Bayes-Netze

Vorlesung im Sommersemester 2012

Organisatorisches

- Zeit und Ort:

  Mo 14-16 Cartesium 0.01

- Prof. Carsten Lutz
  Raum Cartesium 2.59
  Tel. (218)-64431
  clu@uni-bremen.de

- Position im Curriculum:

  Modulbereich Theorie, Wahlbereich Master-Ergänzung, Vertiefungs VL

Organisatorisches

- Voraussetzungen:
  ein bischen Aussagenlogik und Wahrscheinlichkeitstheorie

- Form: K2, 4 Termine mit Übungen
  (Diskussion in VL jederzeit erwünscht!)

- Vorlesungsmaterial:

  Folien und Aufgabenblätter auf:

  http://www.informatik.uni-bremen.de/tdki/lehre/ss12/bayes/

  Beispiele, Beweise, etc an der Tafel (mitschreiben!)

Literatur

Basierend auf:

- Adnan Darwiche, Modelling and Reasoning with Bayesian Networks. Cambridge University Press, 2009.

Weitere Referenzen:

- Daphne Koller and Nir Friedman, Probabilistic Graphical Models - Principles and Techniques. MIT Press, 2009.
- Richard Neapolitan, Learning Bayesian Networks. Prentice Hall, 2003.
- Judea Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988.

Prüfungen

Mündliche Prüfung

oder

Übungen:

- Übungsaufgaben drei mal im Semester
- Werden in Gruppen (2-3 Personen) bearbeitet, abgegeben und korrigiert, mindestens einmal vorrechnen
- Fachgespräche am Ende des Semesters

Bayes-Netze

Vorlesung im Sommersemester 2012

Plan für Heute

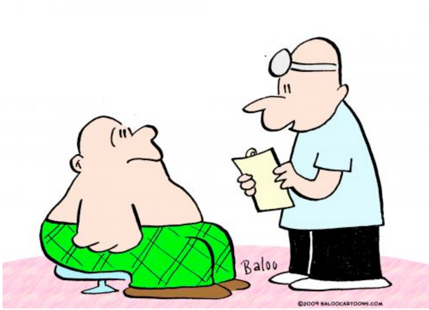
- Einführung und Motivation
- Grundlagen in Aussagenlogik und Wahrscheinlichkeitstheorie

Motivation

Unsicheres Wissen

Unsicheres Wissen

In vielen Anwendung der Informatik und künstlichen Intelligenz spielt unsicheres Wissen eine wesentliche Rolle



Medizinisches Assistenzsystem

- Vorgeschichte
- Symptome
- Testergebnisse

➡

- Diagnose
- Therapieempfehlung

Quellen von Unsicherheit z.B.:

- Vorgeschichte nicht vollständig bekannt, Symptome nicht eindeutig
- Testergebnisse fehlerbehaftet oder erlauben mehrere Schlüsse
- Mehrere Diagnosen und Therapien möglich

Unsicheres Wissen

In vielen Anwendung der Informatik und künstlichen Intelligenz spielt unsicheres Wissen eine wesentliche Rolle

Intelligenter Spamfilter

- Schlüsselwörter
- Absenderadresse
- SMTP-Relay

➟

- Spam-Level
- Aktion

Quellen von Unsicherheit z.B.:

- Sind Wörter wie Via9ra oder Val1um Anzeichen für Spam?
- Ist Mail von *****.ru Spam oder von der russischen Großmutter?
- War Textextraktion aus eingebettetem Bild korrekt?

Unsicheres Wissen

Charakteristische Merkmale:

- Unsicherheit erwächst aus unvollständigen Informationen, fehlerbehafteten “Sensoren”, unzureichender Modellierung
- Management / Beherrschung der Unsicherheit ist Schlüssel zur erfolgreichen Bewältigung der Anwendung.
- Es gibt sehr viele Parameter, die verarbeitet werden müssen (hunderte bis tausende medizinische Parameter / Spam-Merkmale)

Bayes-Netze sind wichtiger Vertreter von probabilistischen graphischen Modellen, die für solche Probleme sehr erfolgreich verwendet werden

