



Künstliche Intelligenz: Grundlagen und Anwendungen

Albayrak, Fricke (AOT) – Oppper, Grünhage (KI)
Wintersemester 2015 / 2016

5. Aufgabenblatt

Musterlösung

Aufgabe 1 – Satz von Bayes (20%)

Angenommen Sie sind Zeuge eines nächtlichen Unfalls mit Fahrerflucht eines Taxis in Athen. Alle Athener Taxis sind grün oder blau. Sie bezeugen unter Eid, dass das Taxi blau war. Tests zeigen allerdings, dass die Unterscheidung zwischen blau und grün bei schlechten Lichtverhältnissen nur zu 75% korrekt ist.

- (a) Ist es mit diesen Angaben möglich zu berechnen, welche Farbe die wahrscheinlichste Farbe des Taxis war?

Ohne eine Angabe zur Häufigkeit von grünen Taxis in Athen ist die gesuchte Wahrscheinlichkeit nicht berechenbar.

- (b) Ist es dann möglich, wenn man zusätzlich weiß, dass 8 aus 10 Athener Taxis grün sind?

- Es werden im Folgenden die Ereignisse $A =$ „Das Taxi erscheint blau“ und $B =$ „Das Taxi ist tatsächlich blau“ betrachtet.
- Wahrscheinlichkeiten aus dem Modell:

$$P(B) = 0.2 \quad P(A|B) = 0.75 \quad P(\bar{A}|\bar{B}) = 0.75$$

- Berechnung der gesuchten Posterior-Wahrscheinlichkeit:

$$\begin{aligned} P(B|A) &= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})} \\ &= \frac{0.75 \cdot 0.2}{0.75 \cdot 0.2 + 0.25 \cdot 0.8} \approx 0.43 \end{aligned}$$

Berechnen Sie gegebenenfalls die Wahrscheinlichkeit! Unterscheiden Sie dabei sorgfältig zwischen den beiden Ereignissen „das Taxi ist blau“ und „das Taxi erscheint blau“!

Aufgabe 2 – Markov Blanket**(30%)**

Das Markov-Blanket $\text{mb}(X)$ einer Zufallsvariable X in einem Bayes-Netz besteht aus den Eltern $E_1, E_2, \dots \in \text{eltern}(X)$, den Kindern $K_1, K_2, \dots \in \text{kinder}(X)$ und den Co-Eltern $C_1, C_2, \dots \in \{\text{eltern}(Y) \mid Y \in \text{kinder}(X)\} \setminus \{X, K_1, K_2, \dots\}$ des Knotens. Außerhalb liegen weitere Zufallsvariablen $Z_1, Z_2, \dots \notin \text{mb}(X) \cup \{X\}$.

- (a) Geben Sie eine möglichst kompakte Faktorisierung der Verbundwahrscheinlichkeit $P(X, E_1, E_2, \dots, K_1, K_2, \dots, C_1, C_2, \dots, Z_1, Z_2, \dots)$ aller im Bayes-Netz enthaltenen Zufallsvariablen an!

$$\begin{aligned} & P(X, E_1, E_2, \dots, K_1, K_2, \dots, C_1, C_2, \dots, Z_1, Z_2, \dots) \\ = & P(X \mid E_1, E_2, \dots) \cdot \prod_n P(E_n \mid \text{eltern}(E_n)) \cdot \prod_n P(K_n \mid \text{eltern}(K_n)) \\ & \cdot \prod_n P(C_n \mid \text{eltern}(C_n)) \cdot \prod_n P(Z_n \mid \text{eltern}(Z_n)) \end{aligned}$$

- (b) Zeigen Sie unter Verwendung des Ergebnisses aus Teilaufgabe (a), dass folgende Aussage gilt:

$$P(X \mid \text{mb}(X), Z_1, Z_2, \dots) = P(X \mid \text{mb}(X))$$

$$\begin{aligned} & P(X \mid \text{mb}(X), Z_1, Z_2, \dots) \\ = & \frac{P(X, E_1, E_2, \dots, K_1, K_2, \dots, C_1, C_2, \dots, Z_1, Z_2, \dots)}{\sum_x P(X = x, E_1, E_2, \dots, K_1, K_2, \dots, C_1, C_2, \dots, Z_1, Z_2, \dots)} \\ & \text{(Beim Kürzen bleiben nur Faktoren übrig, die } X \text{ enthalten.)} \\ = & \frac{P(X \mid E_1, E_2, \dots) \prod_n P(K_n \mid \text{eltern}(K_n))}{\sum_x P(X \mid E_1, E_2, \dots) \prod_n P(K_n \mid \text{eltern}(K_n))} \\ = & P(X \mid \text{mb}(X)) \end{aligned}$$

Aufgabe 3 – Erwartungswerte

(15%)

