1 Theorie

1.1 Wahrscheinlichkeitsraum

1.1.1 Definition

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

$$(i)\ P(\Omega) = 1$$
$$(ii)\ P\Big(\bigcup_i A_i\Big) = \sum_i P(A_i), \text{ mit } A_i \cap A_j = \emptyset \text{ für } i \neq j$$

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.

1.1.2 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:

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

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

1.1.3 Diskreter Wahrscheinlichkeitsraum

Ein diskreter Wahrscheinlichkeitsraum ist ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, bei dem die Grundmenge Ω abzählbar ist und die Menge der Ereignisse $\mathcal{A} := \mathcal{P}(\Omega)$ der Potenzmenge entspricht.

1.1.4 Laplace Experiment

Ein Laplace-Experiment ist ein Zufallsexperiment bei dem der Ereignisraum Ω endlich viele Elemente und ein Ereignis $A \subseteq \Omega$ die Wahrscheinlichkeit $P(A) = \frac{\#A}{\#\Omega}$ hat.

1.2 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).

1.3 Spamfilter / Satz von Bayes

1.3.1 Satz der totalen Wahrscheinlichkeit

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

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

1.3.2 Satz von Bayes

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

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

1.3.3 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 $P(B|A) = P(B)$.

1.3.4 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 handelt.

Aus einer Datenbank kann man das Vorkommen dieser Wörter in allen E-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

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

gilt.

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

$P(E = S|A_1 \cap \cdots \cap A_n)$ `("E =" wird im folgenden weggelassen)`

$$= \frac{P(A_1 \cap \cdots \cap A_n|S) \cdot P(S)}{P(A_1 \cap \cdots \cap A_n)} \quad \texttt{(-> Satz von Bayes)}$$

$$= \frac{P(A_1|S) \cdots P(A_n|S) \cdot P(S)}{P(A_1 \cap \cdots \cap A_n|H) + P(A_1 \cap \cdots \cap A_n|S)} \quad \frac{\texttt{(-> Stoch. Unabhängigkeit)}}{\texttt{(-> Satz der totalen Wahrscheinlichkeit)}}$$

$$= \frac{P(A_1|S) \cdots P(A_n|S) \cdot P(S)}{P(A_1|H) \cdots P(A_n|H) + P(A_1|S) \cdots P(A_n|S)} \quad \frac{}{\texttt{(-> Stoch. Unabhängigkeit)}}$$

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

1.4 Zufallsvariablen

1.4.1 Allgemeine Zufallsvariable

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $(\Omega', \mathcal{A}')$ ein Messraum. Eine Zufallsvariable ist eine Abbildung

$$X : \Omega \to \Omega'$$

so dass für alle Ereignisse $A' \in \mathcal{A}'$

$$X^{-1}(A') \in \mathcal{A}$$

ein Ereignis in $\mathcal{A}$ ist. Urbilder von Ereignissen sind also Ereignisse.

1.4.2 Messraum

Ein Messraum ist ein Paar $(\Omega, \mathcal{A})$ bestehend aus einer Menge Ω und einer Sigma-Algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$.

1.4.3 Reelle Zufallsvariable

Unter einer reellen Zufallsvariable verstehen wir eine Zufallsvariable

$$X : \Omega \to \mathbb{R}^n$$
$$X(\omega) := \Big(X_1(\omega), \cdots, X_n(\omega)\Big),$$

wobei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum ist und $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ der $\mathbb{R}^n$ zusammen mit der Borel'schen Sigma-Algebra ist.

1.5 Erwartungswert

1.5.1 Definition

Für eine reelle, integrierbare Zufallsvariable X ist der Erwartungswert definiert durch

$$\mathbb{E}(X) := \int_\Omega X \, dP .$$

Ist $(\Omega, \mathcal{A}, P)$ ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine eindimensionale reelle Zufallsvariable, so ist

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$$

1.5.2 Eigenschaften

Sind $X, Y : \Omega \to \mathbb{R}^n$ reelle, integrierbare Zufallsvariablen und $a, b \in \mathbb{R}$ konstant, so gilt:

$$\begin{aligned}
&\mathbb{E}(a \cdot X \pm b \cdot Y) = a \cdot \mathbb{E}(X) \pm b \cdot \mathbb{E}(Y) \\
&\forall x \in \Omega : X(x) \leq Y(x) \Rightarrow \mathbb{E}(X) \leq \mathbb{E}(Y) \\
&X, Y \text{ stoch. unabhängig} \Rightarrow \mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) \\
&\mathbb{E}(1_A) = P(A)
\end{aligned}$$

1.6 Varianz

Für eine reelle Zufallsvariable X ist die Varianz definiert durch

$$\mathbb{V}(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

1.7 Verteilungen

1.7.1 Normalverteilung

Die Normalverteilung $N(\mu, \sigma^2)$ auf $\mathbb{R}$ ist definiert durch

$$\begin{aligned}
&\text{Dichte: } f(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \\
&\Rightarrow \text{Verteilung: } F(x) = N(\mu, \sigma^2)(-\infty, x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2} dt
\end{aligned}$$

Erwartungswert und Varianz bei $X \sim N(\mu, \sigma^2)$:

$$\begin{aligned}
\mathbb{E}(X) &= \mu \\
\mathbb{V}(X) &= \sigma^2
\end{aligned}$$

1.7.2 Verteilungsfunktion

Für eine reelle Zufallsvariable X heißt

$$\begin{aligned}
&F_X : \Omega \to [0, 1] \\
&F_X(x) := P(X \leq x) := P_X((-\infty, x)) = P(X^{-1}(-\infty, x))
\end{aligned}$$

Verteilungsfunktion von X.