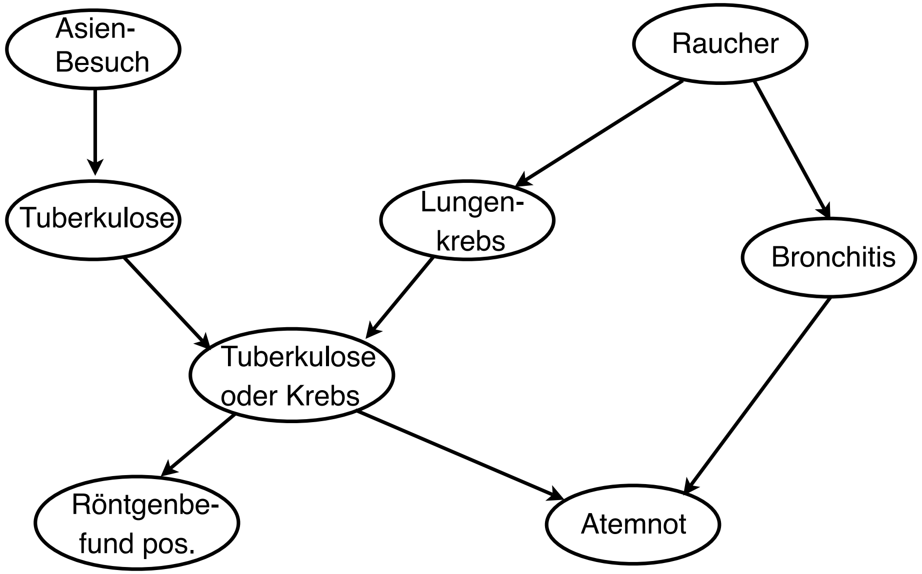
Unsicheres Wissen

Zentrale Eigenschaften von Bayes-Netzen:

- Solide formale Grundlage: Wahrscheinlichkeitstheorie
- Verwendung von Graphen: Intuitive und kompakte Datenstruktur
- Effizientes Schlußfolgern mit generellen Algorithmen möglich
- Lassen sich auch bei sehr vielen Parametern effektiv konstruieren:
  - Realistisch durchführbare Strategien zur Expertenbefragung
  - automatisches Lernen aus historischen / statistischen Daten

Unsicheres Wissen

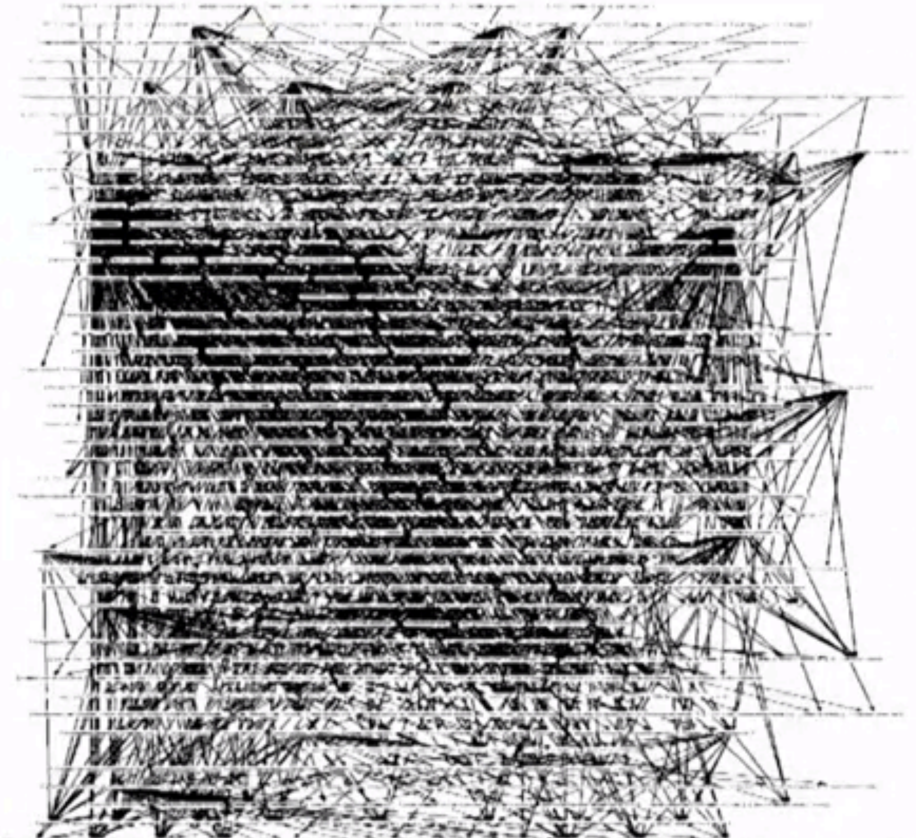
Ein medizinisches Bayes-Netz zum Beispiel:



+ geeignete Annotation mit Wahrscheinlichkeiten

Unsicheres Wissen

In der Praxis dann aber eher:



M. Pradhan, G. Provan, B. Middleton, M. Henrion, UAI 94

Unsicheres Wissen

Anwendungsgebiete sind beispielsweise:

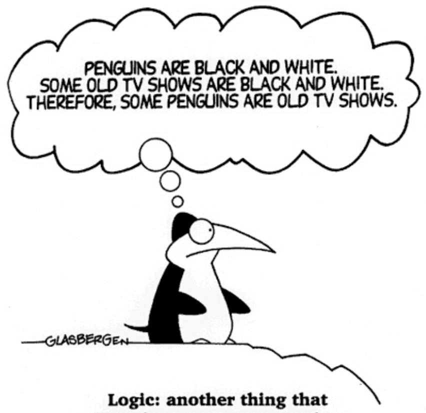
- Medizinische Diagnose
- Fehleranalyse
- Sprachverarbeitung
- Verkehrsanalyse
- Modelle für soziale Netzwerke
- Bildverarbeitung
- Spracherkennung
- Roboter: Lokalisierung und Kartenerstellung
- Bioinformatik
- und viele mehr

Übersicht Vorlesung

- Motivation und Grundlagen
- Kapitel 1: Bayes-Netze
- Kapitel 2: Schlussfolgerungsmechanismen
- Kapitel 3: Komplexität und Approximation
- Kapitel 4: Maschinelles Lernen

Grundlagen

Aussagenlogik



Logic: another thing that penguins aren't very good at.

