Stochastik für Informatiker

Dr. rer. nat. Johannes Riesterer

Highlight

Beispiel: Hash Kollision

Beim Hashing werden zufällig k Daten auf n Speicherplätze verteilt. Bezeichnen wir mit $A_{k,n}$ die Möglichkeiten der Mehrfachbelegungen von Feldern, so ist das komplementäre Ereignis $A_{k,n}^c = Var_k^n(\Omega, o.W.)$, wobei Ω die Menge der verfügbaren Speicherplätze darstellt.

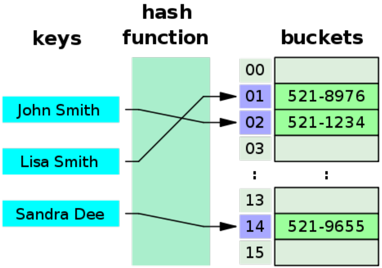


Figure: Quelle: Wikipedia

Beispiel: Hash Kollision

$$P(A^c_{k,n}) = \frac{\# Var^n_k(\Omega, o.W.)}{\# Var^n_k(\Omega, m.W.)} =$$

$$\frac{n_k}{n^k} = \frac{n \cdot (n-1) \cdots (n-(k-1))}{n^k} = \prod_{i=0}^{k-1} (1 - \frac{i}{n})$$

$$= \exp(\sum_{i=0}^{k-1} \ln(1 - \frac{i}{n})) \leq \exp(\sum_{i=0}^{k-1} (-\frac{i}{n}))$$

$(ln(1-x) \leq -x$ für $x < 1)$

$$= \exp(-\frac{(k-1)k}{2n})$$

Beispiel: Hash Kollision (Geburtstags-Paradoxon)

Für $n = 365$ und $k = 23$ ist damit $P(A_{k,n}) > \frac{1}{2}$. Die Wahrscheinlichkeit, dass bei einer Gruppe von mehr als 23 Leuten zwei Leute am gleichen Tag Geburtstag haben, ist also größer als $\frac{1}{2}$.

Axiome von Kolmogorov

σ-Algebra

Es sei Ω eine Menge und $\mathcal{A} \subset \mathcal{P}(\Omega)$ ein System von Teilmengen (= Ereignissen). $\mathcal{A}$ heißt σ-Algebra (Sigma-Algebra) falls gilt:

$$(i)\ \Omega, \emptyset \in \mathcal{A}$$
$$(ii)\ A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$$
$$(iii)\ A_i \in \mathcal{A} \Rightarrow \bigcup_i A_i \in \mathcal{A}$$

$(A^c = \Omega \setminus A)$

Interpretation

Die Grundmenge Ω ist ein Ereignis. Das nicht-Eintreffen eines Ereignisses ist ein Ereignis. Die Vereinigung von Ereignissen ist ein Ereignis.

Axiome von Kolmogorov

Axiome von Kolmogorov

Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathcal{A}, P)$ bestehend aus der Grundmenge Ω, einer σ-Algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$ und einer Abbildung $P : \mathcal{A} \to [0,1]$

$$(i)\ P(\Omega) = 1$$

$$(ii)\ P\left(\bigcup_i A_i\right) = \sum_i P(A_i),\ \text{mit } A_i \cap A_j = \emptyset \text{ für } i \neq j$$

Die Elemente von Ω werden elementare Ereignisse und die von $\mathcal{A}$ Ereignisse genannt. Mengen M mit $P(M) = 0$ werden Nullmengen genannt. Die Abbildung P wird Wahrscheinlichkeitsmaß genannt.

Interpretation

Die Grundmenge und das Wahrscheinlichkeitsmaß wird durch das betrachtete Experiment definiert. Die Menge der Ergebnisse $\mathcal{A}$ beschreibt die Fragestellungen, an denen wir im Rahmen des Experiments interessiert sind. Die Grundmenge ist ein sicheres Ereignis. Das Experiment liefert also sicher einen definierten Ausgang. Eine Münze bleibt Beispielsweise nicht auf der Kante stehen, wenn nur Kopf oder Zahl modelliert ist. Die Wahrscheinlichkeit des Eintreffens von überschneidungsfreien Ereignissen addiert sich.

