



OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

INF

FAKULTÄT FÜR
INFORMATIK

Intelligente Systeme

Unsicheres Wissen

Prof. Dr. R. Kruse C. Moewes G. Ruß

{kruse,cmoewes,russ}@iws.cs.uni-magdeburg.de

Institut für Wissens- und Sprachverarbeitung

Fakultät für Informatik

Otto-von-Guericke Universität Magdeburg

Übersicht

1. Motivation

2. Grundlagen der Wahrscheinlichkeitstheorie

3. Probabilistische Inferenz

4. Bayes'sche Netze

Motivation

In vielen Fällen ist unser Wissen über die Welt

- ▶ unvollständig: **Wissensumfang ist nicht hinreichend**
z.B.: „Ich weiß nicht, wie der Bus an Feiertagen fährt, da ich nur werktags fahre!“
- ▶ unscharf: **Wissen ist nicht präzise genug**
z.B.: „Der Bus fährt ungefähr zu jeder vollen Stunde!“
- ▶ unsicher: **Wissen ist unzuverlässig**
z.B.: „Der Bus fährt wahrscheinlich jede volle Stunde.“

Wir müssen trotzdem agieren!

- ▶ Schließen unter Unsicherheit
- ▶ Schließen über Eintretenswahrscheinlichkeiten
- ▶ ... und Kosten/Nutzen

Beispiel

- ▶ **Ziel:** um 9:15 Uhr in Uni sein, um an Vorlesung teilzunehmen
- ▶ mehrere **Pläne**, um Ziel zu erreichen:
 - ▶ P_1 : 8:00 Uhr aufstehen, 8:55 Uhr Haus verlassen, 9:00 Uhr Bus...
 - ▶ P_2 : 7:30 Uhr aufstehen, 8:25 Uhr Haus verlassen, 8:30 Uhr Bus...
 - ▶ ...
- ▶ alle Pläne sind *korrekt*, aber
 - ▶ sie implizieren verschiedene **Kosten** und verschiedene **Wahrscheinlichkeiten** Ziel *tatsächlich* zu erreichen
 - ▶ P_2 am besten, denn Vorlesung ist wichtig, und Erfolgsrate bei P_1 liegt nur bei ca. 85–95%

Grade der Überzeugung

- ▶ wir (oder andere Agenten) sind von Regeln und Fakten nur bis zu gewissem Grad überzeugt
- ▶ Möglichkeit, **Grad der Überzeugung auszudrücken**: Benutzung von **Wahrscheinlichkeiten**
- ▶ „Agent ist von Sensorinformation zu 0,9 überzeugt.“ heißt: Agent glaubt in 9 von 10 Fällen an Richtigkeit der Information
- ▶ W'keiten fassen „Unsicherheit“ bedingt durch Unwissen zusammen
- ▶ nicht mit *Unschärfe (Vagheit)* verwechseln (→ Fuzzy-Logik)
- ▶ Prädikat *groß* ist **unscharf**
Aussage „*Das könnte Peters Uhr sein.*“ ist **unsicher**

Rationale Entscheidungen unter Unsicherheit

- ▶ verschiedene *Aktionen* oder *Pläne* zur Auswahl
- ▶ können mit verschiedenen *W'keiten* zu verschiedenen Ergebnissen führen
- ▶ *Aktionen* verursachen verschiedene – ggf. subjektive – **Kosten**
- ▶ *Ergebnisse* haben verschiedenen – ggf. subjektiven – **Nutzen**
- ▶ rational: wähle Aktion, die größten **zu erwartenden Gesamtnutzen** hat

Entscheidungstheorie = Nutzentheorie + W'keitstheorie

Entscheidungstheoretischer Agent

Algorithmus 1 DT-Agent

Eingabe: *a percept*

Ausgabe: *an action*

- 1: static: a set of probabilistic beliefs about the state of the world
 - 2: calculate updated probabilities for current state based on available evidence including current percept and previous action
 - 3: calculate outcome probabilities for actions, given action descriptions and probabilities of current states
 - 4: select *action* with highest expected utility, given probabilities of outcomes and utility information
 - 5: **return** *action*
-

Entscheidungstheorie: Agent ist rational genau dann, wenn er die Aktion wählt, die größten zu erwartenden Nutzen gemittelt über alle möglichen Ergebnisse von Aktionen hat

Übersicht

1. Motivation

2. Grundlagen der Wahrscheinlichkeitstheorie

3. Probabilistische Inferenz

4. Bayes'sche Netze

Unbedingte Wahrscheinlichkeiten I

- ▶ $P(A)$ bezeichnet **unbedingte** oder **a priori** Wahrscheinlichkeit, dass Ereignis A eintritt, falls *keine* zusätzliche Information verfügbar, z.B.

