GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Einführung in die Methoden der Künstlichen Intelligenz

Maschinelles Lernen

PD Dr. David Sabel

SoSe 2014

Stand der Folien: 29. Mai 2023

Einführung

- Direkte Programmierung eines intelligenten Agenten nicht möglich (bisher)
- Daher benötigt: **Maschinelles Lernen**
- Viele Ansichten / Begriffe, was maschinelles Lernen ist
- Erfolgreichste Ansätze verwenden statistische / stochastische Methoden
- Basieren auf Adaption von Werten / Gewichten

Lernen und Agenten

Lernen soll Performanz des Agenten verbessern:

- Verbesserung der internen Repräsentation
- Optimierung bzw. Beschleunigung der Erledigung von Aufgaben.
- Erweiterung des Spektrums oder der Qualität der Aufgaben, die erledigt werden können.

Beispiele

- Anpassung / Erweiterung des Lexikons eines computerlinguistischen Systems
  Inhalt wird angepasst, aber auch gleichzeitig die Semantik
- Bewertungsfunktion im Zweipersonenspiel (Adaption der Gewichte), war für Dame und Backgammon erfolgreich
- Lernen einer Klassifikation durch Trainingsbeispiele

Struktur eines lernenden Systems

- **Agent**: (ausführende Einheit, performance element). Soll verbessert werden anhand von Erfahrung
- **Lerneinheit: (learning element)** Steuerung des Lernvorgangs. Vorgaben was schlecht ist. Bewertungseinheit (critic) und Problemgenerator
- **Umwelt**: Umwelt in der aggiert wird (künstlich oder real)

Lernmethoden

- **Überwachtes Lernen (supervised learning)**
  - Es gibt einen „allwissenden Lehrer“
  - Er sagt dem Agent, nach seiner Aktion, ob diese richtig / falsch wahr
  - unmittelbares Feedback
  - Alternative: Gebe positiv/negative Beispiele am Anfang vor

Lernmethoden

- **Überwachtes Lernen (supervised learning)**
  - Es gibt einen „allwissenden Lehrer“
  - Er sagt dem Agent, nach seiner Aktion, ob diese richtig / falsch wahr
  - unmittelbares Feedback
  - Alternative: Gebe positiv/negative Beispiele am Anfang vor
- **Unüberwachtes Lernen (unsupervised learning)**
  - Agent, weiß nicht, was richtig ist
  - Bewertung der Güte der Aktion
  - z.B. Agent misst den Effekt selbst

Lernmethoden

- **Überwachtes Lernen (supervised learning)**
  - Es gibt einen „allwissenden Lehrer"
  - Er sagt dem Agent, nach seiner Aktion, ob diese richtig / falsch wahr
  - unmittelbares Feedback
  - Alternative: Gebe positiv/negative Beispiele am Anfang vor
- **Unüberwachtes Lernen (unsupervised learning)**
  - Agent, weiß nicht, was richtig ist
  - Bewertung der Güte der Aktion
  - z.B. Agent misst den Effekt selbst
- **Lernen durch Belohnung/Bestrafung (reinforcement learning)**
  - Lernverfahren belohnen gute Aktion, bestrafen schlechte
  - D.h. Aktion ist bewertbar, aber man kennt den richtigen Parameter nicht

Lernmethoden (2)

Mögliche Vorgehensweisen:

- inkrementell,
- alle Beispiele auf einmal.

Mögliche Rahmenbedingungen:

- Beispielwerte: exakt / ungefähr (fehlerhaft)
- nur positive bzw. positive und negative Beispiele

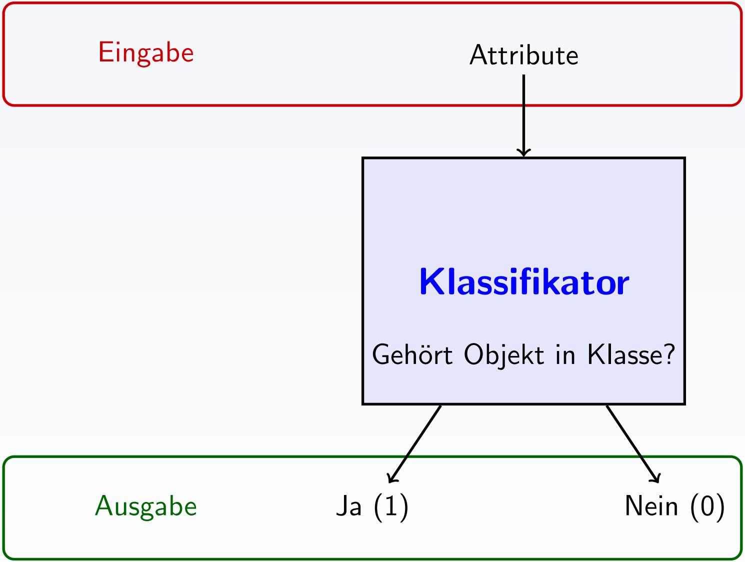
Klassifikationsverfahren

Klassifikation anhand von Eigenschaften (Attributen)

Beispiele:

- Vogel: kann-fliegen, hat-Federn, Farbe,...
- Vorhersage, ob ein Auto im kommenden Jahr einen Defekt hat: Alter, Kilometerstand, letzte Reparatur, Marke, ...
- Medizinischer Test auf Krankheit: Symptome, Blutwerte, ...
- Kreditwürdigkeit e. Bankkunden: Einkommen, Alter, Eigentumsverhältnisse, Adresse

Klassifikator



Abstrakte Situation

- Menge M von Objekten (mit innerer Struktur)
- Programm $P : M \to \{0, 1\}$
- Wahre Klassifikation $K : M \to \{0, 1\}$

Abstrakte Situation

- Menge M von Objekten (mit innerer Struktur)
- Programm $P : M \to \{0, 1\}$
- Wahre Klassifikation $K : M \to \{0, 1\}$

Eingabe: Objekt x

- Wenn $K(x) = P(x)$, dann liegt das Programm richtig
  - **richtig-positiv**: Wenn $P(x) = 1$ und $K(x) = 1$
  - **richtig-negativ**: Wenn $P(x) = 0$ und $K(x) = 0$
- Wenn $K(x) \neq P(x)$, dann liegt das Programm falsch:
  - **falsch-positiv**: Wenn $P(x) = 1$, aber $K(x) = 0$
  - **falsch-negativ**: Wenn $P(x) = 0$, aber $K(x) = 1$

Beispiel: Schwangerschaftstest

Beispieldaten: 200 durchgeführte Tests

Test ergab ...	positiv	negativ
Schwangere Frauen	59	1
Nichtschwangere Frauen	15	125

Beispiel: Schwangerschaftstest

Beispieldaten: 200 durchgeführte Tests

Test ergab ...	positiv	negativ
Schwangere Frauen	59	1
Nichtschwangere Frauen	15	125

- Richtig positiv: Frau schwanger, Test sagt schwanger

Beispiel: Schwangerschaftstest

Beispieldaten: 200 durchgeführte Tests

Test ergab ...	positiv	negativ
Schwangere Frauen	59	1
Nichtschwangere Frauen	15	125

- Richtig positiv: Frau schwanger, Test sagt schwanger
- Falsch negativ: Frau schwanger, Test sagt nicht schwanger

Beispiel: Schwangerschaftstest

Beispieldaten: 200 durchgeführte Tests

Test ergab ...	positiv	negativ
Schwangere Frauen	59	1
Nichtschwangere Frauen	15	125

- Richtig positiv: Frau schwanger, Test sagt schwanger
- Falsch negativ: Frau schwanger, Test sagt nicht schwanger
- Falsch positiv: Frau nicht schwanger, Test sagt schwanger

Beispiel: Schwangerschaftstest

Beispieldaten: 200 durchgeführte Tests

Test ergab ...	positiv	negativ
Schwangere Frauen	59	1
Nichtschwangere Frauen	15	125

- Richtig positiv: Frau schwanger, Test sagt schwanger
- Falsch negativ: Frau schwanger, Test sagt nicht schwanger
- Falsch positiv: Frau nicht schwanger, Test sagt schwanger
- Richtig negativ: Frau nicht schwanger, Test sagt nicht schwanger

Beispiel: Schwangerschaftstest

Beispieldaten: 200 durchgeführte Tests

Test ergab ...	positiv	negativ
Schwangere Frauen	59	1
Nichtschwangere Frauen	15	125

- Richtig positiv: Frau schwanger, Test sagt schwanger
- Falsch negativ: Frau schwanger, Test sagt nicht schwanger
- Falsch positiv: Frau nicht schwanger, Test sagt schwanger
- Richtig negativ: Frau nicht schwanger, Test sagt nicht schwanger

Wie gut ist der Test?