Ereignisse und Logik

In den Beispielen spielen Ereignisse eine fundamentale Rolle, z.B.:

"Patient war in Asien"

"Patient hat Tuberkulose oder Krebs"

"Patient hatte einen positiven Röntgenbefund"

Aussagenlogik ist Sprache zur Beschreibung von Ereignissen:

- Variablen sind atomare (nicht näher beschriebene) Ereignisse
- Formeln ermöglichen Aussagen über Ereignisse
- Wahrheitswert beschreiben, ob ein Ereignis eintritt oder nicht

Aussagenlogik

Beispiel: Einbruchsalarm

Variablen / Ereignisse:

Einbruch	Es findet ein Einbruch statt
Erdbeben	Es gibt ein Erdbeben
Alarm	Der Einbruchsalarm wird ausgelöst
Radio	Im Radio wird über Erdbeben berichtet
Anruf	Nachbar ruft an und berichtet über Alarm

Aussagen über Ereignisse zum Beispiel:

$\text{Erdbeben} \vee \text{Einbruch}$ $\quad$ $(\text{Einbruch} \vee \text{Erdbeben}) \rightarrow \text{Alarm}$

$\text{Einbruch} \wedge \neg\text{Alarm}$ $\quad$ $(\neg\text{Einbruch} \wedge \neg\text{Erdbeben}) \rightarrow \neg\text{Alarm}$

Syntax

Wir nehmen geeignete endliche Menge VAR $= \{x_1, x_2, \ldots, x_n\}$ von *Aussagenvariablen* an.

Intuitiv kann jedes x_i Wahrheitswert wahr oder falsch annehmen, repräsentiert ein Ergeignis wie “es gibt ein Erdbeben”.

Definition Syntax Aussagenlogik

Die Menge AL der *aussagenlogischen Formeln* ist induktiv definiert durch

- jede Aussagenvariable x_i ist in AL
- Wenn $\varphi, \psi \in$ AL, dann auch $\neg\varphi$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$ in AL

Klammern werden weggelassen, wenn das Resultat eindeutig ist, wobei $\neg$ stärker bindet als $\wedge$ und $\vee$

also z.B. $\neg x \wedge y = (\neg x \wedge y) \neq \neg(x \wedge y)$

Semantik

Definition Semantik Aussagenlogik

Eine *Belegung* ist eine Abbildung $\omega : \text{VAR} \to \{\mathsf{true}, \mathsf{false}\}$. Sie definiert einen Wahrheitswert $\omega(\varphi)$ für jede Formel φ:

- $\omega(\neg\varphi) = \begin{cases} \mathsf{true} & \text{falls } \omega(\varphi) = \mathsf{false} \\ \mathsf{false} & \text{falls } \omega(\varphi) = \mathsf{true} \end{cases}$
- $\omega(\varphi \wedge \psi) = \begin{cases} \mathsf{true} & \text{falls } \omega(\varphi) = \mathsf{true} \text{ und } \omega(\psi) = \mathsf{true} \\ \mathsf{false} & \text{sonst} \end{cases}$
- $\omega(\varphi \vee \psi) = \begin{cases} \mathsf{true} & \text{falls } \omega(\varphi) = \mathsf{true} \text{ oder } \omega(\psi) = \mathsf{true} \\ \mathsf{false} & \text{sonst} \end{cases}$

Belegung ω repräsentieren wir auch als Teilmenge von VAR:

$$x \in \omega \quad \text{gdw} \quad \omega(x) = \mathsf{true}$$

Implikation

Weitere Junktoren als Abkürzung definierbar, insbesondere:

Implikation	$\varphi \to \psi$ steht für $\neg\varphi \vee \psi$
Biimplikation	$\varphi \leftrightarrow \psi$ steht für $(\varphi \to \psi) \wedge (\psi \to \varphi)$

Daraus ergibt sich:

$$\omega(\varphi \to \psi) = \begin{cases} \mathsf{true} & \text{falls } \omega(\varphi) = \mathsf{false} \text{ oder } \omega(\varphi) = \omega(\psi) = \mathsf{true} \\ \mathsf{false} & \text{sonst} \end{cases}$$

$$\omega(\varphi \leftrightarrow \psi) = \begin{cases} \mathsf{true} & \text{falls } \omega(\varphi) = \omega(\psi) \\ \mathsf{false} & \text{sonst} \end{cases}$$

Wir nehmen an, dass $\neg, \wedge, \vee$ stärker binden als $\to$ und $\leftrightarrow$,

$x \wedge y \to z$ steht also für $(x \wedge y) \to z$

Semantik

Wenn $\omega(\varphi) = \mathtt{true}$, dann sagen wir: ω *erfüllt* φ .