$$P(\text{Karies}) = 0.1$$

- ▶ *Karies* ist Proposition (Aussage)
- ▶ Gewinnung von a priori W'keiten: statistische Analyse oder aus allgemeinen Regeln

Axiomatische Wahrscheinlichkeitstheorie

Funktion P von aussagenlogischen Formeln in Menge $[0, 1]$ ist W'keitsmaß, falls für alle Aussagen A, B gilt:

1. $0 \leq P(A) \leq 1$,
2. $P(\text{True}) = 1$,
3. $P(\text{False}) = 0$,
4. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$.

alle anderen Eigenschaften lassen sich aus diesen Axiomen ableiten,
z.B.

$$P(\neg A) = 1 - P(A) \text{ folgt aus}$$

$$P(A \vee \neg A) = 1 \text{ und}$$

$$P(A \wedge \neg A) = 0.$$

Wieso sind die Axiome sinnvoll?

- ▶ wenn P objektiv beobachtbare W' keit bezeichnet, machen Axiome natürlich Sinn
- ▶ aber wieso sollte Agent Axiome beachten, wenn er *Grad seiner Überzeugung* modelliert?
- ▶ *objektive vs. subjektive W' keiten*
- ▶ Axiome schränken Menge der Überzeugungen ein, die Agent aufrechterhalten kann
- ▶ eines der überzeugendsten Argumente, warum subjektive Überzeugungen Axiome respektieren sollten: 1931 von de Finetti: basiert auf Zusammenhang zwischen Aktionen und Grad der Überzeugung.
sind Überzeugungen widersprüchlich, dann wird Agent auf lange Sicht in seiner Umwelt scheitern!

Bedingte Wahrscheinlichkeiten I

- ▶ neue Information kann W'keit eines Ereignisses ändern
- ▶ Beispiel: Wahrscheinlichkeit von Zahnlöchern erhöht sich, wenn man weiß, dass Patient Zahnschmerzen hat
- ▶ bei Zusatzinformation: nicht mehr mit a priori W'keiten rechnen!
- ▶ $P(A | B)$ bezeichnet **bedingte** oder **a posteriori** W'keit von A , sofern alleinige Beobachtung (Evidenz) B gegeben:

$$P(\text{Karies} | \text{Zahn}) = 0.8$$

- ▶ $P(X | Y)$: Tabelle aller bedingten W'keiten über alle Werte von X und Y

Bedingte Wahrscheinlichkeiten II

- ▶ $P(\text{Wetter} \mid \text{Kopfschmerzen})$: (4×2) -Tabelle von bedingten W 'keiten aller Kombinationen der Werte einer Zufallsvariablen

	Kopfs. = TRUE	Kopfs. = FALSE
Wetter = sonnig	$P(W = \text{sonnig} \mid \text{Kopfs.})$	$P(W = \text{sonnig} \mid \neg\text{Kopfs.})$
Wetter = Regen		
Wetter = bewölkt		
Wetter = Schnee		

- ▶ bedingte W 'keiten ergeben sich aus unbedingten W 'keiten **per Definition** (für $P(B) > 0$):

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

Bedingte Wahrscheinlichkeiten III

- $P(X, Y) = P(X | Y)P(Y)$ entspricht einem Gleichungssystem:

$$P(W = \text{sonnig} \wedge \text{Kopfs.}) = P(W = \text{sonnig} | \text{Kopfs.})P(\text{Kopfs.})$$

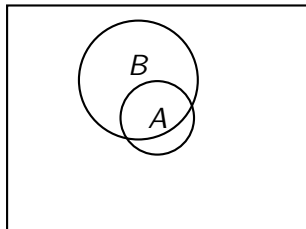
$$P(W = \text{Regen} \wedge \text{Kopfs.}) = P(W = \text{Regen} | \text{Kopfs.})P(\text{Kopfs.})$$

$$\vdots = \vdots$$

$$P(W = \text{Schnee} \wedge \neg\text{Kopfs.}) = P(W = \text{Schnee} | \neg\text{Kopfs.})P(\neg\text{Kopfs.})$$

Bedingte Wahrscheinlichkeiten IV

$$P(A | B) = \frac{P(A \wedge B)}{P(B)}$$



- ▶ Produktregel: $P(A \wedge B) = P(A | B)P(B)$
- ▶ analog: $P(A \wedge B) = P(B | A)P(A)$
- ▶ A und B heißen **unabhängig** voneinander, falls $P(A | B) = P(A)$ und $P(B | A) = P(B)$
- ▶ dann und nur dann gilt $P(A \wedge B) = P(A)P(B)$

Interpretation bedingter Wahrscheinlichkeiten

Achtung: häufiger Interpretationsfehler

„ $P(A | B) = 0.8$ bedeutet, dass $P(A) = 0.8$ ist
falls B wahr ist“

Diese Aussage ist aus mindestens zwei Gründen falsch!