Maßzahlen

M Gesamtmenge aller zu untersuchenden Objekte:

Recall (Richtig-Positiv-Rate, hit rate)

$$\frac{|\{x \in M \mid P(x) = 1 \wedge K(x) = 1\}|}{|\{x \in M \mid K(x) = 1\}|}$$

D.h. Anteil richtig klassifizierter, positiver Objekte

Maßzahlen

M Gesamtmenge aller zu untersuchenden Objekte:

Recall (Richtig-Positiv-Rate, hit rate)

$$\frac{|\{x \in M \mid P(x) = 1 \land K(x) = 1\}|}{|\{x \in M \mid K(x) = 1\}|}$$

D.h. Anteil richtig klassifizierter, positiver Objekte

Beispiel (60 Schwangere, 59 mal positiv)

$$\frac{59}{60} \approx 98,3\%$$

Maßzahlen (2)

Richtig-Negativ-Rate, correct rejection rate

$$\frac{|\{x \in M \mid P(x) = 0 \land K(x) = 0\}|}{|\{x \in M \mid K(x) = 0\}|}$$

D.h. Anteil richtig klassifizierter, negativer Objekte

Maßzahlen (2)

Richtig-Negativ-Rate, correct rejection rate

$$\frac{|\{x \in M \mid P(x) = 0 \land K(x) = 0\}|}{|\{x \in M \mid K(x) = 0\}|}$$

D.h. Anteil richtig klassifizierter, negativer Objekte

Beispiel (140 Nicht-Schwangere, 125 mal negativ)

$$\frac{125}{140} \approx 89{,}3\%$$

Maßzahlen (3)

Precision (Präzision, positiver Vorhersagewert)

$$\frac{|\{x \in M \mid P(x) = 1 \wedge K(x) = 1\}|}{|\{x \in M \mid P(x) = 1\}|}$$

D.h. Anteil der richtigen unten den als scheinbar richtig erkannten

Maßzahlen (3)

Precision (Präzision, positiver Vorhersagewert)

$$\frac{|\{x \in M \mid P(x) = 1 \wedge K(x) = 1\}|}{|\{x \in M \mid P(x) = 1\}|}$$

D.h. Anteil der richtigen unten den als scheinbar richtig erkannten

Beispiel (59 Schwangere richtig erkannt, Test positiv: 59 + 15 = 74)

$$\frac{59}{74} \approx 79,8\%$$

Maßzahlen (4)

Negative-Vorhersage Rate

$$\frac{|\{x \in M \mid P(x) = 0 \land K(x) = 0\}|}{|\{x \in M \mid P(x) = 0\}|}$$

D.h. Anteil der richtig als falsch klassifizierten unter allen als falsch klassifizierten

Maßzahlen (4)

Negative-Vorhersage Rate

$$\frac{|\{x \in M \mid P(x) = 0 \land K(x) = 0\}|}{|\{x \in M \mid P(x) = 0\}|}$$

D.h. Anteil der richtig als falsch klassifizierten unter allen als falsch klassifizierten

Beispiel (125 Nicht-Schwangere richtig erkannt, Test negativ: 125 + 1 = 126)

$$\frac{125}{126} \approx 99,2\%$$

Maßzahlen (5)

F-Maß: Harmonisches Mittel aus Precision und Recall:

$$F = 2 \cdot \frac{(\text{precision} \cdot \text{recall})}{(\text{precision} + \text{recall})}$$

Maßzahlen (5)

F-**Maß**: Harmonisches Mittel aus Precision und Recall:

$$F = 2 \cdot \frac{(\text{precision} \cdot \text{recall})}{(\text{precision} + \text{recall})}$$

Beispiel (Precision = 79,8 % und Recall = 98,3 %)

$$F = 2 \cdot \frac{0,798 \cdot 0,983}{0,798 + 0,983} \approx 2 \cdot \frac{0,784}{1,781} \approx 0,88$$

Weitere Beispiele

Bei seltenen Krankheiten möglich:

- Guter Recall (Anteil der Kranken, die erkannt wurden),
- aber schlechte Präzision (viele false-positives)

Bsp: Klassifikator: Körpertemperatur über 38,5 C $\implies$ Gelbfieber.

- In Deutschland haben 10.000 Menschen Fieber mit 38,5 C aber nur 1 Mensch davon hat Gelbfieber
- Recall $= \frac{1}{1} = 1$
- Precision $= \frac{1}{10.000} = 0,0001$
- F-Wert ≈ 0

Fazit: Man muss immer beide Maße betrachten!

Weiteres Vorgehen

Ziel: Finde effizientes Klassifikatorprogramm

Vorher: Kurzer Exkurs zu Wahrscheinlichkeiten und zur Entropie

Exkurs: Wahrscheinlichkeiten, Entropie

Sei X ein Orakel (n-wertige Zufallsvariable)

- X liefert Wert a_i aus $\{a_1, \ldots, a_n\}$
- p_i = Wahrscheinlichkeit, dass X den Wert a_i liefert
- Folge von Orakelausgaben: $b_1, \ldots, b_m$

  Je länger die Folge: Anteil der a_i in der Folge nähert sich p_i
- $\{p_1, \ldots, p_n\}$ nennt man auch
  diskrete Wahrscheinlichkeitsverteilung
  der Menge $\{a_1, \ldots, a_n\}$ bzw. des Orakels X
- Es gilt stets $\sum_i p_i = 1$
- Sind a_i Zahlen, dann ist der
  Erwartungswert $E(X) = \sum_i p_i \cdot a_i$

Exkurs: Wahrscheinlichkeiten, Entropie (2)

Urnenmodell:

X benutzt einen Eimer mit Kugeln beschriftet mit $a_1, \dots, a_n$ und zieht bei jeder Anfrage zufällig eine Kugel (und legt sie zurück)

Dann gilt:

$$p_i = \textbf{relative Häufigkeit} \text{ von } a_i\text{-Kugeln in der Urne}$$

$$= \frac{a_i\text{-Kugeln in der Urne}}{\text{Anzahl alle Kugeln in der Urne}}$$

Exkurs: Wahrscheinlichkeiten, Entropie (3)

Gegeben: Wahrscheinlichkeitsverteilung p_i, $i = 1, \ldots, n$

Informationsgehalt des Zeichens a_k

$$I(a_k) = \log_2(\frac{1}{p_k}) = -\log_2(p_k) \geq 0$$

Interpretation:

- „Grad der Überraschung beim Ziehen des Symbols a_i“, oder auch:
- „Wie oft muss man das Orakel im Mittel fragen, um a_i zu erhalten“

D.h.

- Selten auftretenden Zeichen: haben hohe Überraschung
- Bei nur einem Zeichen:$p_1 = 1$, $I(p_1) = 0$

Exkurs: Wahrscheinlichkeiten, Entropie (4)

Entropie (Mittlerer Informationsgehalt)

$$I(X) = \sum_{i=1}^{n} p_i * I(a_i) = \sum_{i=1}^{n} p_i * \log_2(\frac{1}{p_i}) = -\sum_{i=1}^{n} p_i * \log_2(p_i) \geq 0$$

entspricht in etwa, der „mittleren Überraschung“

Beispiele

8 Objekte mit gleicher Wahrscheinlichkeit ($p_i = \frac{1}{8}$)

- Informationsgehalt jedes a_i: $\log_2(\frac{1}{\frac{1}{8}}) = \log_2 8 = 3$
- Entropie $\sum_{i=1}^{8} p_i * 3 = \sum_{i=1}^{8} \frac{1}{8} * 3 = 3$

Beispiele

8 Objekte mit gleicher Wahrscheinlichkeit ($p_i = \frac{1}{8}$)

- Informationsgehalt jedes a_i: $\log_2(\frac{1}{\frac{1}{8}}) = \log_2 8 = 3$
- Entropie $\sum_{i=1}^{8} p_i * 3 = \sum_{i=1}^{8} \frac{1}{8} * 3 = 3$

1000 Objekte mit gleicher Wahrscheinlichkeit ($p_i = \frac{1}{1000}$)

- Informationsgehalt jedes a_i: $-\log_2(1/1000) = 9.966$
- Entropie $= 9.966$

Beispiele

8 Objekte mit gleicher Wahrscheinlichkeit ($p_i = \frac{1}{8}$)

- Informationsgehalt jedes a_i: $\log_2(\frac{1}{\frac{1}{8}}) = \log_2 8 = 3$
- Entropie $\sum_{i=1}^{8} p_i * 3 = \sum_{i=1}^{8} \frac{1}{8} * 3 = 3$