Wir schreiben dann auch $\omega \models \varphi$ und nennen ω ein *Modell* von φ.

Die Menge aller Modelle von φ bezeichnen wir mit $\mathsf{Mod}(\varphi)$.

Semantik

Jede Belegung repräsentiert eine mögliche Welt:

Welt	Erdbeben	Einbruch	Alarm
ω_1	false	false	false
ω_2	false	false	true
ω_3	false	true	false
ω_4	false	true	true
ω_5	true	false	false
ω_6	true	false	true
ω_7	true	true	false
ω_8	true	true	true

Es gilt zum Beispiel: $\omega_8 \models$ Einbruch $\vee$ Erdbeben $\rightarrow$ Alarm

$\omega_5 \not\models$ Einbruch $\vee$ Erdbeben $\rightarrow$ Alarm

Beachte: Für n Variablen gibt es 2^n Belegungen / Welten

Zentrale Begriffe

Definition Erfüllbarkeit, Gültigkeit

Eine Formel φ heißt

- *erfüllbar* wenn sie ein Modell hat (also $\mathsf{Mod}(\varphi) \neq \emptyset$)
- *gültig* oder *Tautologie* wenn jede Belegung Modell ist ($\mathsf{Mod}(\varphi) = 2^{\mathsf{VAR}}$)

Unerfüllbare Formeln z.B.:	$x \wedge \neg x$	$x \wedge \neg y \wedge (x \rightarrow y)$
Gültige Formeln z.B.:	$x \vee \neg x$	$(x \wedge y) \leftrightarrow \neg x \vee \neg y$

Definition Implikation, Äquivalenz

Für Formeln φ und ψ sagen wir,

- φ *impliziert* ψ wenn $\mathsf{Mod}(\varphi) \subseteq \mathsf{Mod}(\psi)$
- φ und ψ sind *äquivalent* wenn $\mathsf{Mod}(\varphi) = \mathsf{Mod}(\psi)$

Intantiierungen

Jede Belegung ω lässt sich durch Formel φ_ω beschreiben:

$$\varphi_\omega := \bigwedge_{\omega(x)=\text{true}} x \wedge \bigwedge_{\omega(x)=\text{false}} \neg x$$

Zum Beispiel:

Welt	Erdbeben	Einbruch	Alarm
ω_3	false	true	false

$$\varphi_{\omega_3} = \neg\text{Erdbeben} \wedge \text{Einbruch} \wedge \neg\text{Alarm}$$

Lemma

Für jede Belegung ω gilt $\mathsf{Mod}(\varphi_\omega) = \{\omega\}$.

Formel der Form φ_ω nennen wir Intantiierung von VAR.

Statt VAR verwenden wir dabei oft auch Teilmengen $X \subseteq$ VAR.

Mit $\mathsf{Inst}(X)$ bezeichnen wir die Menge aller Instantiierungen von X.

Mehrwertige Variablen

Anstatt zwei Wahrheitswerte true und false für jede Variable anzunehmen, werden wir manchmal auch mehrwertige Variablen zulassen

Zum Beispiel drei Werte für Alarm: an, aus, defekt

Syntax:

Alarm ist keine atomare Formel mehr, stattdessen drei atomare Formeln

$$\text{Alarm} = \mathsf{an} \qquad \text{Alarm} = \mathsf{aus} \qquad \text{Alarm} = \mathsf{defekt}$$

Zum Beispiel:

$$\neg\text{Erdbeben} \land \neg\text{Einbruch} \rightarrow \text{Alarm} = \mathsf{aus}$$

Für zweiwertige Variablen x ist also x Abkürzung für $x = \mathsf{true}$

$\neg x$ Abkürzung für $x = \mathsf{false}$

Mehrwertige Variablen

Semantik:

Belegung ω weist jeder Variable Wert aus ihrem Wertebereich zu

Mögliche Welten im vorigen Beispiel:

Welt	Erdbeben	Einbruch	Alarm
ω_1	false	false	an
ω_2	false	false	aus
ω_3	false	false	defekt
ω_4	false	true	an
ω_5	false	true	aus
ω_6	false	true	defekt
ω_7	true	false	an
ω_8	true	false	aus
ω_9	true	false	defekt
ω_{10}	true	true	an
ω_{11}	true	true	aus
ω_{12}	true	true	defekt

Mehrwertige Variablen

Alle Begriffe wie Modell, Intantiierung, Erfüllbarkeit, Äquivalenz usw. sind für den zweiwertigen und den mehrwertigen Fall analog

Intantiierung zum Beispiel

$$\text{Erdbeben} \wedge \neg\text{Einbruch} \wedge \text{Alarm} = \mathsf{defekt}$$

In dieser VL:

- arbeiten wir üblicherweise mit zweiwertiger Logik
- gehen wir ohne weiteren Kommentar zu mehrwertiger Logik über wenn das adäquater ist.

