Eine Herleitung des Bayes-Theorems

Wahrscheinlichkeit

Das Bayes-Theorem beruht auf drei als absolut richtig anerkannten Grundsätzen, deren innewohnende Wahrheit keines Beweises bedarf [1]: dem Axiom der Nichtnegativität, dem Axiom der Normiertheit und dem Axiom der Additivität.

Wenn gilt:

(**1**) $0 \leq p(A) \leq 1$ — Axiom der Nichtnegativität

und

(**2**) $p(A) = 1$ für das sichere Ereignis A, — Axiom der Normiertheit

dann heißen die Zahlen p(A) Wahrscheinlichkeiten für das Eintreten des Ereignisses A. [2]

Ist A ein Ereignis, und setzt sich dieses aus einer Menge disjunkter Teilereignisse A_i zusammen als $A = \{A_1, \dots , A_N\}$, dann errechnet sich die Wahrscheinlichkeit für das Eintreten dieses Ereignisses als Summe aller Einzelwahrscheinlichkeiten $p(A_i)$

(**3**) $p(A) = \sum_{i=1}^{N} p(A_i)$. [3] — Axiom der Additivität

Additionssatz

Es seien N eine natürliche Zahl, und e und u natürliche Zahlen aus dem Intervall [1,N]. Das Ereignis A_E sei eine Menge aus e Teilereignissen. Das Ereignis A_U sei eine Menge aus u = N - e Teilereignissen. A_E und A_U seien disjunkt. Der vollständige Ereignisraum sei gegeben durch

$$A_E = \{A_1, \dots , A_e\} \qquad \text{bzw.} \qquad A_U = \{A_{e+1}, \dots , A_N\}$$

Dann lauten nach (3) definitionsgemäß die Wahrscheinlichkeiten für das Eintreten dieser Ereignisse:

---

[1] Duden, „Das Fremdwörterbuch“

[2]) nach: Lambacher Schweizer, „Leistungskurs Stochastik“

[3]) ebd.

$$p(A_E) = \sum_{i=1}^{e} p(A_i) \quad \text{und} \quad p(A_U) = \sum_{i=e+1}^{N} p(A_i).$$

Der aus den beiden Ereignissen A_E und A_U sich zusammensetzende Gesamtraum $A = A_E \cup A_U$ besteht aus der Menge aller N zu den Ereignissen A_E und A_U gehörigen Teilereignisse

$$A = A_E \cup A_U = \{A_1, \dots , A_e\} \cup \{A_{e+1}, \dots , A_N\} = \{A_1, \dots , A_e, A_{e+1}, \dots , A_N\}.$$

Die Wahrscheinlichkeit für einen aus N Teilereignissen bestehenden Gesamtraum A ist definitionsgemäß

$$p(A_E \cup A_U) = \sum_{i=1}^{N} p(A_i).$$

Eine Zerlegung der Summe in Teilsummen führt wieder zurück zu

$$p(A_E \cup A_U) = \sum_{i=1}^{e} p(A_i) + \sum_{i=e+1}^{N} p(A_i),$$

also insgesamt zu

$$p(A_E \cup A_U) = p(A_E) + p(A_U).$$

Es leuchtet ein, dass auf die gleiche Art und Weise jede disjunkte Zerlegung des vollständigen Ereignisraumes A in N einander ausschließender Hauptereignisse zu dem folgenden Additionssatz führt:

(4) $p(A_1 \cup A_2 \cup \dots \cup A_N) = p(A_1) + p(A_2) + \dots + p(A_N)$

Additionssatz für sich ausschließende Hypothesen

Multiplikationssatz

Der Multiplikationssatz für ***bedingte*** Wahrscheinlichkeiten voneinander ***abhängiger*** Ereignisse[4] A und B lautet:

(5) $p(A \cap B) = p(A) * p(B \mid A)$. [5]

Multiplikationssatz für abhängige Ereignisse

[4]) Siehe z. B.: Lambacher Schweizer, „Leistungskurs Stochastik“

[5]) Lies $p(B \mid A)$ wie: „Wahrscheinlichkeit für das abhängige Ereignis B bei gegebenem unabhängigem Ereignis A“.

Herleitung des Theorems

Nach Vorbereitung mit den im vorstehenden Kapitel getroffenen Deduktionen lässt sich das Bayes-Theorem nun herleiten.

Gegeben sei ein unabhängiges Ereignis B. Von diesem abhängig entwickele sich ein weiteres Ereignis A, das aus N disjunkten Teilereignissen bestehe:

$$A = \{A_1, \dots , A_N\}.$$

A und B sollen den Ereignisraumes vollständig darstellen und gegenseitig disjunkt sein. Dann ergibt sich damit

$$B = (A_1 \cap B) \quad \cup \quad \dots \quad \cup \quad (A_N \cap B).$$

Der zugehörige Wahrscheinlichkeitsausdruck

$$p(B) = p\boldsymbol{(}(A_1 \cap B) \quad \cup \quad \dots \quad \cup \quad (A_N \cap B)\boldsymbol{)}$$

ist gemäß Gleichung (4) äquivalent zu

$$p(B) = p(A_1 \cap B) \quad + \quad \dots \quad + \quad p(A_N \cap B).$$

Ersetzt man hierin jeden einzelnen Term ($A_i \cap B$) gemäß (5), dann hat man

$$\textbf{(6)} \qquad p(B) = \sum_{i=1}^{N} p(A_i) * p(B \mid A_i).$$

Weiter lässt sich ($A_i \cap B$) kommutieren und wegen der definitorischen Gleichrangigkeit formal durch den entsprechenden Ausdruck von (5) ersetzen:

$$p(A_i \cap B) = p(B \cap A_i)$$

$$p(A_i) * p(B \mid A_i) = \qquad\qquad\qquad = p(B) * p(A_i \mid B)$$

Aus der letzten Zeile ergibt sich

$$p(A_i \mid B) = \frac{p(A_i)}{p(B)} * p(B \mid A_i).$$

Indem man hierin die Anfangswahrscheinlichkeit p(B) gemäß (6) ersetzt, erhält man das Bayes-Theorem in der folgenden allgemeinen Schreibweise[6]:

[6] Lies $p(A_i \mid B)$ wie „Wahrscheinlichkeit P des i-ten Teilereignisses A_i bei gegebenem Ereignis B.

$$\textbf{(7)}\quad p(A_i \mid B) = \frac{p(A_i) * p(B \mid A_i)}{\sum_{j=1}^{N} p(A_j) * p(B \mid A_j)} \quad – \quad \text{wobei i und j} \in [1,N].$$

Gleichverteilung der Anfangswahrscheinlichkeiten für alle betrachteten Hypothesen ist der Normalfall bei schriftvergleichenden Untersuchungen. Dies bedeutet:

$$p(A_p) = p(A) = p(A_q) \text{ für alle } p \neq q \text{ mit } p, q \in [1,N].$$

Damit schreibt sich Gleichung (7) als

$$p(A_i \mid B) = \frac{p(A) * p(B \mid A_i)}{\sum_{j=1}^{N} p(A) * p(B \mid A_j)},$$

so dass sich p(A) herauskürzt und sich die Hypothesenwahrscheinlichkeiten allein durch die Likelihoods $p(B \mid A_i)$ ergeben:

$$\textbf{(8)}\quad p(A_i \mid B) = \frac{p(B \mid A_i)}{\sum_{j=1}^{N} p(B \mid A_j)}.$$

Vom Theorem zum Algorithmus

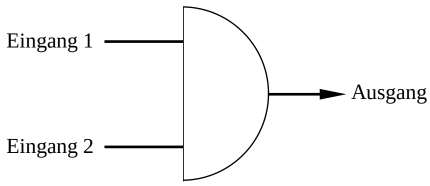
Auch das unabhängige Ereignis B lässt sich zusammengesetzt denken als Menge von M Teilereignissen: $B = \{B_1, \ldots, B_M\}$. Alle Teilereignisse B_j seien in einem Zuge entstanden, so dass sie zugleich manifest nebeneinander stehen. Dies ist gleichbedeutend mit der logischen Und-Verknüpfung, so dass die Likelihood

(9) $$p(B \mid A_i) \quad \text{als} \quad p(\{B_1 \textit{ und } B_2 \textit{ und } B_3 \textit{ und } \ldots\} \mid A_i)$$

geschrieben werden kann.

Die Rechenoperation „Multiplikation“ und die der logischen Und-Verknüpfung sind zueinander äquivalent. Die Äquivalenz zur Multiplikation zeigt sich an der zugehörigen Wahrheitstabelle der logischen Und-Verknüpfung.

Eing. 1		Eing.2		Ausgang
wahr	und	wahr	=	wahr
wahr	und	falsch	=	falsch
falsch	und	wahr	=	falsch
falsch	und	falsch	=	falsch



Sind die Teilereignisse durch Zahlen repräsentiert, kann man darin die Zustände „wahr“ durch die Zahl 0 und „falsch“ durch die Zahl 1 ersetzen. Man erhält auch dann am Ausgang das richtige Ergebnis, wenn man die Eingänge 1 und 2 rechnerisch miteinander multipliziert:

Eing. 1		Eing.2		Ausgang
1	*	1	=	1
1	*	0	=	0
0	*	1	=	0
0	*	0	=	0

Demnach lautet der obige Ausdruck (9), praktikabel als Rechenoperation geschrieben, nunmehr:

(10) $$p(B \mid A_i) = rF_{i,1} * \ldots * rF_{i,j} * \ldots * rF_{i,M} = \prod_{j=1}^{M} rF_{i,j}$$

Darin ist M die Anzahl der unabhängigen Teilereignisse, und $rF_{i,j}$ ist die relative Häufigkeit (engl. „relative frequency“) des j-ten unabhängigen Teilereignisses bei gegebenem i-tem abhängigem Ereignis A_i. Der Operator Π sagt aus, dass die Likelihood $p(B \mid A_i)$ sich als Produkt all dieser A-posteriori-Wahrscheinlichkeiten $rF_{i,j}$ ergibt. [7]

[7] Das geometrische Mittel aus M Werten ist definiert als die M-te Wurzel aus dem Produkt der Werte. Demnach stellt Gleichung 10 die Wahrscheinlichkeit des unabhängigen Ereignisses B bei gegebenem abhängigem Teilereignis A_i als M-te Potenz des geometrischen Mittels der unabhängigen Teilereignisse, sprich: der relativen Häufigkeiten $rF_{i,j}$ dar.

Setzt man in die Bayes-Formel der Gleichung 8 zur Berechnung der Hypothesenwahrscheinlichkeit die Likelihood aus Gleichung 10 ein, so erhält man bei Gleichverteilung der Anfangswahrscheinlichkeiten als Rechenvorschrift:

$$(5) \qquad p(A_i \mid B) = \frac{\prod_{j=1}^{M} rF_{i,j}}{\sum_{i=1}^{N} \prod_{j=1}^{M} rF_{i,j}} = \frac{L_i}{\sum_{i=1}^{N} L_i}$$