1000 Objekte mit gleicher Wahrscheinlichkeit ($p_i = \frac{1}{1000}$)

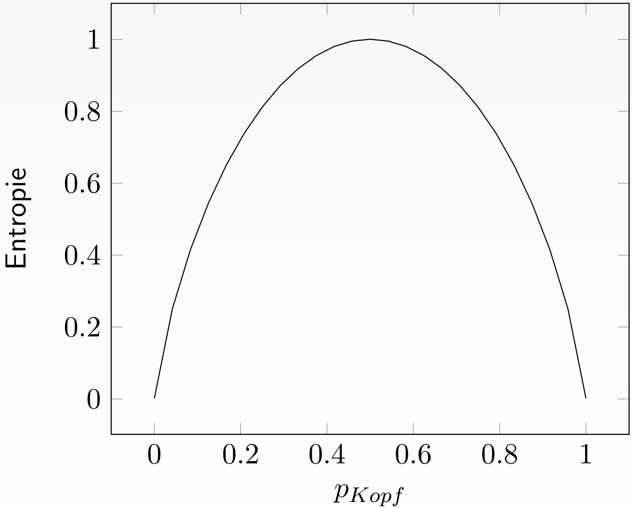
- Informationsgehalt jedes a_i: $-\log_2(1/1000) = 9.966$
- Entropie $= 9.966$

8 Objekte: $p_1 = 0.4994, p_2 = 0.4994, p_i = 0.001$ für $i = 3, \ldots, 8$

- Informationsgehalt a_1, a_2: $-\log_2(0.04994) \approx 1.002$
  a_i: $-\log_2(0.001) \approx 9.966$
- Entropie: $2 * 0.4994 * 1.002 + 6 * 0.001 * 9.996 \approx 1.061$

Beispiele

Bernoulli-Experiment: p_{Kopf} und $p_{Zahl} = 1 - p_{Kopf}$



Entscheidungsbaumlernen

Lernen von Entscheidungsbäumen (1)

Objekt mit Attributen

- Es gibt eine endliche Menge A von Attributen.
- zu jedem Attribut $a \in A$: Menge von möglichen Werten W_a. Wertebereich endlich, oder $\mathbb{R}$.
- Ein Objekt wird beschrieben durch eine Funktion $A \to \times_{a \in A} W_a$. Alternativ: Tupel mit $|A|$ Einträgen
- Ein Konzept K ist repräsentiert durch ein Prädikat P_K auf der Menge der Objekte. $P_K \subseteq$ Alle Objekte

Beispiel:

- Alle Objekte: Bücher
- Attribute: (Autor, Titel, Seitenzahl, Preis, Erscheinungsjahr).
- Konzepte „billiges Buch“ (Preis ≤ 10); „umfangreiches Buch“ (Seitenzahl ≥ 500), „altes Buch“ (Erscheinungjahr < 1950)

Entscheidungsbaum

Entscheidungsbaum zu einem Konzept K:

- endlicher Baum
- innere Knoten: Abfragen eines Attributwerts
  - Bei reellwertigen Attributen $a \leq v$ für $v \in \mathbb{R}$. 2 Kinder: Für Ja und Nein
  - Bei diskreten Attributen a mit Werten $w_1, \ldots, w_n$: n Kinder: Für jeden Wert eines
- Blätter: Markiert mit „Ja“ oder „Nein“ (manchmal auch mit „Ja oder Nein“)
- Pro Pfad: Jedes Attribut (außer rellwertige) nur einmal

Der Entscheidungsbaum gibt gerade an, ob ein Objekt zum Konzept gehört.

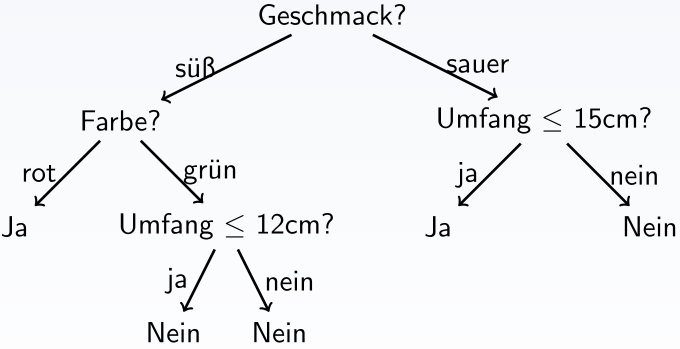
Beispiel

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Objekte: Äpfel mit Attributen:

Geschmack (süß/sauer), Farbe (rot/grün), Umfang (in cm)

Konzept: „guter Apfel“



Ist der Baum „optimal“?

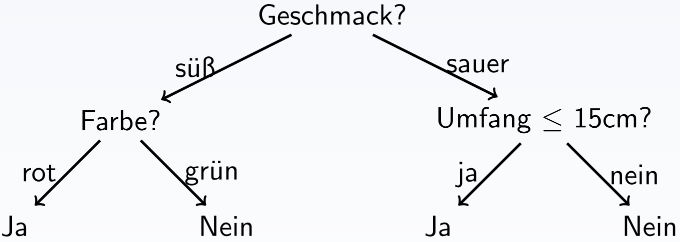
Beispiel

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Objekte: Äpfel mit Attributen:

Geschmack (süß/sauer), Farbe (rot/grün), Umfang (in cm)

Konzept: „guter Apfel“



Ist der Baum „optimal“?Nein

Entscheidungsbäume (2)

Wofür werden sie benutzt?
$\rightarrow$ als Klassifikator für Konzepte

Woher kommt der Baum?
$\rightarrow$ Durch Lernen einer Trainingsmenge

Was ist zu beachten?
$\rightarrow$ Der Baum sollte möglichst **kurze Pfade** haben
$\rightarrow$ Trainingsmenge **muss** positive **und** negative Beispiele beinhalten

Gute Entscheidungsbäume

Ein **guter Entscheidungsbaum** ist ein möglichst kleiner, d.h. der eine **möglichst kleine mittlere Anzahl** von Anfragen bis zur Entscheidung benötigt.

Wir betrachten: Algorithmen zur Konstruktion von guten Entscheidungsbäumen

Ansatz: Verwende die Entropie
(verwandt zur Konstruktion von Huffman-Bäumen)

Vorgehen: Auswahl des nächsten Attributs

Menge M von Objekten mit Attributen geteilt in positive und negative Beispiele. Annahme: Objekte sind gleichverteilt und spiegeln die Wirklichkeit wider.

Sei

- $p =$ Anzahl positiver Beispiele in M
- $n =$ Anzahl negativer Beispiele in M

Entropie der Menge M:

$$I(M) = \frac{p}{p+n} * \log_2(\frac{p+n}{p}) + \frac{n}{p+n} * \log_2(\frac{p+n}{n})$$

Vorgehen: Auswahl des nächsten Attributs (2)

Wie verändert sie die Entropie nach Wahl von Attribut a?

- Sei $m(a)$ der Wert des Attributs a des Objekts $m \in M$.
- Sei a ein Attribut mit Werten $w_1, \ldots, w_k$
- Dann zerlegt das Attribut a die Menge M in Mengen

$$M_i := \{m \in M \mid m(a) = w_i\}$$

- Seien p_i, n_i die Anzahl positiver/negativer Beispiele in M_i.
- Gewichtete Mittelwert des entstandenen Informationsgehalt nach Auswahl des Attributs a

$$I(M|a) = \sum_{i=1}^{k} P(a = w_i) * I(M_i)$$

Vorgehen: Auswahl des nächsten Attributs (2)

Wie verändert sie die Entropie nach Wahl von Attribut a?