1. $P(A)$ ist immer die a priori W'keit, nicht die a posteriori W'keit, gegeben eine Evidenz (B wahr)
2. $P(A | B) = 0.8$ ist nur dann anwendbar, falls keine andere Evidenz außer B vorhanden ist! Wenn C bekannt ist, dann muss $P(A | B \wedge C)$ berechnet oder geschätzt werden. I.A. gilt $P(A | B \wedge C) \neq P(A | B)$; z.B. könnte $C \rightarrow A$ gelten (C impliziert A)

Unbedingte Wahrscheinlichkeiten II

- ▶ i.A. Zufallsvariable kann nicht nur Werte wahr und falsch, sondern mehrere Werte annehmen

$$P(\text{Wetter} = \text{sonnig}) = 0.7$$

$$P(\text{Wetter} = \text{Regen}) = 0.2$$

$$P(\text{Wetter} = \text{bewölkt}) = 0.08$$

$$P(\text{Wetter} = \text{Schnee}) = 0.02$$

$$P(\text{Kopfs.} = \text{wahr}) = 0.1$$

- ▶ Zufallsvariable Wetter mit $P(\text{Wetter} = \text{sonnig})$ für $P(\{\omega \mid \text{Wetter}(\omega) = \text{sonnig}\})$
- ▶ Propositionen auch mit Gleichungen über Zufallsvariablen
- ▶ logische Konnektoren zur Bildung von Propositionen, z.B.

$$P(\text{Karies} \wedge \neg \text{Schmerzen}) = 0.06$$

bedeutet $P(\{\omega \mid \text{Karies}(\omega) = \text{wahr} \wedge \text{Schmerzen}(\omega) = \text{falsch}\})$

Unbedingte Wahrscheinlichkeiten III

- ▶ $P(X)$: Vektor der W 'keiten für (geordneten) Wertebereich der Zufallsvariable X :

$$P(\text{Kopfs.}) = \langle 0.1, 0.9 \rangle$$

$$P(\text{Wetter}) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$$

- ▶ Vektoren definieren W 'keitsverteilung der Zufallsvariablen *Kopfs.* und *Wetter*
- ▶ $P(\text{Kopfs.}, \text{Wetter})$ ist (4×2) -Tabelle von W 'keiten aller Kombinationen der Werte einer Zufallsvariablen

	Kopfs. = TRUE	Kopfs. = FALSE
Wetter = sonnig	$P(W = \text{sonnig} \wedge \text{Kopfs.})$	$P(W = \text{sonnig} \wedge \neg \text{Kopfs.})$
Wetter = Regen		
Wetter = bewölkt		
Wetter = Schnee		

Verbundwahrscheinlichkeit

- ▶ W 'keit, die Agent jeder Aussage in Domäne zuordnet
- ▶ **atomares Ereignis:** Zuweisung von Werten an alle Zufallsvariablen X_1, \dots, X_n (= vollständige Spezifikation eines Zustands)

Beispiel: seien X, Y boolesche Variablen, dann gibt es folgende 4 atomaren Ereignisse:

$$X \wedge Y, \quad X \wedge \neg Y, \quad \neg X \wedge Y, \quad \neg X \wedge \neg Y$$

- ▶ **Verbundw'keitsverteilung** $P(X_1, \dots, X_n)$ weist jedem atomaren Ereignis eine W 'keit zu:

	Zahn	\neg Zahn
Karies	0.04	0.06
\neg Karies	0.01	0.89

- ▶ da alle atomaren Ereignisse disjunkt, ist Summe über alle Felder 1 (Disjunktion der Ereignisse), Konjunktion ist notwendigerweise falsch