Grundlagen

Unsicheres Wissen und Wahrscheinlichkeiten

Wahrscheinlichkeit

Bei unsicherem Wissen ist Aussagenlogik nicht präzise genug:

einzig mögliche Aussagen sind

”das Ereignis findet statt” und ”das Ereignis findet nicht statt”

Stattdessen braucht man Grad persönlicher Überzeugung für das Stattfinden eines Ereignisses (engl: Degree of Belief):

”ich glaube zum Grad g, dass das Ereignis stattgefunden hat”

Wir verwenden verfeinerte Semantik basierend auf Wahrscheinlichkeiten

Semantik: Verteilungen

Semantik für sicheres Wissen: Belegung repräsentiert eine einzelne Welt

Semantik für unsicheres Wissen:
Weise jeder Welt eine Wahrscheinlichkeit zu: Grad der Überzeugung, dass diese Welt der wirklichen Situation entspricht

Definition Wahrscheinlichkeitsverteilung

Eine *(Wahrscheinlichkeits)verteilung über VAR* ist eine Abbildung $\mathsf{Pr} : 2^{\mathsf{VAR}} \rightarrow [0, 1]$ so dass

$$\sum_{\omega} \mathsf{Pr}(\omega) = 1.$$

Die Verteilung Pr bestimmt für jede aussagenlogische Formel φ eine Wahrscheinlichkeit

$$\mathsf{Pr}(\varphi) := \sum_{\omega \models \varphi} \mathsf{Pr}(\omega).$$

Verteilungen

Welt	Erdbeben	Einbruch	Alarm	$\mathsf{Pr}(\cdot)$
ω_1	false	false	false	.7128
ω_2	false	false	true	.0072
ω_3	false	true	false	.0180
ω_4	false	true	true	.1620
ω_5	true	false	false	.0240
ω_6	true	false	true	.0560
ω_7	true	true	false	.0010
ω_8	true	true	true	.0190

$$\mathsf{Pr}(\mathsf{Erdbeben}) = \mathsf{Pr}(\omega_5) + \mathsf{Pr}(\omega_6) + \mathsf{Pr}(\omega_7) + \mathsf{Pr}(\omega_8) = .1$$

$$\mathsf{Pr}(\neg\mathsf{Erdbeben}) = .9$$

$$\mathsf{Pr}(\mathsf{Erdbeben} \wedge \mathsf{Alarm}) =$$

$$\mathsf{Pr}(\neg(\mathsf{Erdbeben} \vee \mathsf{Einbruch} \rightarrow \mathsf{Alarm})) =$$

Problem

Repräsentation:

- Für sicheres Wissen muß nur eine einzige Welt repräsentiert werden. Kann bei n Variablen als Folge von n Bits dargestellt werden.
- Für unsicheres Wissen muß eine Verteilung repräsentiert werden. Das ist bei n Variablen eine Tabelle mit 2^n Wahrscheinlichkeiten.

Problem:

- Unrealistischer Speicherbedarf: 40 Variablen $\rightsquigarrow >$1 Billion Welten
- Modellierung: Wer sollte so viele Wahrscheinlichkeiten angeben?
- Einzelne Welten entsprechen teilweise extrem exotischen Fällen, für die auch ein Experte keine Wahrscheinlichkeit angeben kann.

Diese Probleme werden wir später mit Bayes-Netzen lösen

Logik und Wahrscheinlichkeit

Einige einfache Beobachtungen bezüglich der Wahrscheinlichkeit aussagenlogischer Formeln.

Lemma

Für jede Verteilung Pr und aussagenlogische Formel φ gilt:

1. $\Pr(\varphi) \in [0, 1]$
2. $\Pr(\varphi) = 0$ falls φ unerfüllbar ist
3. $\Pr(\varphi) = 1$ falls φ Tautologie ist
4. $\Pr(\varphi) = 1 - \Pr(\neg\varphi)$ (also auch $\Pr(\varphi) + \Pr(\neg\varphi) = 1$)
5. $\Pr(\varphi \vee \psi) = \Pr(\varphi) + \Pr(\psi) - \Pr(\varphi \wedge \psi)$

Grundlagen

Konditionale Wahrscheinlichkeiten / Konditionierung

Konditionale Wahrscheinlichkeiten

Es gibt vielerlei Abhängigkeiten zwischen unserem Glauben an das Eintreten verschiedener Ereignisse

Zum Beispiel: im allgemeinen glauben wir, dass

$$\Pr(\text{Einbruch}) = 0.2 \qquad \Pr(\text{Erdbeben}) = 0.1$$

Wenn wir allerdings wissen, dass der Alarm ausgelöst wurde, so steigt unser Glauben an Erbeben und Einbruch, z.B.:

$$\Pr(\text{Einbruch}|\text{Alarm}) = 0.7 \qquad \Pr(\text{Erdbeben}|\text{Alarm}) = 0.3$$

Generelle Frage:

Wie sollen sich die Wahrscheinlichkeiten durch neue Evidenz verändern?