- Sei $m(a)$ der Wert des Attributs a des Objekts $m \in M$.
- Sei a ein Attribut mit Werten $w_1, \ldots, w_k$
- Dann zerlegt das Attribut a die Menge M in Mengen

$$M_i := \{m \in M \mid m(a) = w_i\}$$

- Seien p_i, n_i die Anzahl positiver/negativer Beispiele in M_i.
- Gewichtete Mittelwert des entstandenen Informationsgehalt nach Auswahl des Attributs a

$$I(M|a) = \sum_{i=1}^{k} \underbrace{P(a = w_i)}_{\frac{p_i+n_i}{p+n}} * \underbrace{I(M_i)}_{\frac{p_i}{p_i+n_i} * \log_2(\frac{p_i+n_i}{p_i}) + \frac{n_i}{p_i+n_i} * \log_2(\frac{p_i+n_i}{n_i})}$$

Vorgehen: Auswahl des nächsten Attributs (3)

$$I(M|a) = \sum_{i=1}^{k} \frac{p_i + n_i}{p + n} * \left(\frac{p_i}{p_i + n_i} * \log_2\left(\frac{p_i + n_i}{p_i}\right) + \frac{n_i}{p_i + n_i} * \log_2\left(\frac{p_i + n_i}{n_i}\right) \right)$$

Wähle das Attribut a mit bestem **Informationsgewinn**:
$I(M) - I(M|a)$

Zur Wohldefiniertheit, setzen wir: $\frac{0}{a} * \log_2(\frac{a}{0}) := 0$

Das Verfahren ID3 (Iterative Dichotomiser 3)

Algorithmus **ID3-Verfahren**

Eingabe: Menge M von Objekten mit Attributen

Algorithmus:

Erzeuge Wurzel als offenen Knoten mit Menge M

  while es gibt offene Knoten **do**
    wähle offenen Knoten K mit Menge M
      if M enthält nur positive Beispiele **then**
        schließe K mit Markierung „Ja"
      else-if M enthält nur negative Beispiele **then**
        schließe K mit Markierung „Nein"
      else-if $M = \emptyset$ **then**
        schließe K mit Markierung „Ja" oder „Nein"
      else
        finde Attribut a mit maximalem Informationsgewinn: $I(M) - I(M|a)$
        markiere K mit a und schließe K
        erzeuge Kinder von K:
          Ein Kind pro Attributausprägung w_i von a mit Menge M_i
        Füge Kinder zu den offenen Knoten hinzu
      end-if
  end-while

Bemerkungen

- Praktische Verbesserung: Stoppe auch, wenn der Informationsgewinn zu klein
- Jedes diskrete Attribut wird nur einmal pro Pfad abgefragt, da beim zweiten Mal der Informationsgewinn 0 ist
- Wenn man eine Beispielmenge hat, die den ganzen Tupelraum abdeckt, dann wird genau das Konzept gelernt.
- Reellwertige Attribute: Leichte Anpassung möglich.

Beispiel

Äpfel: Geschmack $\in$ {süß, sauer} und Farbe $\in$ {rot, grün}.

Menge $M =$ {(süß, rot), (süß, grün), (sauer, rot), (sauer, grün)}.

Konzept: „guter Apfel“

Positiv: {(süß, rot), (süß, grün)}
Negativ: {(sauer, rot), (sauer, grün)}

$p = 2, n = 2 \Rightarrow I(M) = 0.5 \log_2 2 + 0.5 \log_2 2 = 1$

Beispiel (Forts)

Attribut Geschmack:

- $p_{\text{süß}} = 2, n_{\text{süß}} = 0$
- $p_{\text{sauer}} = 0, n_{\text{sauer}} = 2$
- $I(M|\text{Geschmack}) = \frac{2}{4} * (\frac{2}{2}\log\frac{2}{2} + \frac{0}{2}\log\frac{2}{0}) + \frac{2}{4} * (\frac{0}{2}\log\frac{2}{0} + \frac{2}{2}\log\frac{2}{2}) = 0$
- $I(M) - I(M|\text{Geschmack}) = 1$

Attribut Farbe:

- $p_{\text{rot}} = 1, n_{\text{rot}} = 1$
- $p_{\text{grün}} = 1, n_{\text{grün}} = 1$
- $I(M|\text{Farbe}) = \frac{2}{4} * (\frac{1}{2}\log\frac{2}{1} + \frac{1}{2}\log\frac{2}{1}) + \frac{2}{4} * (\frac{1}{2}\log\frac{2}{1} + \frac{1}{2}\log\frac{2}{1}) = 1$
- $I(M) - I(M|\text{Farbe}) = 0$

Beispiel

süß,rot	süß,grün	sauer,rot	sauer,grün
10	20	4	6

Ergibt:

- $I(M) = 0.8112$
- $I(M|\text{Geschmack}) = 0$
- $I(M) - I(M|\text{Geschmack}) = 0.8112$

Attribut Farbe:

- $I(M|\text{Farbe}) = 0.8086$
- $I(M) - I(M|\text{Farbe}) = 0.0026$

Grund: Die Farben sind in positiv / negativ nicht relativ gleich

Beispiel

süß,rot	süß,grün	sauer,rot	sauer,grün
10	20	3	6

- $I(M) = 0.7793$
- $I(M|\text{Geschmack}) = 0$
- $I(M) - I(M|\text{Geschmack}) = 0.7793$

Attribut Farbe:

- $I(M|\text{Farbe}) = 0.7793$
- $I(M) - I(M|\text{Farbe}) = 0$

Beispiel

Äpfel: Geschmack $\in$ {süß, sauer} und Farbe $\in$ {rot, grün}, Nr $\in \{1, \dots, 4\}$

Menge $M = \{(\text{süß}, \text{rot}, 1), (\text{süß}, \text{grün}, 2), (\text{sauer}, \text{rot}, 3), (\text{sauer}, \text{grün}, 4)\}$.

Dann:

- $I(M) = 1$
- $I(M|\text{Geschmack}) = 0$, Informationsgewinn: $I(M) - I(M|\text{Geschmack}) = 1$
- $I(M|\text{Farbe}) = 1$, Informationsgewinn: $I(M) - I(M|\text{Farbe}) = 0$
- $I(M|\text{Nr}) = 0$, Informationsgewinn: $I(M) - I(M|\text{Nr}) = 1$

Unfair: Apfelnr ist eindeutig, und stellt implizit mehr Ja/Nein Fragen dar.
Abhilfe: Weglassen der Apfelnr
Allgemein: ID3 bevorzugt Attribute mit vielen Werten
Daher: C4.5 als Anpassung von ID3

Beispiel: Konzept Apfel schmeckt wie er aussieht

Äpfel: Geschmack $\in \{$süß, sauer$\}$ und Farbe $\in \{$rot, grün, gelb$\}$

Menge $M =$ einmal jede Kombination
positiv: (rot,süß), (grün,sauer), (gelb,süß), (gelb,sauer)

- $I(M) = 0.9183$
- $I(M|\text{Farbe} = 0.6666$ und $I(M) - I(M|\text{Farbe}) = 0.2516$
- $I(M|\text{Geschmack}) = 0.9183$ und $I(M) - I(M|\text{Geschmack}) = 0$

Wähle Farbe und dann Geschmack.

Die Variante C4.5

- ID3 bevorzugt Attribute mit vielen Ausprägungen
- C4.5 ändert dies, und normiert daher den Informationsgewinn
- Algorithmus wie ID3 mit einem Unterschied:

$$\text{normierter Informationsgewinn} = (I(M) - I(M|a)) * \text{Normierungsfaktor}$$

Normierungsfaktor für Attribut a Werten $w_i, i = 1, \ldots, k$:

$$\frac{1}{\sum_{i=1}^{k} P(a = w_i) * \log_2(\frac{1}{P(a = w_i)})}$$

Beispiel

Äpfel: Geschmack $\in \{$süß, sauer$\}$ und Farbe $\in \{$rot, grün$\}$, Nr $\in \{1, \ldots, 4\}$

Menge $M = \{$(süß, rot, 1), (süß, grün, 2), (sauer, rot, 3), (sauer, grün, 4)$\}$.

Dann:

- $I(M) = 1$
- $I(M|\text{Geschmack}) = 0$, Informationsgewinn $= 1$, normiert $= 1 * 1 = 1$
- $I(M|\text{Farbe}) = 1$, Informationsgewinn $= 0$, normiert $= 0 * 1 = 0$
- $I(M|\text{Nr}) = 0$, Informationsgewinn $= 1$, normiert $= 1 * \frac{1}{2} = \frac{1}{2}$

Normierungsfaktoren:

- Geschmack: $\frac{1}{2/4 * \log_2(4/2) + 2/4 * \log_2(4/2)} = \frac{1}{1} = 1$
- Farbe: $\frac{1}{2/4 * \log_2(4/2) + 2/4 * \log_2(4/2)} = \frac{1}{1} = 1$
- Nr: $\frac{1}{1/4 * \log_2(4/1) + 1/4 * \log_2(4/1) + 1/4 * \log_2(4/1) + 1/4 * \log_2(4/1)} = \frac{1}{2}$

Übergeneralisierung

- **Effekt**: Beispiele werden eingeordnet, aber der Entscheidungsbaum unterscheidet zu fein
- **Grund**: Beispiele nicht repräsentativ bzw. ausreichend.

