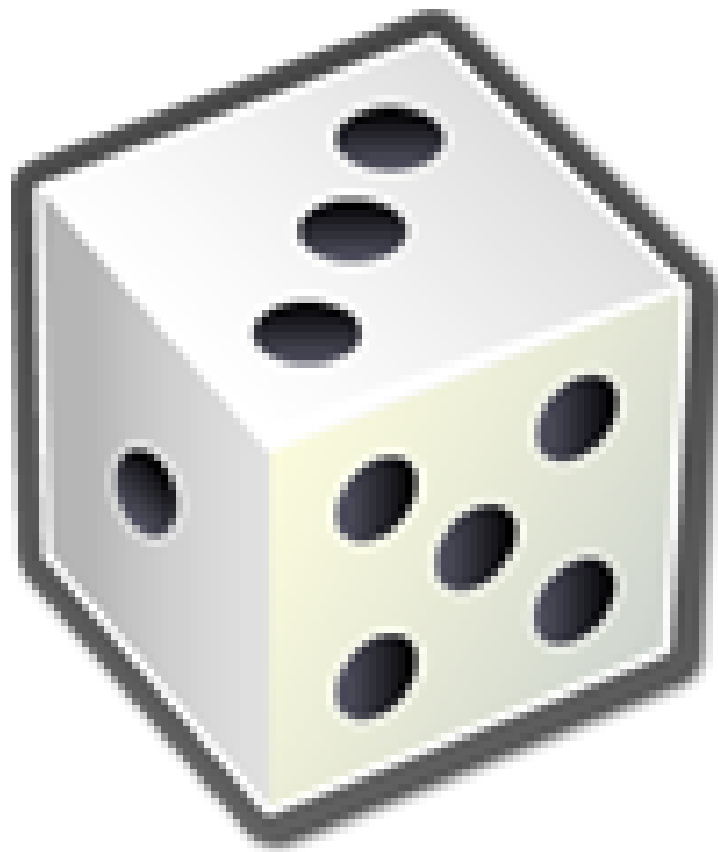


Stochastik

Dr. Johannes Riesterer

11. April 2021



©Johannes Riesterer

Vorwort

Kann jeder Mathematik lernen? Als Antwort auf diese Frage möchte ich auf den interessanten Lebenslauf einer der bedeutendsten Mathematikerinnen aller Zeiten eingehen (Auszug aus Wikipedia):

Emmy Noether war eine deutsche Mathematikerin, die grundlegende Beiträge zur abstrakten Algebra und zur theoretischen Physik lieferte. Insbesondere hat Noether die Theorie der Ringe, Körper und Algebren revolutioniert. Das nach ihr benannte Noether-Theorem gibt die Verbindung zwischen Symmetrien von physikalischen Naturgesetzen und Erhaltungsgrößen an.

Sie zeigte in mathematischer Richtung keine besondere Frühreife, sondern hatte in ihrer Jugend Interesse an Musik und Tanzen. Sie besuchte die Städtische Höhere Töchterschule – das heutige Marie-Therese-Gymnasium – in der Schillerstraße in Erlangen. Mathematik wurde dort nicht intensiv gelehrt. Im April 1900 legte sie die Staatsprüfung zur Lehrerin der englischen und französischen Sprache an Mädchenschulen in Ansbach ab. 1903 holte sie in Nürnberg die externe Abiturprüfung am Königlichen Realgymnasium – dem heutigen Willstätter-Gymnasium – nach.

Inhaltsverzeichnis

1	Notationen	6
2	Diskrete Modelle	7
2.1	Naiver Bayes'scher Spam Filter	9
	Tabellenverzeichnis	10
	Abbildungsverzeichnis	11

1 Notationen

Für eine Menge Ω bezeichnet $\#\Omega$ die Anzahl ihrer Elemente und $\mathcal{P}(\Omega) := \{A \mid A \subset \Omega\}$ die Potenzmenge von Ω .

2 Diskrete Modelle

Definition 1 (Laplace Wahrscheinlichkeit). Sei $\Omega = \{\omega_1, \dots, \omega_n\}$ eine endliche Menge. Für $A \subset \Omega$ definieren wir die Wahrscheinlichkeit durch

$$P(A) := \frac{\#A}{\#\Omega}.$$

Die Elemente $\omega_1, \dots, \omega_n \in \Omega$ nennen wir Elementarereignisse und Teilmengen $A \subset \Omega$ Ereignisse. Die Abbildung

$$P : \mathcal{P}(\Omega) \rightarrow [0, 1]$$

nennen wir die Laplace-Verteilung bzw. Gleichverteilung auf Ω .

Definition 2 (Permutationen und Kombinationen). • $\text{Perm}_k^n(\Omega, m.W.) :=$

$\{\omega_1, \dots, \omega_k \in \Omega^k\}$ Menge aller Permutationen mit Wiederholung.

• $\text{Perm}_k^n(\Omega, o.W.) := \{\omega_1, \dots, \omega_k \in \Omega^k \mid \omega_i \neq \omega_j\}$ Menge aller Permutationen ohne Wiederholung.

• $\text{Kom}_k^n(\Omega, m.W.) := \{\omega_{i_1}, \dots, \omega_{i_k} \in \Omega^k \mid 1 \leq i_1 \leq \dots \leq i_k\}$ Menge aller Kombinationen mit Wiederholung.

• $\text{Kom}_k^n(\Omega, o.W.) := \{\omega_{i_1}, \dots, \omega_{i_k} \in \Omega^k \mid 1 \leq i_1 \leq \dots \leq i_k; \omega_{i_i} \neq \omega_{i_j}\}$ Menge aller Kombinationen ohne Wiederholung.