1.7.3 Gleichverteilung

Die Gleichverteilung $U(a,b)$ auf einem Intervall $(a,b) \subset \mathbb{R}$ ist definiert durch

$$\text{Dichte: } f(x) := \frac{1_{(a,b)}}{|b-a|}$$

$$\text{Verteilung: } F(x) = P_f((-\infty,x)) = \int_{-\infty}^{x} \frac{1_{(a,b)}}{|b-a|} dt$$

$$= \begin{cases} 0 \text{ für } x \leq a \\ \frac{x-a}{|b-a|} \text{ für } a \leq x \leq b \\ 1 \text{ für } x \geq b \end{cases}$$

Erwartungswert und Varianz bei $X \sim U(a,b)$:

$$\mathbb{E}(X) = \frac{a+b}{2}$$

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{1}{3}\frac{b^3-a^3}{b-a} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{1}{12}(b-a)^2$$

1.7.4 Dichte

Sei $\Omega \subset \mathbb{R}^n$ und $(\Omega, \mathcal{A})$ ein Messraum, wobei alle $A \in \mathcal{A}$ Lebesgue-messbar sind. Eine Funktion $f : \Omega \to \mathbb{R}$ heißt Dichte, falls für ihr Lebesgue-Integral $\int_\Omega f d\mu = 1$ gilt.

1.8 Schwaches Gesetz der großen Zahlen

1.8.1 Definition

Seien $X_i : \Omega \to \mathbb{R}$ unabhängige, reelle Zufallsvariablen mit $\mathbb{E}(X_i) = \mu < \infty$ und $\mathbb{V}(X_i) = \sigma < \infty$, dann gilt

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \leq \frac{\sigma}{n \cdot \epsilon^2} \xrightarrow[n\to\infty]{} 0$$

(stochastische Konvergenz).

1.8.2 Bedeutung

Das schwache Gesetz der großen Zahlen besagt, dass das arithmetische Mittel einer großen Stichprobe einer Zufallsvariable mit einer beliebig kleinen Wahrscheinlichkeit dem Erwartungswert der Zufallsvariable entspricht.

Gegenteilige (äquivalente) Formulierung:
Die Wahrscheinlichkeit, dass die Differenz zwischen beobachteter relativer Häufigkeit und theoretischer Wahrscheinlichkeit kleiner ist als eine beliebig kleine positive Zahl ϵ, ist für eine unendlich große Stichprobe praktisch 1.

1.9 Zentraler Grenzwertsatz

1.9.1 Definition

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X_n : \Omega \to \mathbb{R}$ eine folge stochastisch unabhängiger, identisch verteilter, reeller Zufallsvariablen mit $\mathbb{E}(X_n) = \mu$ und $\mathbb{V}(X_n) = \sigma^2$. Dann gilt für das arithmetische Mittel $S_n := \frac{1}{n}\sum_{i=1}^{n} X_i$

$$P_{\frac{\sqrt{n}}{\sigma}(S_n-\mu)} \to P_{N(0,1)}$$

wobei $P_{N(0,1)}$ das Wahrscheinlichkeits-Maß mit der Dichte $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ ist.

1.9.2 Bedeutung

Die Summe von n identisch verteilten, stochastisch unabhängigen Zufallsvariablen ist näherungsweise normalverteilt.

Beispiel Würfel:
Die Augensumme von $n \to \infty$ Würfeln ist normalverteilt, wenn alle Würfel von einander stochastisch unabhängig und gleichverteilt sind.

1.10 Schätzer

1.10.1 Ausgangslage

Angenommen man findet einen Apparat, der zufällig Zahlen in einem Intervall $[0, \rho]$ ausgibt. Anhand von Beobachtungen der Zahlen möchte man ρ schätzen. Wir machen die Annahme, dass alle Zahlen in dem Intervall gleich wahrscheinlich auftreten und nehmen n Stichproben $X_1, \cdots, X_n$. Einen Schätzer für ρ bezeichnen wir mit T_n.

1.10.2 Maximalwert-Schätzer

Eine einfache und einleuchtende Idee ist es, ρ durch die größte beobachtete Zahl zu schätzen, also $T_n^{max} := \max(X_1, \cdots, X_n)$.
Dieser Schätzer konvergiert für $n \to \infty$ gegen ρ, also $P(|T_n^{max} - \rho| \geq \epsilon) \xrightarrow[n\to\infty]{} 0$.

1.10.3 Erwartungswert-Schätzer

Da das Auftreten der Zahlen gleich wahrscheinlich ist, ist der Erwartungswert des Zufallsexperiments $\rho/2$. Unter Berufung auf das schwache Gesetz der großen Zahlen erscheint der Schätzer $T_n^E := 2 \cdot \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$ sehr plausibel.
Dieser Schätzer konvergiert für $n \to \infty$ gegen ρ, also $P(|T_n^E - \rho| \geq \epsilon) \xrightarrow[n\to\infty]{} 0$.