Beispiel: Krankheitsdiagnose: Alle positiven Beispiele sind jünger als 25 oder älter als 30
Übergeneralisierung: Alter zwischen 25 und 30 $\Longrightarrow$ nicht krank.

Übergeneralisierung (2)

Lösung: **Pruning** des Entscheidungbaums

- Stoppe Aufbau des Baums ab einer gewissen Schranke,
  da alle weiteren Attribute vermutlich irrelevant.
- Blatt-Markierung: Jenachdem welche Beispiele signifikant sind bisher
- Stoppen kann durch statistische Tests gesteuert werden

Verrauschte Daten: Gleiches Verfahren, d.h. Pruning

Versionenraum-Lernverfahren

Objekte

Was wird gelernt?

- Objekte habe mehrwertige Attributsausprägungen (nur diskrete Attribute)
- Schreibweise als Tupel $(w_1, \ldots, w_n)$ wobei $w_i =$ Attributsausprägung des Attributs A_i
- Ziel: Lerne Konzept

Beispiel: Äpfel

Attribut	Attributsausprägungen
Geschmack	süß, sauer
Farbe	rot, grün, gelb
Herkunft	Italien, Spanien, Deutschland
Größe	S, M, L

Roter, süßer Apfel aus Deutschland mittlerer Größe wird dargestellt als: (süß,rot,Deutschland,M)

Konzepte - Konzeptsprache

- Konzepte (und Hypothesen) entsprechen Mengen von Objekten
- Syntaktische Darstellung von Konzepten und Hypothesen
- $\langle w_1, \ldots, w_n \rangle$ entspricht der einelementigen Menge $\{(w_1, \ldots, w_n)\}$
- statt w_i auch erlaubt: $?$ oder $\emptyset$
- **Semantik von** $\emptyset$: Unabhängig von der Ausprägung: Objekt gehört nicht zum Konzept, d.h. $\langle w_1, \ldots, \emptyset, \ldots, w_n \rangle$ entspricht der leeren Menge
- **Semantik von** $?$: Jede Attributsausprägung ist erlaubt.

Beispiele:

- Süße Äpfel: ⟨süß,?,?,?⟩
- Rote, süße Äpfel aus Italien: ⟨süß,rot,Italien,?⟩

Enthaltensein und Subsumtion

- $e \in k$: Objekt e ist im Konzept k **enthalten**

  Beispiel: (süß,rot,Italien,M) $\in$ $\langle$süß,rot,Italien,?$\rangle$

- Es gibt eine (partielle) **Ordnung** auf Konzepten

  $k_1 \leq k_2$: wenn die zugehörigen Mengen die Teilmengenbeziehung erfüllen („$k_1 \subseteq k_2$“)

  Beispiel: $\langle$süß,rot,Italien,?$\rangle$ $\leq$ $\langle$süß,?,?,?$\rangle$

- k_1 **ist spezieller als** k_2 bzw. k_2 **ist allgemeiner als** k_1
- $<$ ist der strikte Anteil (echt spezieller / allgemeiner)

Konsistenz

Sei B eine Menge von Beispielen und h eine Hypothese.

h **ist konsistent für** B gdw. für alle $e \in B$:

- Wenn e ein positives Beispiel, dann gilt $e \in h$.
- Wenn e ein negatives Beispiel, dann gilt $e \notin h$.

Z.B. $B = \{(\text{süß,rot,Italien,M})^+, (\text{sauer,rot,Deutschland,L})^-\}$

- $h_1 = \langle\text{süß,?,?,?}\rangle$ konsistent für B
- $h_2 = \langle\text{?,?,?,M}\rangle$ konsistent für B
- $h_3 = \langle\text{?,rot,?,?}\rangle$ nicht konsistent für B
- $h_4 = \langle\text{?,?,?,?}\rangle$ nicht konsistent für B.

Versionenraum-Lernverfahren

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Ideen:

- Inkrementelles Lernverfahren:
  Beispiele werden nach und nach abgearbeitet
- Verwalte die folgende Menge (= **Versionenraum**):

> Maximale Menge aller Hypothesen, die konsistent bezüglich aller bisher verarbeiteten Beispiele ist.

- Beispiele selbst können nach ihrer Arbeitung vergessen werden.
- Alte Beispiele werden nicht nochmals gelesen.

Versionenraum-Lernverfahren

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Ideen:

- Inkrementelles Lernverfahren:
  Beispiele werden nach und nach abgearbeitet
- Verwalte die folgende Menge (= **Versionenraum**):

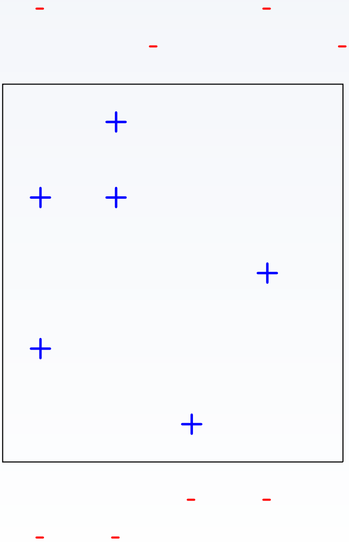
> Maximale Menge aller Hypothesen, die konsistent bezüglich aller bisher verarbeiteten Beispiele ist.

- Beispiele selbst können nach ihrer Arbeitung vergessen werden.
- Alte Beispiele werden nicht nochmals gelesen.

Probleme / Fragen

- Wie wird die Menge aller konsistenter Hypothesen dargestellt?
- Wie wird die Menge angepasst beim nächsten Beispiel?

Versionenraum



+ bisher verarbeitete positive Beispiele
- bisher verarbeitete negative Beispiele

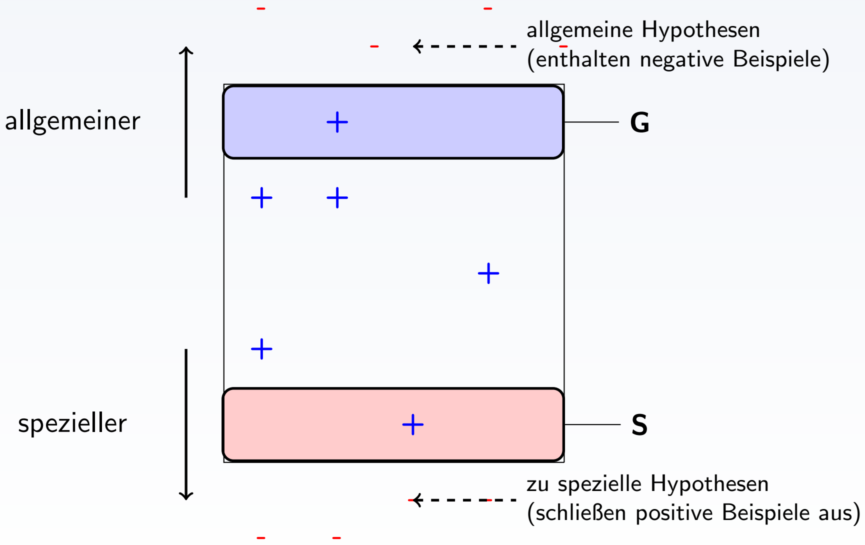
Darstellung des Versionenraums

Speichere zwei Grenzmengen S und G:

- $S =$ **untere Grenze** des Versionenraums (die **speziellsten** Hypothesen):
  Für jede Hypothese $h \in S$ gilt:
  - h ist konsistent für die Menge der bisher gesehenen Beispiele
  - es gibt kein h' mit $h' < h$, das konsistent für die Menge der bisher gesehenen Beispiele ist.
- $G =$ **obere Grenze** des Versionenraums (die **allgemeinsten** Hypothesen):
  Für jede Hypothese $h \in G$ gilt:
  - h ist konsistent für die Menge der bisher gesehenen Beispiele
  - es gibt kein h' mit $h' > h$, das konsistent für die Menge der bisher gesehenen Beispiele ist.

$$\text{Versionenraum} = \{h \mid \exists s \in S, g \in G : s \leq h \leq g\}$$

Grenzen im Bild



Iteratives Versionenraum-Lernen

Start des Algorithmus:

- Wähle die speziellsten Hypothesen für S
- Wähle die allgemeinsten Hypothesen für G
- Versionenraum: Enthält alle Hypothesen

Daher starte mit:

$$S = \{\langle\emptyset, \ldots, \emptyset\rangle\}$$

$$G = \{\langle ?, \ldots, ?\rangle\}$$

Beachte: Für jede Hypothese h gilt $\langle\emptyset, \ldots, \emptyset\rangle \leq h \leq \langle ?, \ldots, ?\rangle$

Grundgerüst des Algorithmus

Solange $G \neq \emptyset$ und $S \neq \emptyset$ und es noch Beispiele gibt:
Für neues Beispiel e:

- Passe Menge S an.
- Passe Menge G an.

