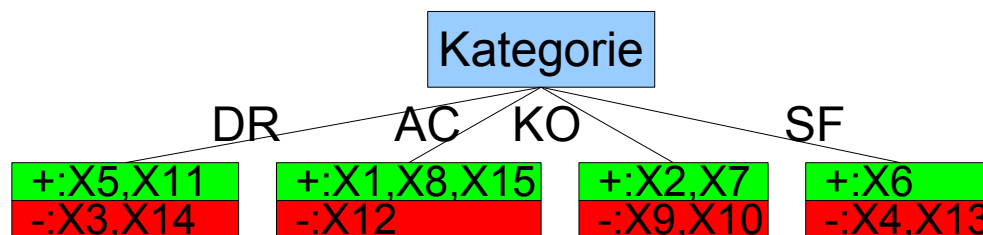
Bsp.	Attr.	€	Loge	Wetter	Warten	Bes	Kat.	Land	Res.	Gruppe	Erg.
X2	O	€	ja	O	nein	O	KO	int.	nein	Paar	ja
X5	O	€	ja	O	nein	O	DR	int.	nein	Paar	ja
X10	O	€	ja	+	nein	O	KO	int.	nein	Paar	nein
X11	+	€	ja	O	ja	+	DR	int.	nein	Paar	ja
X15	O	€	ja	-	nein	O	AC	int.	nein	Paar	ja



# Mögliche Auswahl-Attribute für den Kinobesuch - Gruppe 2



Bsp.	Attr.	€	Loge	Wetter	Warten	Bes	Kat.	Land	Res.	Gruppe	Erg.
X6	+	€€	ja	+	nein	+	SF	int.	ja	Freunde	ja
X9	-	€	ja	+	nein	O	KO	nat.	nein	Freunde	nein
X14	O	€	ja	+	ja	+	DR	int.	ja	Freunde	nein



# Algorithmus zur Baumgenerierung

```
function DT(Beispiele E, Attribute A, Klassifikation default) {  
  if (E={}) then return default  
  else if (alle Elemente in E haben gleiche Klassif. c) then return c  
  else if A={} then Fehler „gleiche Bsp. mit untersch. Klassifikation“  
  else {  
    a:= ChooseAttribute(A,E)  
    T:= neuer Entscheidungsbaum mit Wurzelmarkierung a  
    for each Attributwert wi von a do {  
      Ei:= {e ∈ E | a(e) = wi}  
      Ti:=DT(Ei,A\{a}, MajorityVal(E))  
      füge Ti an Baum T an mit Markierung wi  
    }  
  }  
  return T  
}
```

# Algorithmus zur Baumgenerierung

- **default** wird berechnet durch *MajorityValue* :  
Bestimmung des häufigsten Ergebnis für die Beispiele
- **Attributauswahl** ausgelagert in *ChooseAttribute*
  - gute Auswahl ist schwierig
  - gute Auswahl sollte einen verallgemeinerten Baum generieren  
--> Kompakte Bäume sind allgemeiner
  - **Occam's Razor** :  
*„Bevorzuge die einfachste Hypothese, die konsistent mit allen Beobachtungen ist“*

# ID3 - Entropie und Informationsgehalt

- Def. **Entropie**(=mittlerer Informationsgehalt) :

$$H(P) = - \sum_{i=1}^n p_i \log_2 p_i \quad [H] = \frac{\text{Bit}}{\text{Zeichen}}$$

für W.-keitsverteilung **P** mit den W.-keiten **p<sub>i</sub>**

- Beispiele :

- p(a)=1 :  $H(P) = -1 \cdot \log_2(1) = \underline{0}$

- p(a)=1; p(b)=0 :  $H(P) = -(1 \cdot \log_2(1) + 0 \cdot \log_2(0)) \rightarrow \underline{0}$

- p(a)=0,5; p(b)=0,5 :  $H(P) = -2 \cdot (0,5 \cdot \log_2(0,5)) = \underline{1}$

