Künstliche Intelligenz

Wahrscheinlichkeitsrechnung

Dr.-Ing. Stefan Lüdtke

Universität Leipzig

Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI)

Warum Wahrscheinlichkeiten?

- Sei A_t: Fahre zum Flughafen t Minuten bevor der Flug geht
- Werde ich mit A_t rechtzeitig zum Flughafen kommen?
- Problem:
  - Partielle Beobachtbarkeit: Straßen, Pläne anderer Verkehrsteilnehmer, ...
  - Sensorrauschen (Akgekündigte Flugzeit)
  - Unsicherheit in Effekten der Aktionen (Platter Reifen...)
  - Generelle Modellierungs-Komplexität
- Rein logikbasierte Ansätze riskieren entweder, falsch zu sein (A_25 wird reichen) oder nicht hilfreich (A_25 reicht, wenn kein Stau und keine Panne)
- Lösung: Wahrscheinlichkeitsrechnung erlaubt quantifizierte Aussagen über das Eintreten von unsicheren Ereignissen

Warum Wahrscheinlichkeiten?

Wahrscheinlichkeitsrechnung erlaubt:

- Faulheit
  - Zu aufwendig, alle möglichen Aktionen und Effekte zu spezifizieren
  - Extrem große entstehende Zustandsräume, sodass wir keine Lösung in begrenzter Zeit finden
- Theoretische Ignoranz
  - Unvollständiges Wissen / Verständnis von Zusammenhängen, z.B. Medizin, Wirtschaft, ...
- Praktische Ignoranz
  - Selbst wenn vollständige Zusammenhänge im Prinzip verstanden, können wir nicht immer aufwendig zunächst alle Informationen erheben

Grundlagen der Wahrscheinlichkeitsrechnung

- Achtung: Im Folgenden vereinfachte Definitionen, die mathematisch nicht ganz sauber sind, für unsere Zwecke aber ausreichen
- Sei Ω eine Menge (Sample Space) und $\omega \in \Omega$ eine mögliche Welt
- z.B. Gleichzeitiges Werfen eines Würfels und einer Münze: $\Omega = \{(1, K), (1, Z), (2, K), \dots, (6, Z)\}$
- Sei $P : \Omega \to \mathbb{R}_{\geq 0}$ mit
  - $0 \leq P(\omega) \leq 1$
  - $\sum_{\omega \in \Omega} P(\omega) = 1$
- Nennen P eine Wahrscheinlichkeitsverteilung
- Ein Ereignis A ist eine Teilmenge von Ω, mit

$$P(A) = \sum_{\omega \in A} P(A)$$

Grundlagen der Wahrscheinlichkeitsrechnung

- Eine Zufallsvariable ist eine Funktion, die mögliche Welten auf "Teile" der Welt abbildet, z.B. $X((w, m)) = m$
- D.h. eine Zufallsvariable extrahiert Bestandteile der Welt, für die wir uns interessieren
- Eine Wahrscheinlichkeitsverteilung P über Welten definiert eine Wahrscheinlichkeitsverteilung über jeder Zufallsvariablen:

$$P(X = x_i) = \sum_{\omega : X(\omega) = x_i} P(\omega)$$

- Schreibweise: Großbuchstaben für Zufallsvariablen, Kleinbuchstaben für Belegungen

Grundlagen der Wahrscheinlichkeitsrechnung

- Seien X und Y Zufallsvariablen im Sample Space Ω. Dann ist die gemeinsame Wahrscheinlichkeitsverteilung von X und Y:

$$P(X = x, Y = y) = \sum_{\{\omega : X(\omega) = x \wedge Y(\omega) = y\}} P(\omega)$$

- z.B. Sei Ω das gleichzeitige Werfen von einem grünen, einem blauen und einem roten Würfel. Dann ist

$$P(G = 3, R = 1) = 1/36$$

Grundlagen der Wahrscheinlichkeitsrechnung

- Bedingte Wahrscheinlichkeit:

$$P(X \mid Y) = P(X, Y)/P(Y)$$

- Interpretation: Wahrscheinlichkeit, dass x eintritt, wenn y bereits eingetreten ist
- Durch Umformen ergibt sich die Produktregel:

$$P(X, Y) = P(X \mid Y)\, P(Y)$$

- Außerdem ergibt sich direkt der Satz von Bayes:

$$P(X \mid Y) = P(Y \mid X)\, P(X)/P(Y)$$

- Zufallsvariablen X und Y heißen unabhängig, wenn

$$P(X, Y) = P(X)\, P(Y)$$

Grundlagen der Wahrscheinlichkeitsrechnung

- Summenregel (ergibt sich aus der Definition der gemeinsamen Verteilung):

$$P(X = x) = \sum_y P(X = x, Y = y)$$

- Gesetz der totalen Wahrscheinlichkeit (Summenregel und Def. bedingte Wahrscheinlichkeiten):

$$P(X = x) = \sum_y P(X = x \,|\, Y = y) P(Y = y)$$

Beispiel

- Wahrscheinlichkeit einer Krankheit: $P(krank) = 0.0002$
- Sensitivität des Tests: $P(positiv \,|\, krank) = 1$
- Wahrscheinlichkeit eines falsch-positiven Ergebnisses: $P(positiv \,|\, gesund) = 0.01$
- Sie haben ein positives Testergebnis erhalten. Was ist die Wahrscheinlichkeit, tatsächlich erkrankt zu sein?