Iteriere weiter.
Beachte: Wenn eine der Mengen S oder G leer wird, dann ist der Versionenraum zusammengefallen und es gibt keine konsistente Hypothese zu den Beispielen passt.

Anpassungen bei neuen Beispielen

Einfache Fälle:

- e ist **positives** Beispiel und $e \in h \in (G \cup S)$:
  keine Anpassung für h erforderlich!
- e ist **negatives** Beispiel und $e \notin h \in (G \cup S)$:
  keine Anpassung erforderlich!

Anpassungen bei neuen Beispielen (2)

Wenn e ein **positives** Beispiel ist:

- Die Hypothesen $h \in S \cup G$ mit $e \in h$ müssen nicht verändert werden.
- Für jedes $h \in G$ mit $e \notin h$: h **ist zu speziell** aber h kann nicht verallgemeinert werden (da $h \in G$). Daher: **Lösche** h aus G.
- Für jedes $h \in S$ mit $e \notin h$: h ist **zu speziell** und **muss verallgemeinert** werden. Ersetze h durch alle h' wobei
  - h' ist eine minimale Verallgemeinerung von h bzgl. e , d.h. $h' > h$, $e \in h'$ und es gibt keine Verallgemeinerung h'' von h bzgl. e mit $h' > h''$.
  - Zudem werden nur solche h' eingefügt, die nicht zu allgemein sind, d.h. es muss eine Hypothese $g \in G$ geben mit $g \geq h'$.

Schließlich (**um S kompakt zu halten**): Entferne alle Hypothesen aus S, die echte allgemeiner sind als andere Hypothesen aus S

Anpassungen bei neuen Beispielen (3)

Wenn e ein **negatives** Beispiel ist:

- Die Hypothesen $h \in S \cup G$ mit $e \notin h$ müssen nicht verändert werden.
- Für jedes $h \in S$ mit $e \in h$: h **ist zu allgemein** aber h kann nicht spezialisiert werden (da $h \in S$). Daher: **Lösche** h aus S.
- Für jedes $h \in G$ mit $e \in h$: h ist **zu allgemein** und **muss spezialisiert** werden. Ersetze h durch alle h' wobei
  - h' ist eine minimale Spezialisierung von h bzgl. e, d.h. $h' < h$, $h' < h$, $e \notin h'$ und es gibt keine Spezialisierung h'' von h bzgl. e mit $h'' > h'$.
  - Zudem werden nur solche h' eingefügt, die nicht zu speziell sind, d.h. es muss eine Hypothese $s \in S$ geben mit $s \leq h'$.

Schließlich (**um G kompakt zu halten**): Entferne alle Hypothesen aus G, die echt spezieller sind als andere Hypothesen aus G

Resultat des Algorithmus

Verschiedene Möglichkeiten:

- Wenn $S = G = \{h\}$, dann ist h das gelernte Konzept.
- Wenn $S = \emptyset$ oder $G = \emptyset$: Versionenraum ist kollabiert, es gibt keine konsistente Hypothese für die vorgegebenen Trainingsbeispiele.
- S und G sind nicht leer aber auch nicht einelementig und gleich: Der aufgespannte Versionenraum enthält das Konzept aber dieses ist nicht eindeutig identifizierbar. D.h. alle Hypothesen im Versionenraum sind konsistent für die Beispiele.

Beispiel

Beispielmenge:

(süß,rot,Italien,L)$^+$
(süß,gelb,Spanien,S)$^-$
(süß,gelb,Spanien,L)$^+$
(sauer,rot,Deutschland,L)$^-$

Starte mit

$$S = \{\langle\emptyset,\emptyset,\emptyset,\emptyset\rangle\}$$
$$G = \{\langle ?,?,?,?\rangle\}$$

Beispiel (2)

$$\begin{aligned} S &= \{\langle\emptyset,\emptyset,\emptyset,\emptyset\rangle\} \\ G &= \{\langle ?,?,?,? \rangle\} \end{aligned}$$

Aktuelles Beispiel: (süß,rot,Italien,L)$^+$

Beispiel (2)

$$S = \{\langle\emptyset,\emptyset,\emptyset,\emptyset\rangle\}$$
$$G = \{\langle ?,?,?,?\rangle\}$$

Aktuelles Beispiel: (süß,rot,Italien,L)$^+$

- Da (süß,rot,Italien,L) $\in \langle ?,?,?,?\rangle$, wird G nicht verändert.

Beispiel (2)

$$\begin{aligned} S &= \{\langle\emptyset,\emptyset,\emptyset,\emptyset\rangle\} \\ G &= \{\langle ?,?,?,?\rangle\} \end{aligned}$$

Aktuelles Beispiel: (süß,rot,Italien,L)$^+$

- Da (süß,rot,Italien,L) $\in \langle ?,?,?,?\rangle$, wird G nicht verändert.
- Da (süß,rot,Italien,L) $\not\in \langle\emptyset,\emptyset,\emptyset,\emptyset\rangle$ muss diese Hypothese verallgemeinert werden:

Beispiel (2)

$$
\begin{aligned}
S &= \{\langle\emptyset,\emptyset,\emptyset,\emptyset\rangle\}\\
G &= \{\langle ?,?,?,?\rangle\}
\end{aligned}
$$

Aktuelles Beispiel: (süß,rot,Italien,L)$^+$

- Da (süß,rot,Italien,L) $\in \langle ?,?,?,?\rangle$, wird G nicht verändert.
- Da (süß,rot,Italien,L) $\not\in \langle\emptyset,\emptyset,\emptyset,\emptyset\rangle$ muss diese Hypothese verallgemeinert werden:
  Minimale Verallgemeinerung $\langle$süß,rot,Italien,L$\rangle$.
  D.h. $S = \{\langle$süß,rot,Italien,L$\rangle\}$.

Beispiel (3)

$$\begin{aligned} S &= \{\langle \text{süß,rot,Italien,L} \rangle\}. \\ G &= \{\langle ?, ?, ?, ? \rangle\} \end{aligned}$$

Aktuelles Beispiel: (süß,gelb,Spanien,S)$^{-}$

Beispiel (3)

$$S = \{\langle \text{süß,rot,Italien,L} \rangle\}.$$
$$G = \{\langle ?, ?, ?, ? \rangle\}$$

Aktuelles Beispiel: (süß,gelb,Spanien,S)$^-$

- Da (süß,gelb,Spanien,S) $\not\in$ ⟨süß,rot,Italien,L⟩, wird S nicht verändert.

Beispiel (3)

$S = \{\langle \text{süß,rot,Italien,L} \rangle\}$.
$G = \{\langle ?, ?, ?, ? \rangle\}$

Aktuelles Beispiel: $(\text{süß,gelb,Spanien,S})^-$

- Da (süß,gelb,Spanien,S) $\notin \langle$süß,rot,Italien,L$\rangle$, wird S nicht verändert.
- Da (süß,gelb,Spanien,S) $\in \langle ?, ?, ?, ? \rangle$, muss diese Hypothese spezialisiert werden.

  Minimale Spezialisierungen:
  $\langle$sauer$, ?, ?, ? \rangle$ $\langle ?,$rot$, ?, ? \rangle$ $\langle ?,$gruen$, ?, ? \rangle$ $\langle ?, ?,$Italien$, ? \rangle$
  $\langle ?, ?,$Deutschland$, ? \rangle$ $\langle ?, ?, ?,$L$\rangle$ $\langle ?, ?, ?,$M$\rangle$

Beispiel (3)

$$S = \{\langle\text{süß,rot,Italien,L}\rangle\}.$$
$$G = \{\langle ?,?,?,? \rangle\}$$

Aktuelles Beispiel: (süß,gelb,Spanien,S)$^-$

- Da (süß,gelb,Spanien,S) $\not\in$ ⟨süß,rot,Italien,L⟩, wird S nicht verändert.
- Da (süß,gelb,Spanien,S) $\in \langle ?,?,?,? \rangle$, muss diese Hypothese spezialisiert werden.