- p(a)=0,5; p(b)=p(c)=0,25

$$H(P) = -(0,5 \cdot \log_2(0,5) + 2 \cdot (0,25 \cdot \log_2(0,25))) = \underline{1,5}$$

- p(a)=p(b)=p(c)=p(d)=0,25 :

$$H(P) = -4 \cdot (0,25 \cdot \log_2(0,25)) = \underline{2}$$

# ID3 - Entropie und Informationsgehalt 2

- Bei unseren Klassifikationen ( $p$  positive und  $n$  negative Beispiele) ergibt das :

$$H\left(\frac{p}{p+n}; \frac{n}{p+n}\right) = -\left(\frac{p}{p+n} \cdot \log_2\left(\frac{p}{p+n}\right) + \frac{n}{p+n} \cdot \log_2\left(\frac{n}{p+n}\right)\right)$$

- Annahme : Alle Beispiele sind gleich wahrscheinlich, d.h. z.B.:

$$P(\text{positives Beispiel}) = \frac{p}{p+n}$$

# Entropie beim Kinobesuch

- Alle Beispiele zum Kinobesuch :

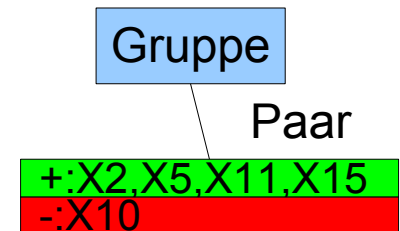
p=8, n=7 : +:X1,X2,X5,X6,X7,X8,X11,X15  
-:X3,X4,X9,X10,X12,X13,X14

$$H\left(\frac{8}{15}; \frac{7}{15}\right) = -\left(\frac{8}{15} \cdot \log_2 \frac{8}{15} + \frac{7}{15} \cdot \log_2 \frac{7}{15}\right) \text{Bit}$$

$$= -\log_2\left(\frac{8}{15} \frac{8}{15} \cdot \frac{7}{15} \frac{7}{15}\right) \text{Bit} \approx \underline{0,9968 \text{ Bit}}$$

- Für den Fall, das Gruppe=Paar gilt :

p=4, n=1 :



$$H\left(\frac{4}{5}; \frac{1}{5}\right) = -\left(\frac{4}{5} \cdot \log_2 \frac{4}{5} + \frac{1}{5} \cdot \log_2 \frac{1}{5}\right) \text{Bit} \approx \underline{0,7219 \text{ Bit}}$$

# ID3 - Bedingte mittlere Information und Informationsgewinn

- Bedingte mittlere Information :

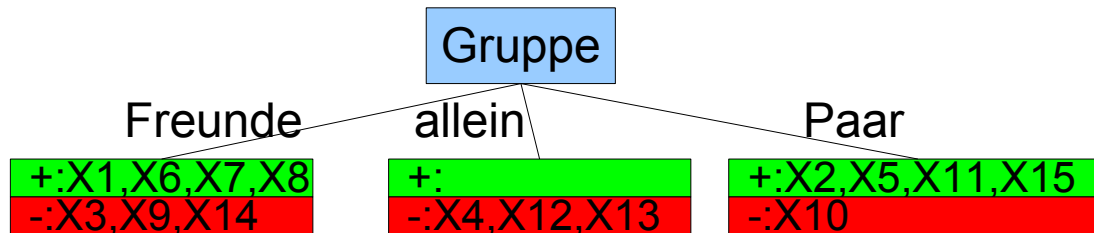
$$\begin{aligned} I(E|a \text{ bekannt}) &= \sum_{i=1}^k P(a = w_i) \cdot I(E_i) \\ &= \sum_{i=1}^k \underbrace{\frac{p_i + n_i}{p + n}}_{\text{Anteil von Wert } i} \cdot \underbrace{H\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)}_{\text{Information im Fall } i} \end{aligned}$$