Konditionierung liefert systematischen Ansatz!

Konditionale Wahrscheinlichkeiten

Wir suchen nach Verteilung $\mathsf{Pr}(\cdot|\psi)$, die sich aus Pr durch Evidenz ψ ergibt

Einige Postulate für “vernünftige” Resultate $\mathsf{Pr}(\cdot|\psi)$:

1. Es soll $\mathsf{Pr}(\psi|\psi) = 1$ und $\mathsf{Pr}(\neg\psi|\psi) = 0$ gelten; also:

   wenn $\omega \not\models \psi$, dann $\mathsf{Pr}(\omega|\psi) = 0$

2. Wkten sollen so wenig wie möglich verändert werden (I)

   wenn $\mathsf{Pr}(\omega) = 0$, dann $\mathsf{Pr}(\omega|\psi) = 0$

3. Wkten sollen so wenig wie möglich verändert werden (II)

   wenn $\omega \models \psi$ und $\omega' \models \psi$, dann bleibt die relative Wkt von ω und ω' gleich:

$$\frac{\mathsf{Pr}(\omega|\psi)}{\mathsf{Pr}(\omega'|\psi)} = \frac{\mathsf{Pr}(\omega)}{\mathsf{Pr}(\omega')}$$

Konditionale Wahrscheinlichkeiten

Interessanterweise ist $\mathsf{Pr}(\cdot|\psi)$ durch diese Postulate eindeutig bestimmt
d.h. es gibt nur einen Weg, $\mathsf{Pr}(\cdot|\psi)$ ”vernünftig” zu definieren.

Definition Konditionierung

Für jede Verteilung $\mathsf{Pr}(\cdot)$ und jede aussagenlogische Formel ψ ist die Verteilung $\mathsf{Pr}(\cdot|\psi)$ definiert durch

$$\begin{aligned} \mathsf{Pr}(\omega|\psi) &= 0 && \text{wenn } \omega \not\models \psi \\ \mathsf{Pr}(\omega|\psi) &= \frac{\mathsf{Pr}(\omega)}{\mathsf{Pr}(\psi)} && \text{wenn } \omega \models \psi. \end{aligned}$$

Wir nennen $\mathsf{Pr}(\varphi|\psi)$ die *konditionale Wahrscheinlichkeit von* φ *gegeben* ψ und $\mathsf{Pr}(\cdot|\psi)$ das *Ergebnis der Konditionierung von* Pr *mit* ψ.

Also:

Welten ω, die ψ falsch machen, bekommen Wahrscheinlichkeit 0

Die anderen Welten werden von $\sum_{\omega\models\psi} \mathsf{Pr}(\omega)$ auf 1 hochskaliert

Konditionierung

Welt	Erdbeben	Einbruch	Alarm	$Pr(\cdot)$
ω_1	false	false	false	.7128
ω_2	false	false	true	.0072
ω_3	false	true	false	.0180
ω_4	false	true	true	.1620
ω_5	true	false	false	.0240
ω_6	true	false	true	.0560
ω_7	true	true	false	.0010
ω_8	true	true	true	.0190

Wir möchten konditionieren mit der neuen Evidenz Alarm

Konditionierung

Welt	Erdbeben	Einbruch	Alarm	$Pr(\cdot)$	$Pr(\cdot \mid Alarm)$
~~ω_1~~	~~false~~	~~false~~	~~false~~	~~.7128~~	0
ω_2	false	false	true	.0072	$\approx$ .029
~~ω_3~~	~~false~~	~~true~~	~~false~~	~~.0180~~	0
ω_4	false	true	true	.1620	$\approx$ .663
~~ω_5~~	~~true~~	~~false~~	~~false~~	~~.0240~~	0
ω_6	true	false	true	.0560	$\approx$ .229
~~ω_7~~	~~true~~	~~true~~	~~false~~	~~.0010~~	0
ω_8	true	true	true	.0190	$\approx$ .078