Rechnen mit der Verbundwahrscheinlichkeit

alle interessanten Wahrscheinlichkeiten: aus Verbundw'keit errechenbar
durch Formulierung als Disjunktion von atomaren Ereignissen

$$\begin{aligned}\text{Beispiel: } P(\text{Karies} \vee \text{Zahn}) &= P(\text{Karies} \wedge \text{Zahn}) \\ &+ P(\neg \text{Karies} \wedge \text{Zahn}) \\ &+ P(\text{Karies} \wedge \neg \text{Zahn})\end{aligned}$$

unbedingte W'keiten durch Aufsummieren von Zeile oder Spalte
(Marginalisierung):

$$P(\text{Karies}) = P(\text{Karies} \wedge \text{Zahn}) + P(\text{Karies} \wedge \neg \text{Zahn})$$

womit wieder bedingte Wahrscheinlichkeiten ableitbar:

$$P(\text{Karies} \mid \text{Zahn}) = \frac{P(\text{Karies} \wedge \text{Zahn})}{P(\text{Zahn})} = \frac{0.04}{0.04 + 0.01} = 0.80$$

Die Bayes'sche Regel

wir wissen (Produktregel):

$$P(A \wedge B) = P(A | B)P(B) \quad \text{und} \quad P(A \wedge B) = P(B | A)P(A)$$

durch Gleichsetzen der rechten Seiten folgt:

$$P(A | B)P(B) = P(B | A)P(A)$$

$$\implies P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

für mehrwertige Variablen (Menge von Gleichungen):

$$P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)}$$

Verallgemeinerung (bzgl. Hintergrundevidenzen):

$$P(Y | X, E) = \frac{P(X | Y, E)P(Y | E)}{P(X | E)}$$

Anwendung der Bayes'schen Regel

$$P(\text{Zahnschmerz} \mid \text{Karies}) = 0.4$$

$$P(\text{Karies}) = 0.1$$

$$P(\text{Zahnschmerz}) = 0.05$$

$$P(\text{Karies} \mid \text{Zahnschmerz}) = \frac{0.4 \cdot 0.1}{0.05} = 0.8$$

warum nicht gleich $P(\text{Karies} \mid \text{Zahnschmerz})$ schätzen?

- ▶ kausales Wissen wie $P(\text{Zahnschmerz} \mid \text{Karies})$ ist i.A. robuster als diagnostisches Wissen $P(\text{Karies} \mid \text{Zahnschmerz})$:
- ▶ Kausalität $P(\text{Zahnschmerz} \mid \text{Karies})$ ist unabhängig von a priori W'keiten $P(\text{Zahnschmerz})$ und $P(\text{Karies})$
- ▶ nähme $P(\text{Karies})$ bei Karies-Epidemie zu, so bliebe Kausalität $P(\text{Zahnschmerz} \mid \text{Karies})$ unverändert, während sich $P(\text{Zahnschmerz})$ proportional mit $P(\text{Zahnschmerz})$ und $P(\text{Karies})$ änderte
- ▶ Arzt, der $P(\text{Karies} \mid \text{Zahnschmerz})$ geschätzt hätte, wüsste keine Regel zum Update von $P(\text{Karies} \mid \text{Zahnschmerz})$
- ▶ Arzt, der Bayes'sche Regel benutzt hätte, wüsste um proportionales Verhalten zwischen $P(\text{Zahnschmerz} \mid \text{Karies})$ und $P(\text{Karies})$

Normierung I

wenn absolute Wahrscheinlichkeit von $P(C | T)$ zu bestimmen und $P(T)$ nicht bekannt, dann vollständige Fallanalyse durchführen (z.B. für C und $\neg C$) und Zusammenhang $P(C | T) + P(\neg C | T) = 1$ (hier boole'sche Variable) ausnutzen:

$$P(C | T) = \frac{P(T | C)P(C)}{P(T)}$$

$$P(\neg C | T) = \frac{P(T | \neg C)P(\neg C)}{P(T)}$$

$$P(C | T) + P(\neg C | T) = \frac{P(T | C)P(C)}{P(T)} + \frac{P(T | \neg C)P(\neg C)}{P(T)}$$

$$P(T) = P(T | C)P(C) + P(T | \neg C)P(\neg C)$$

Normierung II

durch Einsetzen in oberste Gleichung:

$$P(C | T) = \frac{P(T | C)P(C)}{P(T | C)P(C) + P(T | \neg C)P(\neg C)}$$

für mehrwertige Zufallsvariablen:

$$P(Y | X) = \alpha P(X | Y)P(Y)$$

wobei α Normierungskonstante, welche Werte in $P(Y | X)$ zu 1 aufsummieren lässt, z.B.