Beispiel

- Wahrscheinlichkeit einer Krankheit: $P(krank) = 0.0002$
- Sensitivität des Tests: $P(positiv \mid krank) = 1$
- Wahrscheinlichkeit eines falsch-positiven Ergebnisses: $P(positiv \mid gesund) = 0.01$
- Sie haben ein positives Testergebnis erhalten. Was ist die Wahrscheinlichkeit, tatsächlich erkrankt zu sein?

$$
\begin{aligned}
P(krank \mid positiv) &= \frac{P(positiv \mid krank)\, P(krank)}{P(positiv)} \\
&= \frac{P(positiv \mid krank)\, P(krank)}{P(positiv \mid krank) P(krank) + P(positiv \mid gesund) P(gesund)} \\
&= \frac{1 * 0.0002}{1 * 0.0002 + 0.01 * 0.9998} \approx 0.02
\end{aligned}
$$

Gemeinsame Wahrscheinlichkeitsverteilung

- Die gemeinsame Wahrscheinlichkeitsverteilung über alle Zufallsvariablen (*full joint*) definiert alle marginalen und bedingten Wahrscheinlichkeiten

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *cavity*	.016	.064	.144	.576

Gemeinsame Wahrscheinlichkeitsverteilung

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

$$P(To = true) = \sum_{x \in Cat, y \in Cav} P(To = true, Cat = x, Cav = y)$$

$$= 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

Gemeinsame Wahrscheinlichkeitsverteilung

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

$$P(Cav = false \,|\, To = true)$$

$$= P(Cav = false, To = true)/P(To = true)$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Probabilistische Inferenz durch Aufzählen

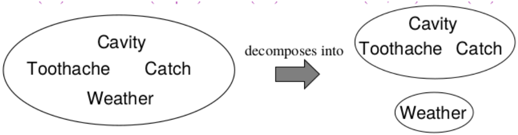
- Sei $\mathbf{X}$ die Sequenz aller Zufallsvariablen
- Im Allgemeinen wollen wir die gemeinsame Verteilung von *Query*-Variablen $\mathbf{Y}$, gegeben Werte $\mathbf{e}$ der *Evidence*-Variablen $\mathbf{E}$ berechnen
- Bezeichnen $\mathbf{H} = \mathbf{X} \setminus \mathbf{Y} \setminus E$ als *Hidden* Variablen.
- Generelle Idee: Marginalisiere die Hidden Variables:

$$P(\mathbf{Y} \,|\, \mathbf{E} = \mathbf{e}) = \alpha \sum_{\mathbf{h}} P(\mathbf{Y}, \mathbf{E} = e, \mathbf{H} = h)$$

- Joint besteht aus $\mathcal{O}(d^n)$ Werten ($n =$ Anzahl der Variablen, $d =$ Anzahl der Belegungen jeder Variablen)
- Problem: Speichern und summieren über so viele Werte nicht mögich in der Praxis
- Brauchen schlauere Methoden zur Repräsentation und Rechnen mit Wahrscheinlichkeitsverteilungen!

Unabhängigkeit

- Was uns hilft: Unabhängigkeit von Zufallsvariablen



- $P(To, Cav, Cat, W) = P(To, Cav, Cat)\, P(W)$
- Brauchen nur 12 Zahlen statt 32 zu speichern!
- Weiteres Beispiel: n Münzen: $2^n \to n$ Werte!
- Aber: Vollständige Unabhängigkeit ist selten...

Bedingte Unabhängigkeit

- Wenn ich Karies hab, ist die Wahrscheinlichkeit, dass die Zahnärztin diesen entdeckt (Catch) unabhängig von meinen Zahnschmerzen:
  $P(Cat \,|\, To, Cav = true) = P(Cat \,|\, Cav = true)$
- Das gleiche, wenn ich kein Karies hab:
  $P(Cat \,|\, To, Cav = false) = P(Cat \,|\, Cav = false)$
- Insgesamt: Cat ist bedingt unabhängig von To, gegeben Cav:
  $P(Cat \,|\, To, Cav) = P(Cat \,|\, Cav)$
- Erlaubt es, die Full Joint zu schreiben als
  $P(Cat, To, Cav) = P(To \,|\, Ca)\, P(Ca \,|\, Cat)\, P(Ca)$
- Nur 5 Werte statt ursprünglich 7

Zusammenfassung

- Wahrscheinlichkeitsrechnung erlaubt die Formalisierung von unsicherem Wissen
- Die Full Joint spezifiziert ein Wahrscheinlichkeitsmodell vollständig
- Queries können durch Summierung über die nicht relevanten Variablen (Marginalisierung) beantwortet werden
- Full Joint ist in der Praxis zu groß, müssen kompaktere Darstellung finden
- Unabhängigkeit und bedingte Unabhängigkeit erlauben dies