  Minimale Spezialisierungen:
  ⟨sauer,?,?,?⟩ ⟨?,rot,?,?⟩ ⟨?,gruen,?,?⟩ ⟨?,?,Italien,?⟩
  ⟨?,?,Deutschland,?⟩ ⟨?,?,?,L⟩ ⟨?,?,?,M⟩

  Kompaktifizierung:
  Nehme nur solche h mit: Es gibt $h' \in S$ und $h \geq h'$:
  Das erfüllen nur ⟨?,rot,?,?⟩ ⟨?,?,Italien,?⟩ ⟨?,?,?,L⟩
  Daher: $G = \{\langle ?,\text{rot},?,? \rangle, \langle ?,?,\text{Italien},? \rangle, \langle ?,?,?,\text{L} \rangle\}$.

Beispiel (4)

$$S = \{\langle \text{süß,rot,Italien,L} \rangle\}.$$
$$G = \{\langle ?,\text{rot},?,? \rangle, \langle ?,?,\text{Italien},? \rangle, \langle ?,?,?,\text{L} \rangle\}.$$

Aktuelles Beispiel: (süß,gelb,Spanien,L)$^{+}$

Beispiel (4)

$S \quad = \{\langle \text{süß,rot,Italien,L} \rangle\}.$
$G \quad = \{\langle ?,\text{rot},?,? \rangle, \langle ?,?,\text{Italien},? \rangle, \langle ?,?,?,\text{L} \rangle\}.$

Aktuelles Beispiel: (süß,gelb,Spanien,L)$^+$

- Da (süß,gelb,Spanien,L) $\not\in \langle ?,\text{rot},?,? \rangle \in G$ muss die Hypothese gelöscht werden (da G die allgemeinsten Hypothesen enthält, darf nicht verallgemeinert werden).

Beispiel (4)

$$S = \{\langle \text{süß}, \text{rot}, \text{Italien}, \text{L}\rangle\}.$$
$$G = \{\langle ?, \text{rot}, ?, ?\rangle, \langle ?, ?, \text{Italien}, ?\rangle, \langle ?, ?, ?, \text{L}\rangle\}.$$

Aktuelles Beispiel: $(\text{süß}, \text{gelb}, \text{Spanien}, \text{L})^{+}$

- Da $(\text{süß}, \text{gelb}, \text{Spanien}, \text{L}) \not\in \langle ?, \text{rot}, ?, ?\rangle \in G$ muss die Hypothese gelöscht werden (da G die allgemeinsten Hypothesen enthält, darf nicht verallgemeinert werden).
- Da $(\text{süß}, \text{gelb}, \text{Spanien}, \text{L}) \not\in \langle ?, ?, \text{Italien}, ?\rangle \in G$ muss die Hypothese gelöscht werden

Beispiel (4)

$S \quad = \{\langle\text{süß,rot,Italien,L}\rangle\}.$
$G \quad = \{\langle ?,\text{rot}, ?, ?\rangle, \langle ?, ?,\text{Italien}, ?\rangle, \langle ?, ?, ?,\text{L}\rangle\}.$

Aktuelles Beispiel: (süß,gelb,Spanien,L)$^+$

- Da (süß,gelb,Spanien,L) $\not\in \langle ?,\text{rot}, ?, ?\rangle \in G$ muss die Hypothese gelöscht werden (da G die allgemeinsten Hypothesen enthält, darf nicht verallgemeinert werden).
- Da (süß,gelb,Spanien,L) $\not\in \langle ?, ?,\text{Italien}, ?\rangle \in G$ muss die Hypothese gelöscht werden
- Da (süß,gelb,Spanien,L) $\in \langle ?, ?, ?,\text{L}\rangle \in G$ bleibt Hypothese in G.

Beispiel (4)

$$S = \{\langle \text{süß},\text{rot},\text{Italien},\text{L}\rangle\}.$$
$$G = \{\langle ?,\text{rot},?,?\rangle, \langle ?,?,\text{Italien},?\rangle, \langle ?,?,?,\text{L}\rangle\}.$$

Aktuelles Beispiel: (süß,gelb,Spanien,L)$^{+}$

- Da (süß,gelb,Spanien,L) $\not\in \langle ?,\text{rot},?,?\rangle \in G$ muss die Hypothese gelöscht werden (da G die allgemeinsten Hypothesen enthält, darf nicht verallgemeinert werden).
- Da (süß,gelb,Spanien,L) $\not\in \langle ?,?,\text{Italien},?\rangle \in G$ muss die Hypothese gelöscht werden
- Da (süß,gelb,Spanien,L) $\in \langle ?,?,?,\text{L}\rangle \in G$ bleibt Hypothese in G.
- Ergibt $G = \{\langle ?,?,?,\text{L}\rangle\}$

Beispiel (4)

$$S = \{\langle \text{süß,rot,Italien,L} \rangle\}.$$
$$G = \{\langle ?,\text{rot},?,? \rangle, \langle ?,?,\text{Italien},? \rangle, \langle ?,?,?,\text{L} \rangle\}.$$

Aktuelles Beispiel: (süß,gelb,Spanien,L)$^+$

- Da (süß,gelb,Spanien,L) $\not\in \langle ?,\text{rot},?,? \rangle \in G$ muss die Hypothese gelöscht werden (da G die allgemeinsten Hypothesen enthält, darf nicht verallgemeinert werden).
- Da (süß,gelb,Spanien,L) $\not\in \langle ?,?,\text{Italien},? \rangle \in G$ muss die Hypothese gelöscht werden
- Da (süß,gelb,Spanien,L) $\in \langle ?,?,?,\text{L} \rangle \in G$ bleibt Hypothese in G.
- Ergibt $G = \{\langle ?,?,?,\text{L} \rangle\}$
- Da (süß,gelb,Spanien,L) $\not\in \langle \text{süß,rot,Italien,L} \rangle \in S$ muss die Hypothese verallgemeinert werden.

Beispiel (4)

$$S = \{\langle \text{süß,rot,Italien,L} \rangle\}.$$
$$G = \{\langle ?,\text{rot},?,? \rangle, \langle ?,?,\text{Italien},? \rangle, \langle ?,?,?,\text{L} \rangle\}.$$

Aktuelles Beispiel: (süß,gelb,Spanien,L)$^+$

- Da (süß,gelb,Spanien,L) $\notin \langle ?,\text{rot},?,? \rangle \in G$ muss die Hypothese gelöscht werden (da G die allgemeinsten Hypothesen enthält, darf nicht verallgemeinert werden).
- Da (süß,gelb,Spanien,L) $\notin \langle ?,?,\text{Italien},? \rangle \in G$ muss die Hypothese gelöscht werden
- Da (süß,gelb,Spanien,L) $\in \langle ?,?,?,\text{L} \rangle \in G$ bleibt Hypothese in G.
- Ergibt $G = \{\langle ?,?,?,\text{L} \rangle\}$
- Da (süß,gelb,Spanien,L) $\notin \langle \text{süß,rot,Italien,L} \rangle \in S$ muss die Hypothese verallgemeinert werden.
  Minimale Verallgemeinerung: $\langle$süß,?,?,L$\rangle$.
  Kompaktifizierung ändert nichts. Daher $S = \{\langle \text{süß,?,?,L} \rangle\}$

Beispiel (5)

$$\begin{aligned} S &= \{\langle \text{süß},?,?,\text{L}\rangle\} \\ G &= \{\langle ?, ?, ?,\text{L}\rangle\}. \end{aligned}$$

Aktuelles Beispiel: $(\text{sauer},\text{rot},\text{Deutschland},\text{L})^-$

Beispiel (5)

$$S = \{\langle \text{süß,?,?,L} \rangle\}$$
$$G = \{\langle ?, ?, ?, \text{L} \rangle\}.$$

Aktuelles Beispiel: (sauer,rot,Deutschland,L)$^-$

- Da (sauer,rot,Deutschland,L) $\not\in$ $\langle$süß,?,?,L$\rangle$, bleibt S unverändert.

Beispiel (5)

$$S = \{\langle \text{süß},?,?,\text{L}\rangle\}$$
$$G = \{\langle ?,?,?,\text{L}\rangle\}.$$

Aktuelles Beispiel: $(\text{sauer},\text{rot},\text{Deutschland},\text{L})^-$

- Da $(\text{sauer},\text{rot},\text{Deutschland},\text{L}) \not\in \langle \text{süß},?,?,\text{L}\rangle$, bleibt S unverändert.
- Da $(\text{sauer},\text{rot},\text{Deutschland},\text{L}) \in \langle ?,?,?,\text{L}\rangle$, muss die Hypothese spezialisiert werden.