$$\alpha(1, 1, 3) = \left(\frac{1}{5}, \frac{1}{5}, \frac{3}{5}\right)$$

Übersicht

1. Motivation

2. Grundlagen der Wahrscheinlichkeitstheorie

3. Probabilistische Inferenz

4. Bayes'sche Netze

Multiple Evidenzen I

Patient bejaht Frage nach Zahnschmerzen.
Aus dieser ersten Evidenz schließt Zahnarzt:

$$P(\text{Karies} \mid \text{Zahn}) = 0.8$$

Zahnarzt ertastet bei Untersuchung mit Haken Bruchstelle im Zahn
aus dieser zweiten Evidenz schließt er:

$$P(\text{Karies} \mid \text{Bruch}) = 0.95$$

beide Schlüsse könnten z.B. mit Bayes'scher Regel berechnet werden
aber was bringt kombinierte Evidenz?
mit Bayes'scher Regel könnte er weiter ermitteln:

$$P(\text{Karies} \mid \text{Zahn} \wedge \text{Bruch}) = \frac{P(\text{Zahn} \wedge \text{Bruch} \mid \text{Karies}) \cdot P(\text{Karies})}{P(\text{Zahn} \wedge \text{Bruch})}$$

Multiple Evidenzen II

Problem: er braucht $P(\text{Zahn} \wedge \text{Bruch} \mid \text{Karies})$, d.h. Diagnosewissen für alle Kombinationen von Symptomen i.a. Fall.

besser: Evidenzen mit Hilfe der Evidenzenregel schrittweise aufnehmen

$$P(Y \mid X, E) = \frac{P(X \mid Y, E)P(Y, E)}{P(X \mid E)}$$

mit bestimmter a priori W'keit hat Patient Loch: $P(\text{Karies})$
er berichtet von Zahnschmerzen (Bayes'sche Regel):

$$P(\text{Karies} \mid \text{Zahn}) = P(\text{Karies}) \cdot \frac{P(\text{Zahn} \mid \text{Karies})}{P(\text{Zahn})}$$

Multiple Evidenzen III

$$P(\text{Karies} \mid \text{Zahn}) = P(\text{Karies}) \cdot \frac{P(\text{Zahn} \mid \text{Karies})}{P(\text{Zahn})} \quad (1)$$

Untersuchung ergibt *Bruch*, also

$$P(\text{Karies} \mid \text{Bruch}, \text{Zahn}) = \frac{P(\text{Bruch} \mid \text{Karies}, \text{Zahn})}{P(\text{Bruch} \mid \text{Zahn})} \cdot P(\text{Karies} \mid \text{Zahn}) \quad (2)$$

(1) in (2) einsetzen ergibt

$$P(\text{Karies} \mid \text{Bruch}, \text{Zahn}) = P(\text{Karies}) \cdot \frac{P(\text{Zahn} \mid \text{Karies})}{P(\text{Zahn})} \cdot \frac{P(\text{Bruch} \mid \text{Karies}, \text{Zahn})}{P(\text{Bruch} \mid \text{Zahn})}$$

Multiple Evidenzen IV

Annahme bedingter Unabhängigkeit von *Zahn* und *Bruch* gegeben *Karies* (vereinfachtes Diagnosewissen):

$$P(\text{Bruch} \mid \text{Karies}, \text{Zahn}) = P(\text{Bruch} \mid \text{Karies})$$

$$P(\text{Zahn} \mid \text{Karies}, \text{Zahn}) = P(\text{Zahn} \mid \text{Karies})$$

$$P(\text{Karies} \mid \text{Bruch}, \text{Zahn}) = P(\text{Karies}) \cdot \frac{P(\text{Zahn} \mid \text{Karies})}{P(\text{Zahn})} \cdot \frac{P(\text{Bruch} \mid \text{Karies})}{P(\text{Bruch} \mid \text{Zahn})}$$

Multiple Evidenzen V

$$P(\text{Karies} \mid \text{Bruch}, \text{Zahn}) = P(\text{Karies}) \cdot \frac{P(\text{Zahn} \mid \text{Karies})}{P(\text{Zahn})} \cdot \frac{P(\text{Bruch} \mid \text{Karies})}{P(\text{Bruch} \mid \text{Zahn})}$$

immer noch nötig: Betrachtung von Paaren (Tripeln etc.) von Symptomen wie $P(\text{Bruch} \mid \text{Zahn})$

aber mit Produktregel folgt aus Nennern der Brüche:

$$P(\text{Zahn}) \cdot P(\text{Bruch} \mid \text{Zahn}) = P(\text{Bruch} \wedge \text{Zahn})$$

führt zu:

$$P(\text{Karies} \mid \text{Bruch}, \text{Zahn}) = P(\text{Karies}) \cdot \frac{P(\text{Zahn} \mid \text{Karies}) \cdot P(\text{Bruch} \mid \text{Karies})}{P(\text{Bruch} \wedge \text{Zahn})}$$

für Bestimmung von $P(\text{Karies} \mid \text{Bruch}, \text{Zahn})$ ist $P(\text{Bruch} \wedge \text{Zahn})$ Normierungsfaktor und eliminierbar, sofern $P(\text{Bruch} \mid \neg \text{Karies})$ und $P(\text{Zahn} \mid \neg \text{Karies})$ bekannt