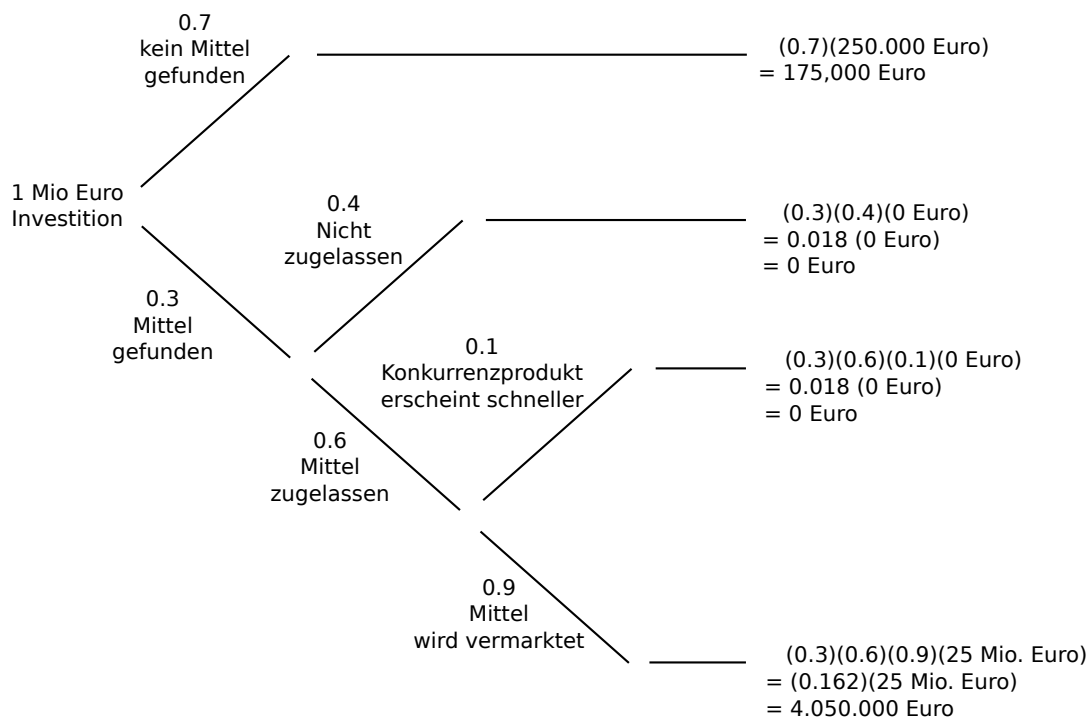
Stellen Sie sich vor, ein Freundin schlägt Ihnen eine Investition in ein Forschungsprojekt vor, in Rahmen dessen ein neues Mittel gegen Glatzköpfigkeit erforscht wird. Sie müssten 1 Mio. Euro investieren. Wahrscheinlich möchten Sie wissen, wie die Erfolgswahrscheinlichkeit aussieht. Sie bekommen folgende Antwort:

Da es sich um ein Forschungsprojekt handelt, liegt die Wahrscheinlichkeit, dass die Forschungsgruppe ein funktionierendes Mittel entdeckt, bei 30%. Falls das Team kein Mittel findet, werden Ihnen 250.000 Euro zurück erstattet. Wenn die Gruppe ein Mittel findet, dann ist die Chance, dass das neue Wundermittel für den Markt zugelassen wird, 60%. Und selbst in diesem Fall, wenn das Mittel wirksam und sicher ist, gibt es noch die 10% Chance, dass ein Mitwettbewerber ein besseres Mittel auf den Markt bringen wird, so dass alle potentiellen Einnahmen wegfallen. Falls alles gut läuft, das Mittel ist sicher und effektiv und es gibt kein Konkurrenzprodukt, können Sie mit Einnahmen von 25 Mio. Euro rechnen.

Sollten Sie investieren?

Hinweis: Sie können annehmen, dass die Ereignisse voneinander unabhängig sind.

Am besten ist es, alle Wahrscheinlichkeiten und zugeordneten Einnahmen oder Verluste in einem Entscheidungsbaum zu sortieren.



Der Erwartungswert der Investition ist die Summe aller möglichen Gewinne multipliziert mit den dazu gehörigen Wahrscheinlichkeiten. Die Wahrscheinlichkeiten können multipliziert werden, da die Ereignisse unabhängig sind.

$$\begin{aligned} \langle \text{Einnahmen} \rangle &= 4.050.000 \text{ Euro} + 0 \text{ Euro} + 0 \text{ Euro} + 175.000 \text{ Euro} \\ &= 4.225.000 \text{ Euro} \end{aligned}$$

Dies ist also im Prinzip eine gute Investition, der Erwartungswert ist deutlich höher als die Investition. Der erwartete Gewinn ist der Erwartungswert minus