Beispiel (5)

$$S = \{\langle \text{süß,?,?,L} \rangle\}$$
$$G = \{\langle ?, ?, ?, \text{L} \rangle\}.$$

Aktuelles Beispiel: $(\text{sauer,rot,Deutschland,L})^-$

- Da $(\text{sauer,rot,Deutschland,L}) \not\in \langle \text{süß,?,?,L} \rangle$, bleibt S unverändert.
- Da $(\text{sauer,rot,Deutschland,L}) \in \langle ?, ?, ?, \text{L} \rangle$, muss die Hypothese spezialisiert werden.
  Minimale Spezialiserungen:
  $\langle$süß,?,?,L$\rangle$ $\langle$?,gelb,?,L$\rangle$ $\langle$?,grün,?,L$\rangle$
  $\langle$?,?,Italien,L$\rangle$ $\langle$?,?,Spanien,L$\rangle$

Beispiel (5)

$$
\begin{aligned}
S &= \{\langle \text{süß,?,?,L} \rangle\} \\
G &= \{\langle ?,?,?,\text{L} \rangle\}.
\end{aligned}
$$

Aktuelles Beispiel: $(\text{sauer,rot,Deutschland,L})^-$

- Da (sauer,rot,Deutschland,L) $\not\in$ ⟨süß,?,?,L⟩, bleibt S unverändert.
- Da (sauer,rot,Deutschland,L) $\in \langle ?,?,?,\text{L} \rangle$, muss die Hypothese spezialisiert werden.
  Minimale Spezialiserungen:
  ⟨süß,?,?,L⟩ ⟨?,gelb,?,L⟩ ⟨?,grün,?,L⟩
  ⟨?,?,Italien,L⟩ ⟨?,?,Spanien,L⟩

  Kompaktifizierung: Nur ⟨süß,?,?,L⟩ bleibt übrig

  Daher: $G = \langle \text{süß,?,?,L} \rangle$

Beispiel (6)

$$\begin{aligned} S &= \{\langle \text{süß,?,?,L} \rangle\} \\ G &= \{\langle \text{süß,?,?,L} \rangle\} \end{aligned}$$

Keine Beispiele mehr, und $S = G = \{\langle \text{süß,?,?,L} \rangle\}$.

Daher ist $\langle \text{süß,?,?,L} \rangle$ das gelernte Konzept.

Noch ein Beispiel

Beispielmenge:

(süß,rot,Italien,L)$^+$
(süß,gelb,Italien,L)$^-$

Starte mit

$$\begin{aligned} S &= \{\langle\emptyset,\emptyset,\emptyset,\emptyset\rangle\} \\ G &= \{\langle ?,?,?,? \rangle\} \end{aligned}$$

Noch ein Beispiel (2)

$$
\begin{aligned}
S &= \{\langle\emptyset,\emptyset,\emptyset,\emptyset\rangle\} \\
G &= \{\langle ?,?,?,?\rangle\}
\end{aligned}
$$

Aktuelles Beispiel: (süß,rot,Italien,L)$^+$

Noch ein Beispiel (2)

$$\begin{aligned} S &= \{\langle\emptyset,\emptyset,\emptyset,\emptyset\rangle\} \\ G &= \{\langle ?,?,?,?\rangle\} \end{aligned}$$

Aktuelles Beispiel: (süß,rot,Italien,L)$^+$

- Da (süß,rot,Italien,L) $\in \langle ?,?,?,?\rangle$, wird G nicht verändert.

Noch ein Beispiel (2)

$$\begin{aligned} S &= \{\langle\emptyset,\emptyset,\emptyset,\emptyset\rangle\} \\ G &= \{\langle ?,?,?,?\rangle\} \end{aligned}$$

Aktuelles Beispiel: (süß,rot,Italien,L)$^+$

- Da (süß,rot,Italien,L) $\in \langle ?,?,?,?\rangle$, wird G nicht verändert.
- Da (süß,rot,Italien,L) $\not\in \langle\emptyset,\emptyset,\emptyset,\emptyset\rangle$ muss diese Hypothese verallgemeinert werden:

Noch ein Beispiel (2)

$$S = \{\langle\emptyset,\emptyset,\emptyset,\emptyset\rangle\}$$
$$G = \{\langle ?,?,?,?\rangle\}$$

Aktuelles Beispiel: (süß,rot,Italien,L)$^+$

- Da (süß,rot,Italien,L) $\in \langle ?,?,?,?\rangle$, wird G nicht verändert.
- Da (süß,rot,Italien,L) $\not\in \langle\emptyset,\emptyset,\emptyset,\emptyset\rangle$ muss diese Hypothese verallgemeinert werden:
  Minimale Verallgemeinerung ⟨süß,rot,Italien,L⟩.
  D.h. $S = \{\langle$süß,rot,Italien,L$\rangle\}$.

Noch ein Beispiel (3)

$$S = \{\langle \text{süß}, \text{rot}, \text{Italien}, \text{L}\rangle\}.$$
$$G = \{\langle ?, ?, ?, ?\rangle\}$$

Aktuelles Beispiel: $(\text{süß}, \text{gelb}, \text{Italien}, \text{L})^-$

- $(\text{süß}, \text{gelb}, \text{Italien}, \text{L}) \not\in \langle \text{süß}, \text{rot}, \text{Italien}, \text{L}\rangle$. S wird nicht verändert
- $(\text{süß}, \text{gelb}, \text{Italien}, \text{L}) \in \langle ?, ?, ?, ?\rangle$, Hypothese wird spezialisiert:
  Minimale Spezialisierungen
  $\langle \text{sauer}, ?, ?, ?\rangle$ $\langle ?, \text{grün}, ?, ?\rangle$ $\langle ?, \text{rot}, ?, ?\rangle$ $\langle ?, ?, \text{Deutschland}, ?\rangle$
  $\langle ?, ?, \text{Spanien}, ?\rangle$ $\langle ?, ?, ?, \text{S}\rangle$ $\langle ?, ?, ?, \text{M}\rangle$
  Kompaktifizerung: Es verbleibt nur $\langle ?, \text{rot}, ?, ?\rangle$
  D.h. $G = \{\langle ?, \text{rot}, ?, ?\rangle\}$.

Noch ein Beispiel (4)

$$S \quad = \{\langle \text{süß,rot,Italien,L} \rangle\}.$$
$$G \quad = \{\langle ?,\text{rot}, ?, ? \rangle\}$$

Keine weiteren Beispiele: Alle Hypothesen h mit

$$\langle \text{süß,rot,Italien,L} \rangle \leq h \leq \langle ?,\text{rot}, ?, ? \rangle$$

sind konsistent.

Weiteres Beispiel

Wie vorher aber mit drittem Beispiel:
Beispielmenge:

(süß,rot,Italien,L)$^{+}$
(süß,gelb,Italien,L)$^{-}$
(süß,grün,Italien,L)$^{+}$

Beim letzten Beispiel

$S = \{\langle \text{süß,rot,Italien,L} \rangle\}.$
$G = \{\langle ?,\text{rot}, ?, ? \rangle\}$

Aktuelles Beispiel: $(\text{süß,grün,Italien,L})^+$

- $(\text{süß,grün,Italien,L}) \not\in \langle ?,\text{rot}, ?, ? \rangle$.
  Da $\langle ?,\text{rot}, ?, ? \rangle$ nicht verallgemeinert werden darf, wird die Hypothese gelöscht.
  D.h. $G = \emptyset$
- $(\text{süß,grün,Italien,L}) \not\in \langle \text{süß,rot,Italien,L} \rangle$.
  Verallgemeinere zu $\langle \text{süß,?,Italien,L} \rangle$.
  Kompaktifizierung: Löscht alle Hypothesen in S, da es keine in G gibt.
  D.h. $S = \emptyset$.
- $G = S = \emptyset$: Es gibt keine konsistente Hypothese.

Nachteile

- Verrauschte Daten können nicht adäquat behandelt werden
- Bei verrauschten Daten findet die Methode nie eine konsistente Hypothese
- Abhilfe: Stärke Konzeptsprachen
- Problem dabei: Overfitting

Erweiterungen

- Erlaube „M_a“ für Attribute, wobei M_a eine Teilmenge der möglichen Ausprägungen von a ist ist.
- Erlaube Disjunktionen dieser Konzepte. Damit kann man bereits alle Konzepte darstellen, wenn die Menge der Attribute und Ausprägungen endlich ist.