Multiple Evidenzen VI

somit:

$$P(\text{Bruch} \wedge \text{Zahn}) = P(\text{Bruch} \wedge \text{Zahn} \mid \text{Karies})P(\text{Karies}) \\ + P(\text{Bruch} \wedge \text{Zahn} \mid \neg\text{Karies})P(\neg\text{Karies})$$

mit bedingter Unabhängigkeit:

$$P(\text{Bruch} \wedge \text{Zahn}) = P(\text{Zahn} \mid \text{Karies})P(\text{Bruch} \mid \text{Karies})P(\text{Karies}) \\ + P(\text{Zahn} \mid \neg\text{Karies})P(\text{Bruch} \mid \neg\text{Karies})P(\neg\text{Karies})$$

einsetzen in:

$$P(\text{Karies} \mid \text{Bruch}, \text{Zahn}) = P(\text{Karies}) \cdot \frac{P(\text{Zahn} \mid \text{Karies}) \cdot P(\text{Bruch} \mid \text{Karies})}{P(\text{Bruch} \wedge \text{Zahn})}$$

also Ableitung möglich aus unbedingten Wahrscheinlichkeiten und einfachen kausalen Regeln!

Multiple Evidenzen VII

mehrfache Evidenzen sind berechenbar durch Reduktion auf:

- ▶ a priori Wahrscheinlichkeiten und
- ▶ (kausale) bedingte Wahrscheinlichkeiten für Evidenz
- ▶ unter Annahme von gewissen bedingten Unabhängigkeiten

Allgemeine Kombinationsregel:

$$P(Z | X, Y) = \alpha P(Z)P(X | Z)P(Y | Z)$$

für X und Y bedingt unabhängig bei gegebenem Z und mit Normierungskonstante α

Zusammenfassung

- ▶ **Unsicherheit** ist unvermeidbar in komplexen und dynamischen Welten, in denen Agenten zur Ignoranz gezwungen sind
- ▶ **Wahrscheinlichkeiten** formulieren Unfähigkeit eines Agenten, definitive Entscheidung zu fällen
- ▶ drücken Grad seiner Überzeugung aus
- ▶ **bedingte** und **unbedingte** W'keiten über Aussagen formulierbar
- ▶ falls Agent die w'keitstheoretischen **Axiome** verletzt, so wird er unter Umständen irrationales Verhalten zeigen
- ▶ **Bayessche Regel** ermöglicht es, unbekannte W'keiten aus bekannten W'keiten zu berechnen
- ▶ **Multiple Evidenzen** können bei bedingter Unabhängigkeit effektiv in Berechnung einbezogen werden

Beispiel I

▶ **Expertenwissen:**

„Metastatischer Krebs ist eine mögliche Ursache von Hirntumor, und ist auch eine Erklärung für erhöhte Kalziumwerte im Blut. Beide Phänomene zusammen können erklären, dass ein Patient ins Koma fällt. Schwere Kopfschmerzen sind auch möglicherweise verbunden mit einem Hirntumor.“

▶ **Spezieller Fall:**

“Der Patient hat schwere Kopfschmerzen.”

▶ **Anfrage:**

Wird der Patient ins Koma fallen?

Wahl des Zustandsraums

Ω ist endliche Menge von erschöpfenden, sich gegenseitig ausschließenden Werten

Beispiel:

Explizite Spezifikation der Zufallsvariablen und ihrer Werte

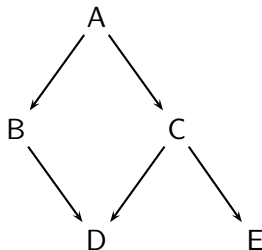
Zufallsvariable	Mögliche Werte
A metastatischer Krebs	$a, \neg a$
B erhöhter Calciumwert im Blut	$b, \neg b$
C Hirntumor	$c, \neg c$
D Koma	$d, \neg d$
E schwere Kopfschm.	$e, \neg e$

Wahrscheinlichkeit für gegebene Abhängigkeiten

I

$$\begin{aligned}
 &P(A = a, B = b, C = c, D = d, E = e) \\
 &= P(E = e \mid A = a, B = b, C = c, D = d) \\
 &\cdot P(D = d \mid A = a, B = b, C = c) \\
 &\cdot P(C = c \mid A = a, B = b) \\
 &\cdot P(B = b \mid A = a) \cdot P(A = a)
 \end{aligned}$$

$$\begin{aligned}
 &= P(E = e \mid C = c) \\
 &\cdot P(D = d \mid B = b, C = c) \\
 &\cdot P(C = c \mid A = a) \\
 &\cdot P(B = b \mid A = a) \\
 &\cdot P(A = a)
 \end{aligned}$$



Was muss vom Experten spezifiziert werden?