Lemma 1. • $\#\text{Perm}_k^n(\Omega, m.W.) = n^k = \underbrace{n \cdot n \cdots n}_{k\text{-mal}}$

• $\#\text{Perm}_k^n(\Omega, o.W.) = n_k = n \cdot (n-1) \cdots (n-(k-1)) = \frac{n!}{(n-k)!}$

• $\#\text{Kom}_k^n(\Omega, o.W.) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$

• $\#\text{Kom}_k^n(\Omega, m.W.) = \binom{n+k-1}{k}$

Beweis. □

Definition 3. Diskrete Wahrscheinlichkeitsverteilung Es sei Ω eine (nicht leere) abzählbare Menge. Eine Abbildung $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ heißt diskrete Wahrscheinlichkeitsverteilung (Wahrscheinlichkeitsmaß), falls gilt:

$$P(\Omega) = 1$$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n), \text{ mit } A_i \cap A_j = \emptyset \text{ für } i \neq j$$

Beispiel 1. Beispiel: Laplace Wahrscheinlichkeit Ω endlich und $P(A) = \frac{\#A}{\#\Omega}$.

Definition 4. Bedingte Wahrscheinlichkeit Für $A, B \subset \mathcal{P}(\Omega)$ und $P(B) > 0$ heißt

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

die bedingte Wahrscheinlichkeit (von A unter B).

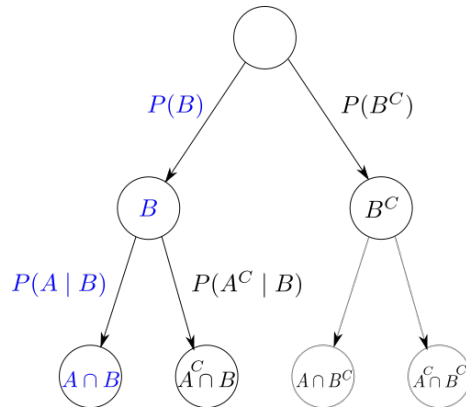


Abbildung 1: Quelle: Wikipedia

Satz 1. *Satz der totalen Wahrscheinlichkeit* Für eine Zerlegung $\Omega = \bigcup_{j=1}^n B_j$, mit $B_i \cap B_k = \emptyset$ für $i \neq k$

$$P(A) = \sum_{j=1}^n P(A | B_j) \cdot P(B_j)$$

Beweis. $A = A \cap B \cup A \cap \bar{B}$ und $A \cap B \cap A \cap \bar{B} = \emptyset$.

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(B) \cdot P(A | B) + P(\bar{B}) \cdot P(A | \bar{B})$$

□

Satz 2. *Satz von Bayes* Für $A, B \subset \mathcal{P}(\Omega)$ mit $P(B) > 0$ gilt

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

Beweis.

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

□

Definition 5. *Stochastische Unabhängigkeit* Zwei Ereignisse A, B heißen stochastisch unabhängig, falls

$$P(A \cap B) = P(A) \cdot P(B)$$

gilt. Somit ist auch $P(A|B) = P(A)$ und $P(B|A) = P(B)$.

2.1 Naiver Bayes'scher Spam Filter

Gegeben ist eine E-Mail E . Wir möchten anhand des Vorkommens bestimmter Wörter A_1, \dots, A_n in der Mail entscheiden, ob es sich um eine erwünschte Mail H oder eine unerwünschte Mail S (Ham or Spam) handelt. (Typische Wörter wären zum Beispiel "reichwerden", "önlencasino"...). Aus einer Datenbank kann man das Vorkommen dieser Wörter in Spam und Ham Mails zählen und damit empirisch die Wahrscheinlichkeiten $P(A_i|S)$ und $P(A_i|H)$ des Vorkommens dieser Wörter in Spam und Ham Mails ermitteln. Wir gehen davon aus, dass es sich bei der Mail prinzipiell mit Wahrscheinlichkeit $P(E = S) = P(E = H) = \frac{1}{2}$ um eine erwünschte Mail H oder eine unerwünschte Mail S handeln kann. Wir machen zudem die (naive) Annahme, dass das Vorkommen der Wörter stochastisch unabhängig ist, also

$$\begin{aligned} P(A_1 \cap \dots \cap A_n | S) &= P(A_1 | S) \cdots P(A_n | S) \\ P(A_1 \cap \dots \cap A_n | H) &= P(A_1 | H) \cdots P(A_n | H) \end{aligned}$$

gilt.

Mit der Formel von Bayes und der totalen Wahrscheinlichkeit können wir somit berechnen

$$\begin{aligned} P(E = S | A_1 \cap \dots \cap A_n) &= \frac{P(A_1 \cap \dots \cap A_n | S) \cdot P(S)}{P(A_1 \cap \dots \cap A_n)} \\ &= \frac{P(A_1 | S) \cdots P(A_n | S) \cdot P(S)}{P(A_1 \cap \dots \cap A_n | H) + P(A_1 \cap \dots \cap A_n | S)} \\ &= \frac{P(A_1 | S) \cdots P(A_n | S) \cdot P(S)}{P(A_1 | H) \cdots P(A_n | H) + P(A_1 | S) \cdots P(A_n | S)} \end{aligned}$$

Bemerkung: $P(E = H | A_1 \cap \dots \cap A_n) = 1 - P(E = S | A_1 \cap \dots \cap A_n)$

Tabellenverzeichnis

Abbildungsverzeichnis

1	Quelle: Wikipedia	8
---	-----------------------------	---