Wir möchten konditionieren mit der neuen Evidenz Alarm

Es gilt $Pr(Alarm) = .2442$

Konditionierung

Welt	Erdbeben	Einbruch	Alarm	$\mathrm{Pr}(\cdot)$	$\mathrm{Pr}(\cdot \mid \mathrm{Alarm})$
~~ω_1~~	~~false~~	~~false~~	~~false~~	~~.7128~~	0
ω_2	false	false	true	.0072	$\approx .029$
~~ω_3~~	~~false~~	~~true~~	~~false~~	~~.0180~~	0
ω_4	false	true	true	.1620	$\approx .663$
~~ω_5~~	~~true~~	~~false~~	~~false~~	~~.0240~~	0
ω_6	true	false	true	.0560	$\approx .229$
~~ω_7~~	~~true~~	~~true~~	~~false~~	~~.0010~~	0
ω_8	true	true	true	.0190	$\approx .078$

Konditionierung

Welt	Erdbeben	Einbruch	Alarm	$Pr(\cdot)$	$Pr(\cdot \mid Alarm)$
~~ω_1~~	~~false~~	~~false~~	~~false~~	~~.7128~~	0
ω_2	false	false	true	.0072	$\approx .029$
~~ω_3~~	~~false~~	~~true~~	~~false~~	~~.0180~~	0
ω_4	false	true	true	.1620	$\approx .663$
~~ω_5~~	~~true~~	~~false~~	~~false~~	~~.0240~~	0
ω_6	true	false	true	.0560	$\approx .229$
~~ω_7~~	~~true~~	~~true~~	~~false~~	~~.0010~~	0
ω_8	true	true	true	.0190	$\approx .078$

$$Pr(\text{Einbruch}) = .2 \qquad Pr(\text{Einbruch} \mid \text{Alarm}) \approx .741$$

$$Pr(\text{Erdbeben}) = .1 \qquad Pr(\text{Erdbeben} \mid \text{Alarm}) \approx .307$$

$$Pr(\neg\text{Erdbeben} \wedge \neg\text{Einbruch}) = .72$$

$$Pr(\neg\text{Erdbeben} \wedge \neg\text{Einbruch} \mid \text{Alarm}) = .029$$

Konditionierung

In der Definition von Konditionierung legen wir die Wkten von Welten fest.

Daraus folgt folgendes Verhalten der Wkten von Formeln

Lemma

Sei Pr eine Verteilung und φ, ψ Formeln. Dann gilt:

$$\mathsf{Pr}(\varphi|\psi) = \frac{\mathsf{Pr}(\varphi \wedge \psi)}{\mathsf{Pr}(\psi)}$$

Beachte: Konditionierung ist nur definiert wenn $\mathsf{Pr}(\varphi) > 0$

Man nennt $\mathsf{Pr}(\varphi|\psi)$ die konditionale Wahrscheinlichkeit von φ gegeben ψ

Grundlagen

Unabhängigkeit

Unabhängigkeit

Wir haben gerade gesehen, dass die Wkten mancher Ereignisse voneinander abhängen (z.B. Alarm und Einbruch)

Andere Ereignisse sind intuitiv unabhängig (z.B. Einbruch und Erdbeben)

Wir würden also erwarten, dass

$$\Pr(\text{Einbruch}|\text{Erdbeben}) = \Pr(\text{Einbruch})$$

$$\text{und } \Pr(\text{Erdbeben}|\text{Einbruch}) = \Pr(\text{Erdbeben})$$

Solche Unabhängigkeiten werden wir in Bayes-Netzen explizit modellieren und auf diese Weise das Repräsentationsproblem für Verteilungen lösen.

Unabhängigkeit

Definition Unabhängig

Seien φ, ψ aussagenlogische Formeln. Wir nennen φ *unabhängig* von ψ wenn $\mathsf{Pr}(\varphi|\psi) = \mathsf{Pr}(\varphi)$ oder $\mathsf{Pr}(\psi) = 0$

Beachte:

- wenn $\mathsf{Pr}(\psi) = 0$, dann kann man mit ψ nicht konditionieren
  intuitiv sollte φ aber nicht von einem unmöglichen Ereignis ψ abhängen
- φ is unabhängig von ψ

  gdw.

  $\mathsf{Pr}(\varphi \wedge \psi) = \mathsf{Pr}(\varphi) \cdot \mathsf{Pr}(\psi)$ — oft auch als Definition von Unabhängigkeit verwendet

  gdw.

  ψ ist unabhänging von φ

Unabhängigkeit

In unserer ursprünglichen Verteilung sind Erdbeben und Einbruch in der Tat unabhängig:

$\text{Pr}(\text{Erdbeben}) = .1$

Welt	Erdbeben	Einbruch	Alarm	$\text{Pr}(\cdot)$		$\text{Pr}(\cdot \mid \text{Erdbeben})$
ω_1	false	false	false	.7128		0
ω_2	false	false	true	.0072		0
ω_3	false	true	false	.0180	.1800	0
ω_4	false	true	true	.1620		0
ω_5	true	false	false	.0240		.2400
ω_6	true	false	true	.0560		.5600
ω_7	true	true	false	.0010	.0200	.0100
ω_8	true	true	true	.0190		.1900

$\text{Pr}(\text{Einbruch}) = .2$ $\qquad$ $\text{Pr}(\text{Einbruch} \mid \text{Erdbeben}) = .2$

Unabhängigkeit

Einfache Beobachtungen bzgl der Unabhängigkeit von AL-Formeln.