- ▶ a priori W'keiten der Knoten ohne Eltern
- ▶ bedingte W'keiten eines Knotens bezogen auf Menge von elementaren Knoten (Markov-Feld)

W'keit für gegebene Abhängigkeiten II

$P(e c) = 0.8$ $P(e \neg c) = 0.6$	Kopfschm. sind häufig, aber viel häufiger, falls Tumor vorhanden ist
$P(d b, c) = 0.8$ $P(d b, \neg c) = 0.8$ $P(d \neg b, c) = 0.8$ $P(d \neg b, \neg c) = 0.05$	Koma ist selten, aber häufig, falls einer der Gründe vorliegt
$P(b a) = 0.8$ $P(b \neg a) = 0.2$	erhöhtes Kalzium ist ungewöhnlich, aber eine häufige Konsequenz von Metastasen
$P(c a) = 0.2$ $P(c \neg a) = 0.05$	Hirntumor ist selten, aber häufige Konsequenz von Metastasen
$P(a) = 0.2$	Vorkommen von metastatischem Krebs in relevanten Studien

- ▶ 11 Werte (anstatt 31) spezifiziert, andere berechenbar
- ▶ Wahrscheinlichkeiten anhand von Fallstudien, Lehrbuchinformationen und massiven Tests

Mathematische Lösung des Problems

Gegeben:

Wahrscheinlichkeit P

Evidenz:

Patient hat starke Kopfschmerzen

Frage:

Wie groß ist die Wahrscheinlichkeit, dass der Patient ins Koma fallen wird?

Evidenz:

$E = e$

Bedingte Wahrscheinlichkeit:

$P(\bullet \mid E = e)$

Frage:

$D = d?$

Randwahrscheinlichkeit:

$P(D = d \mid E = e)$

Analyse dieses speziellen Falles

„Der Patient hat starke Kopfschmerzen“

Idee: Beschreibung der bedingten Wahrscheinlichkeiten $P(\bullet \mid E = e)$

$$P(A = a', B = b', C = c', D = d', E = e' \mid E = e) \\ = \begin{cases} \frac{P(A=a', B=b', C=c', D=d', E=e')}{\sum_{a \in A} \sum_{b \in B} \sum_{c \in C} \sum_{d \in D} P(a, b, c, d, e)} & \text{falls } e' = e \\ 0 & \text{falls } e' = \neg e \end{cases}$$

Auswertung der Daten des Patienten

Wird der Patient ins Koma fallen?

Idee: berechne Randw'keiten $P(\{d\} | \{e\})$

$$P(\{d\} | \{e\}) = \sum_{a' \in A} \sum_{b' \in B} \sum_{c' \in C} \sum_{e' \in E} P(a', b', c', d, e' | e)$$

Probleme:

- ▶ Komplexität der Berechnungen der bedingten Randw'keiten (100 Zufallsvariablen mit 3 Merkmalen: 3^{99} Summationen von kleinen Werten)
- ▶ können (oder müssen) wir erwarten, dass Nutzer mit W'keiten im Produktraum zurechtkommen?

Übersicht

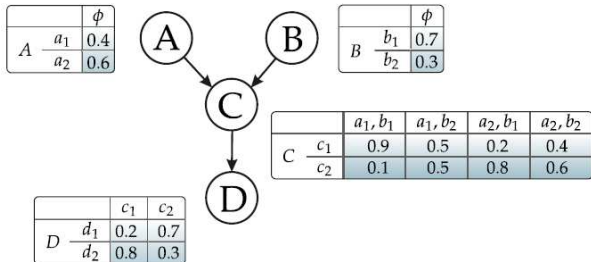
1. Motivation
2. Grundlagen der Wahrscheinlichkeitstheorie
3. Probabilistische Inferenz
- 4. Bayes'sche Netze**

Bayes'sche Netze

- ▶ graphische Modelle
- ▶ Formalismus zur Darstellung kausaler Abhängigkeiten
- ▶ Möglichkeit, einen hochdimensionalen Raum durch mehrere niederdimensionale Räume zu repräsentieren
- ▶ Grundlage: Bayes'scher Satz