- Informationsgewinn :

$$\text{gain}(a) = I(E) - I(E|a \text{ bekannt})$$

- ID3 maximiert anhand des gain (80er)

# Bedingte mittlere Information und Informationsgewinn beim Kinobesuch - Gruppe bekannt

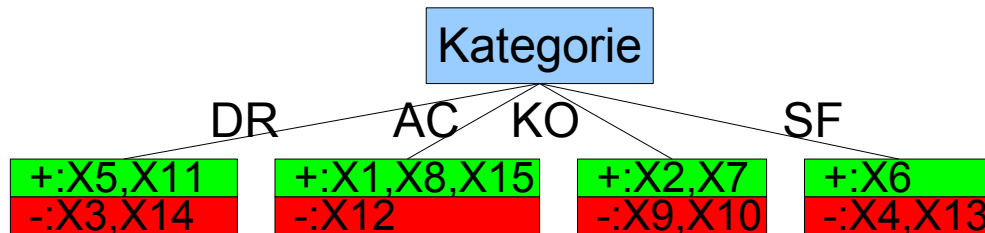


$$\begin{aligned}
 I(E/\text{Gruppe bekannt}) &= \sum_{i \in \left\{ \begin{array}{l} \text{allein,} \\ \text{Freunde,} \\ \text{Paar} \end{array} \right\}} \frac{p_i + n_i}{p + n} \cdot H\left(\frac{p_i}{p_i + n_i}; \frac{n_i}{p_i + n_i}\right) \text{ bit} \\
 &= \underbrace{\frac{3}{15} \cdot \underbrace{H(0; 1)}_{=0}}_{\text{allein}} + \underbrace{\frac{7}{15} \cdot \underbrace{H\left(\frac{4}{7}; \frac{3}{7}\right)}_{\approx 0,9852}}_{\text{Freunde}} + \underbrace{\frac{5}{15} \cdot \underbrace{H\left(\frac{4}{5}; \frac{1}{5}\right)}_{\approx 0,7219}}_{\text{Paar}} \text{ bit} \approx 0,7004 \text{ bit}
 \end{aligned}$$

$$\text{gain}(\text{Gruppe}) = I(E) - I(E/\text{Gruppe bekannt}) \approx 0,2964 \text{ bit}$$



# Bedingte mittlere Information und Informationsgewinn beim Kinobesuch - Kategorie bekannt



$$\begin{aligned}
 & I(E/\text{Kategorie bekannt}) = \\
 = & \underbrace{\frac{4}{15} \cdot H\left(\frac{3}{4}; \frac{1}{4}\right)}_{\approx 0,8113} + \underbrace{\frac{4}{15} \cdot H\left(\frac{2}{4}; \frac{2}{4}\right)}_{=1} + \underbrace{\frac{4}{15} \cdot H\left(\frac{2}{4}; \frac{2}{4}\right)}_{=1} + \underbrace{\frac{3}{15} \cdot H\left(\frac{1}{3}; \frac{2}{3}\right)}_{\approx 0,9183} \text{ bit} \\
 & \underbrace{\hspace{10em}}_{\text{Action}} \quad \underbrace{\hspace{10em}}_{\text{Komödie}} \quad \underbrace{\hspace{10em}}_{\text{Drama}} \quad \underbrace{\hspace{10em}}_{\text{Science-Fiction}} \\
 & \approx 0,9333 \text{ bit}
 \end{aligned}$$

$$\text{gain}(\text{Kategorie}) = I(E) - I(E/\text{Kategorie bekannt}) \approx 0,0634 \text{ bit}$$

# C4.5

- ID3 bevorzugt Attribute mit zahlreichen Werten, daher normalisiert C4.5 (90er) den Informationsgewinn :

$$\textit{gain-ratio}(a) = \frac{\textit{gain}(a)}{\textit{split-info}(a)}$$

$$\textit{split-info}(a) = H(a) = - \sum_{i=1}^k P(a = w_i) \cdot \log_2 P(a = w_i)$$

Fragen?