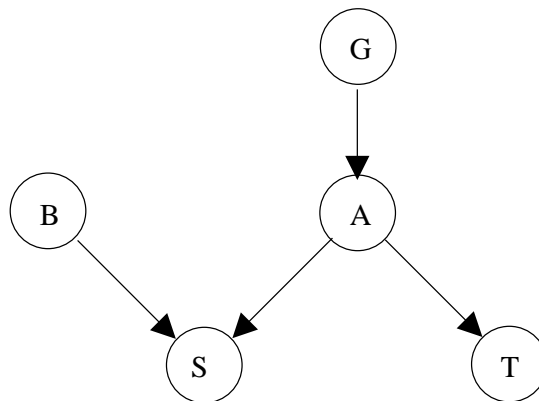
die Anfangsinvestition, also 3.255.000 Euro. Andererseits ist der *wahrscheinlichste* Ausgang, dass sie nur 250.000 Euro zurück gekommen, also 750.000 Euro verlieren. Hedge Fonds oder Multimillionäre können viele solche Investitionen tätigen, manche werden gelingen, andere nicht. Aber das Gesetz der großen Zahlen sagt, dass der Gewinn im Mittel positiv sein wird. Wenn Sie allerdings nur eine einzige Investition tätigen können, sollten Sie sich wohl besser dagegen entscheiden.

Aufgabe 4 – Probabilistische Inferenz

(35%)

Es existieren zwei Krankheiten, die das gleiche Symptom hervorrufen. Folgende Erkenntnisse konnten in wissenschaftlichen Studien festgestellt werden:

- Es wurde ein Gendefekt entdeckt, der bei einem Fünftel der Bevölkerung auftritt ($G = w$). Bei 10% aller Betroffenen führt er zum Ausbruch von Krankheit A ($A = w$). Ohne diesen Defekt ($G = f$) tritt Krankheit A nur in einem von 1000 Fällen auf.
 - Unabhängig von den Genen tritt Krankheit B ($B = w$) bei 2% der Bevölkerung auf.
 - Das Symptom tritt bei jedem 20. Menschen, der keine der beiden Krankheiten hat, auf ($S = w$). Während 50% der Patienten mit (ausschließlich) Krankheit A das Symptom haben, sind es bei Patienten mit (ausschließlich) Krankheit B nur 30%. Wenn beide Krankheiten gleichzeitig auftreten, wird das Symptom zu 90% festgestellt.
 - Für Krankheit A ist ein Test vorhanden. Bei einem erkrankten Menschen fällt er zu 99% positiv aus ($T = w$), bei einem gesunden nur in 0.3% der Fälle.
- (a) Zeichnen Sie ein Bayes-Netz, das zu diesem Modell passt! Die Wahrscheinlichkeitstabellen brauchen Sie hierfür nicht anzugeben.



- (b) Wie groß ist die Wahrscheinlichkeit, dass der Test positiv ausfällt, wenn der Gendefekt vorliegt?

$$P(T = w | G = w) = \sum_{x=w,f} P(T = w | A = x) P(A = x | G = w)$$

$$\begin{aligned}
&= 0.99 \cdot 0.1 + 0.003 \cdot 0.9 \\
&= 0.099 + 0.0027 \\
&= 0.1017
\end{aligned}$$

- (c) Wie wahrscheinlich ist es, an Krankheit A erkrankt zu sein, wenn der Test negativ ausfällt und der Gendefekt vorliegt?

$$\begin{aligned}
P(A = w|T = f, G = w) &= \frac{P(T = f|A = w, G = w)P(A = w|G = w)}{P(T = f|G = w)} \\
&= \frac{P(T = f|A = w)P(A = w|G = w)}{1 - P(T = w|G = w)} \\
&= \frac{0.01 \cdot 0.1}{1 - 0.1017} \\
&\approx 0.00111 = 1.11 \cdot 10^{-3} \approx 0.1\%
\end{aligned}$$

- (d) Wie wahrscheinlich tritt das Symptom auf, wenn man nicht unter dem Gendefekt leidet?

$$\begin{aligned}
P(S = w|G = f) &= \sum_{x=w,f} \sum_{y=w,f} P(S = w, B = x, A = y|G = f) \\
&= \sum_{x=w,f} \sum_{y=w,f} P(S = w|B = x, A = y)P(B = x)P(A = y|G = f) \\
&= P(S = w|B = w, A = w)P(B = w)P(A = w|G = f) \\
&+ P(S = w|B = w, A = f)P(B = w)P(A = f|G = f) \\
&+ P(S = w|B = f, A = w)P(B = f)P(A = w|G = f) \\
&+ P(S = w|B = f, A = f)P(B = f)P(A = f|G = f) \\
&= 0.9 \cdot 0.02 \cdot 0.001 + 0.3 \cdot 0.02 \cdot 0.999 \\
&+ 0.5 \cdot 0.98 \cdot 0.001 + 0.05 \cdot 0.98 \cdot 0.999 \\
&= 0.055453 \approx 5.5\%
\end{aligned}$$