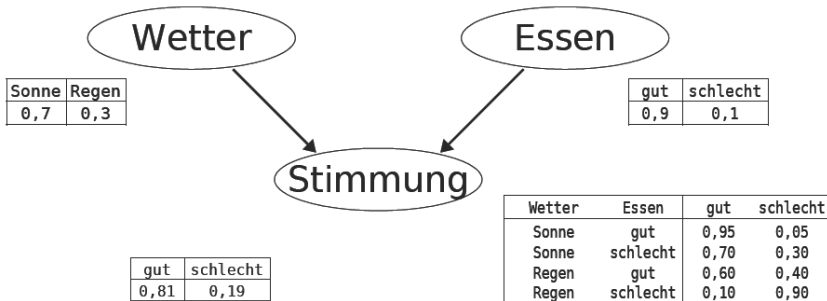
Beispiel I

- ▶ gerichteter, azyklischer Graph
- ▶ Knoten bedeuten Zufallsvariable
- ▶ Kanten symbolisieren (prob.) Abhängigkeiten
- ▶ beinhalten **qualitative** und **quantitative** Informationen



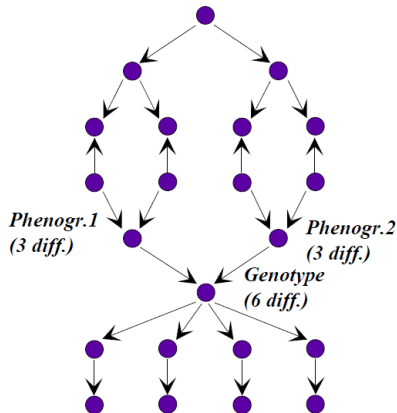
Beispiel II

- ▶ Knoten \rightarrow Zufallsvariable
- ▶ Kanten \rightarrow Abhängigkeiten der Variablen



Beispiel: Genotyp-Bestimmung des Jersey-Rinds

22 Variablen, $6 \cdot 10^{13}$ Zustände, 324 Parameter



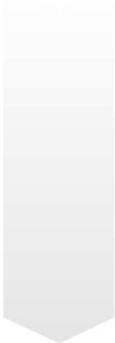
- ▶ graphisches Modell
- ▶ Knoten \rightarrow Zustandsvariablen
- ▶ Kanten \rightarrow Bedingte Abhängigkeiten
- ▶ Zerlegung \rightarrow

$$P(X_1, \dots, X_{22}) = \prod_{i=1}^{22} P(X_i \mid \text{Eltern}(X_i))$$
- ▶ Diagnose $\rightarrow P(\bullet \mid \text{Evidenz})$

Bayes'sche Netze

Top-Down-Ansatz

Expertenwissen



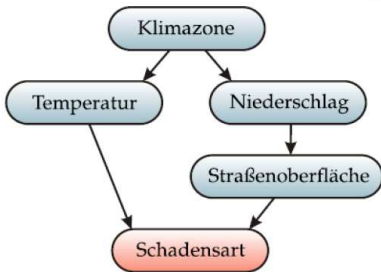
Daten

Bottom-Up-Ansatz

Expertenwissen



Daten



Erlernen von Bayes'schen Netzen aus Daten

- ▶ Erstellung eines Bayes'schen Netzes: von Hand sehr aufwändig
- ▶ viele Zusammenhänge müssen erkannt werden
- ▶ häufig keine Experten mit ausreichendem Wissen verfügbar
- ▶ Daten oft in großen Mengen vorhanden
- ▶ Lernverfahren ermöglichen Entwicklung Bayes'scher Netzwerke aus vorhandenen Daten
- ▶ Arten des Lernens:
 - Parameterlernen:
 - ▶ Netzstruktur bekannt, lediglich Parameter der W 'keitsverteilungen müssen geschätzt werden
 - Strukturlernen:
 - ▶ keine Netzstruktur vorgegeben
 - ▶ Abhängigkeiten aus den Daten bestimmen
- ▶ Erweiterungen auf Lernverfahren für Hyperbaum-Strukturen

Beispiel IV: Planungssystem für VW

- ▶ etwa **200 item families** (Variablen, Merkmale)
- ▶ von **2 bis 50 items pro item family** (Werte pro Variable)
- ▶ mehr als 2^{200} mögliche Fahrzeug-Spezifikationen
- ▶ Wahl der gültigen Spezifikationen ist beschränkt durch **Regelsysteme (10.000 technische Regeln**, noch mehr marketing- und produktionsorientierte Regeln)

Beispiel:

if engine = e_1 then transmission = t_3

*if engine = e_4 and auxiliary heater = h_1
then generator $\in \{g_3, g_4, g_5\}$*



Reales Netz