Lemma

Für jede Verteilung $\Pr$ und alle AL-Formeln φ, ψ_1, ψ_2 gilt:

1. wenn φ unabhängig von ψ, dann $\neg\varphi$ unabhängig von ψ
2. wenn φ unabhängig von ψ, dann φ unabhängig von $\neg\psi$
3. wenn φ unabhängig von ψ, dann $\neg\varphi$ unabhängig von $\neg\psi$

Lemma

Im allgemeinen gilt nicht:

1. wenn φ unabh. von ψ_1 und von ψ_2, dann φ unabh. von $\psi_1 \wedge \psi_2$
2. wenn φ unabh. von $\psi_1 \wedge \psi_2$, dann φ unabh. von ψ_1 und von ψ_2

Unabhängigkeit

Ein (recht extremer) Spezialfall ist Unabhängigkeit aller Ereignisse:

$(*)$ für alle Variablen x und $\Gamma \subseteq \mathsf{VAR} \setminus \{x\}$:

x ist unabhängig von φ_ω für alle $\varphi_\omega \in \mathsf{Inst}(\Gamma)$

Dann definieren Wkten der Variablen in eindeutiger Weise eine Verteilung:

$$\mathsf{Pr}(\omega) = \prod_{\omega \models x} \mathsf{Pr}(x) \cdot \prod_{\omega \not\models x} (1 - \mathsf{Pr}(x))$$

Kompakt speicherbar! Bayes-Netze basieren auf dieser Beobachtung, nehmen aber nicht alle Ereignisse als unabhängig an.

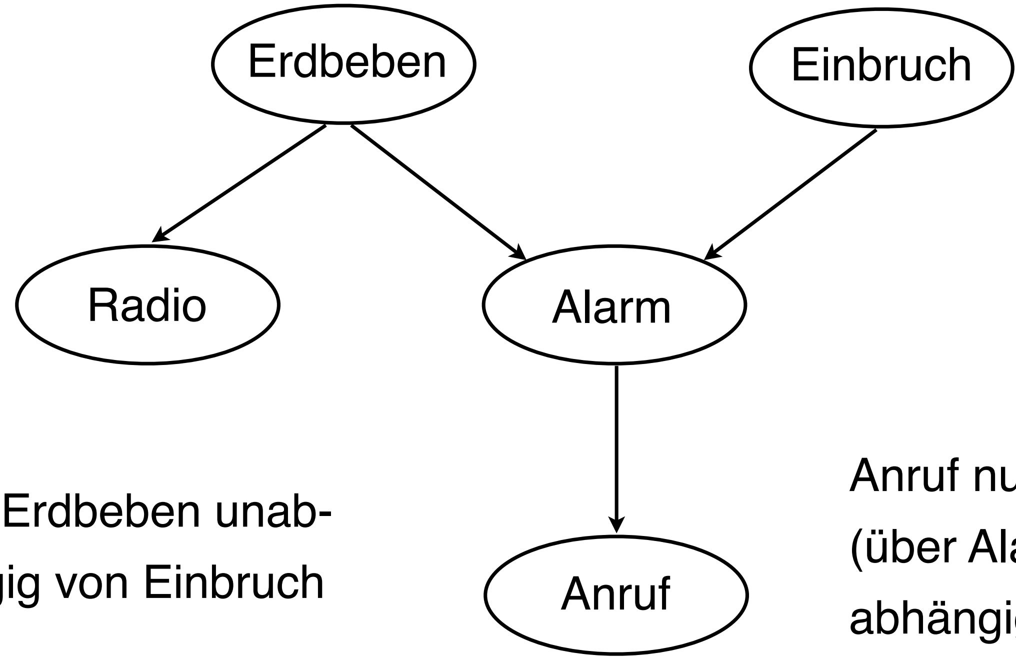
Beachte:

$(*)$ nicht dasselbe wie: alle Aussagenvariablen paarweise unabhängig

(vergl. Beobachtungen zu Konjunktion und Unabhängigkeit auf voriger Folie)

Vorschau Bayes-Netze

Bayes-Netze verwenden Graph, um Unabhängigkeiten zu spezifizieren:



Z.B. Erdbeben unabhängig von Einbruch

Anruf nur mittelbar (über Alarm) abhängig von Erdbeben

Basierend auf einer solchen Struktur kann man dann Verteilungen in (meist) kompakter Weise beschreiben

Übersicht Vorlesung

- Motivation und Grundlagen
- Kapitel 1: Bayes-Netze
- Kapitel 2: Schlussfolgerungsmechanismen
- Kapitel 3: Komplexität und Approximation
- Kapitel 4: Maschinelles Lernen