Würfel

$\Omega = \{1, 2, 3, 4, 5, 6\}$. $\mathcal{A} := \mathcal{P}(\Omega)$ entspricht dem Laplace-Experiment. Wir sind an allen möglichen Fragen interessiert. $\mathcal{A} := \{\{1, 2, 3\}, \{4, 5, 6\}, \{1, 2, 3, 4, 5, 6\}, \{\}\}$ entspricht dem Interesse daran, ob die Augenzahl größer oder kleiner gleich 3 ist.

Diskrete Wahrscheinlichkeitsverteilungen

Diskreter Wahrscheinlichkeitsraum

Ein diskreter Wahrscheinlichkeitsraum ist ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, bei dem die Grundmenge Ω abzählbar.

Beispiel: Laplace Experiment

Ω endlich, $\mathcal{A} = \mathcal{P}(\Omega)$, und $P(A) := \frac{\#A}{\#\Omega}$.

Lemma

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Dann ist für $A \in \mathcal{A}$

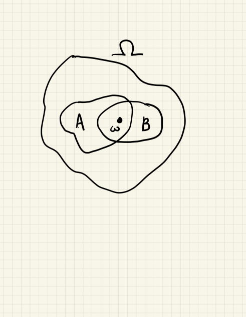
$$P(A^c) = 1 - P(A)$$
$$P(\emptyset) = 0$$

Herleitung der bedingten Wahrscheinlichkeit

$$\tilde{\Omega} := B$$

$$\tilde{\mathcal{A}} := \{C \cap B \mid C \in \mathcal{A}\}$$

$$\tilde{P}(X) = \frac{P(X)}{P(B)} \quad \text{mit } X \in \tilde{\mathcal{A}}$$



Bedingte Wahrscheinlichkeit

Für $A, B \in \mathcal{A}$ und $P(B) > 0$ heißt

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

die bedingte Wahrscheinlichkeit (von A unter B).

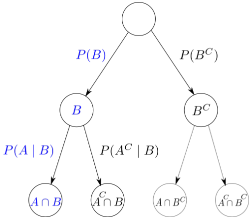


Figure: Quelle: Wikipedia

Diskrete Wahrscheinlichkeitsverteilung

Satz der totalen Wahrscheinlichkeit

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

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

Diskrete Wahrscheinlichkeitsverteilung

Satz von Bayes

Für $A, B \in \mathcal{A}$ mit $P(B) > 0$ gilt

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

Beweis

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

Stochastische Unabhängigkeit

Zwei Ereignisse A, B heißen stochastisch unabhängig, falls

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

gilt. Gleichbedeutend damit ist $P(A|B) = P(A)$ und $P(B|A) = P(B)$.

Diskrete Wahrscheinlichkeitsverteilung

Naiver Bayes'scher Spam Filter

Gegeben ist eine E-Mail E. Wir möchten anhand des Vorkommens bestimmter Wörter $A_1, \ldots A_n$ 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 "reich", "casino", "Vergrösserung" ...)

Naiver Bayes'scher Spam Filter

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.

Naiver Bayes'scher Spam Filter

Wir machen zudem die (naive) Annahme, dass das Vorkommen der Wörter stochastisch unabhängig ist, also

$$P(A_1 \cap \cdots \cap A_n|S) = P(A_1|S) \cdot P(A_2|S) \cdots P(A_n|S)$$
$$P(A_1 \cap \cdots \cap A_n|H) = P(A_1|H) \cdot P(A_2|H) \cdots P(A_n|H)$$

gilt.

Diskrete Wahrscheinlichkeitsverteilung

Naiver Bayes'scher Spam Filter

Mit der Formel von Bayes, der totalen Wahrscheinlichkeit und kürzen, da $P(S) = P(H)$ können wir somit berechnen

$$P(E = S|A_1 \cap \cdots \cap A_n) = \frac{P(A_1 \cap \cdots \cap A_n|S) \cdot P(S)}{P(A_1 \cap \cdots \cap A_n)}$$

$$= \frac{P(A_1|S) \cdots P(A_n|S)}{P(A_1 \cap \cdots \cap A_n|H) + P(A_1 \cap \cdots \cap A_n|S)}$$

$$= \frac{P(A_1|S) \cdots P(A_n|S)}{P(A_1|H) \cdots P(A_n|H) + P(A_1|S) \cdots P(A_n|S)}$$

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