

Mathe Helfer

Simon Roske

7. Dezember 2023

Inhaltsverzeichnis

1	Zahlen	3
1.1	Die natürlichen Zahlen \mathbb{N}	3
1.2	Die ganzen Zahlen \mathbb{Z}	3
1.3	Rationale und irrationale Zahlen \mathbb{Q} , $\mathbb{R} \setminus \mathbb{Q}$	3
1.4	Die reellen Zahlen \mathbb{R}	4
1.5	Komplexe Zahlen \mathbb{C}	4
1.6	Besondere Zahlen	5
1.6.1	π - Die Kreiszahl	5
1.6.2	e - Euler's Zahl	5
1.6.3	i - die imaginäre Zahl	6
2	Grundlagen & Rechengesetze	6
2.1	Kommutativgesetz	6
2.2	Assoziativgesetz	6
2.3	Distributivgesetz	6
2.4	Die binomischen Formeln	6
2.5	Potenzen	7
2.6	Wurzeln	7
2.7	Logarithmus	7
3	Geometrie	8
3.1	Trigonometrie	8
3.2	Polarkoodinaten	10
3.2.1	Komplexe Zahlen mittels Polarkoodinaten	11
3.3	Bogenmaß und Radian	11
3.4	Phytagoras	12
4	Statistik	12
4.1	Durchschnitt	12
4.2	Modus	12
4.3	Median	12
4.4	Erwartungswert	12
4.5	Varianz	12
4.6	Standardabweichung	13
4.7	Korrelation	13
4.8	Pearson's chi-square test	13
5	Wahrscheinlichkeit	13
5.1	Abhängige und unabhängige Variable	13
5.2	Kolmogorow' Axiome	13
5.3	Wahrscheinlichkeitskonzepte	13
5.4	Wahrscheinlichkeitsverteilungen	14
5.4.1	Normalverteilung	14
5.4.2	Exponentialverteilung	15

5.4.3	Gleichverteilung	15
5.4.4	Poissonverteilung	16
5.4.5	t-Verteilung oder Student-t-Verteilung	16
5.4.6	χ^2 Verteilung	17
5.4.7	Binominalverteilung	17
5.4.8	Bernoulliverteilung	17
5.5	Bayes	18
6	Algebra	18
6.1	Determinante	18
7	Analysis	19
7.1	Funktionen/Abbildungen	19
7.1.1	Definitionsbereich und Wertebereich	19
7.1.2	Spezielle Funktionen	19
7.1.3	Inversibilität	20
7.1.4	Nullstellen	20
7.2	Ableitungen	21
7.2.1	Spezielle Ableitungen	21
7.2.2	Ableitungsregeln	21
7.2.3	Partielle Ableitungen ∂	21
7.3	Polynome	22
7.4	Integrale	22
7.5	Der Grenzwert	22
7.6	Differentialgleichungen	22
7.7	Divergenz	22
7.7.1	Kullback-Leibler	22
7.7.2	Jensen-Shannon	22
7.8	Lösungsstrategien	22
7.9	Regression	23
7.9.1	Lineare Regression	23
7.9.2	Polynomialae Regression	23
7.9.3	Logistische Regression	23
8	Modellierung	23

Abbildungsverzeichnis

1	short	9
---	-----------------	---

1 Zahlen

1.1 Die natürlichen Zahlen \mathbb{N}

Der Zahlenbereich der natürlichen Zahlen bildet das Zählen als natürlichen Prozess ab. Ob die 0 Teil von \mathbb{N} ist, ist Definitionssache und nicht überall gleich. Angenommen, das sei nicht der Fall, dann gilt:

$$\mathbb{N} = \{1, 2, 3, 4, \dots, n, n+1, \dots\}.$$

Wie kann man mit natürlichen Zahlen rechnen? Man darf uneingeschränkt addieren und multiplizieren. Man sagt \mathbb{N} ist bezüglich der Addition und Multiplikation abgeschlossen. Alle anderen Rechenoperationen sind nicht uneingeschränkt durchführbar, da negative Zahlen nicht unter die Natürlichen fallen. Eine Untermenge von \mathbb{N} ist die Menge der **Primzahlen**, definiert als:

$$\mathbb{P} = \{1, 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, \dots\}.$$

Primzahlen sind nicht durch 1 und sich selbst teilbar!

1.2 Die ganzen Zahlen \mathbb{Z}

Erweitert man den Zahlenbereich der natürlichen Zahlen mit negativen Zahlen, hat man die ganzen Zahlen:

$$\mathbb{Z} = \{\dots - 3, -2, -1, 0, 1, 2, 3, \dots\}.$$

Nun kann man auch uneingeschränkt subtrahieren.

1.3 Rationale und irrationale Zahlen \mathbb{Q} , $\mathbb{R} \setminus \mathbb{Q}$

Will man uneingeschränkt dividieren, braucht man Bruchzahlen:

- \mathbb{Q}_+ enthält alle positiven Brüche: $\mathbb{Q}_+ = \{\frac{a}{b} \mid a, b \text{ sei eine natürliche Zahl und } b \neq 0\}$.

Nimmt man negative Brüche dazu, hat man die rationalen Zahlen.

- $\mathbb{Q} = \{\frac{a}{b} \mid a, b \text{ sei eine ganze Zahl und } b \neq 0\}$.
- In \mathbb{Q} sind alle Grundrechenarten uneingeschränkt ausführbar.
- \mathbb{Q} enthält alle positiven und negativen Brüche, sowie alle abbrechenden Dezimalbrüche z.B. $-3,75$ und periodischen Dezimalbrüche (z.B. $0,66666\dots 0,66666\dots$).

Bei den rationalen Zahlen ist eines nicht vollständig erlaubt: Das Wurzelziehen, da es zu unendlichen Zahlen führt, die sich nicht als Bruch darstellen lassen, den **irrationalen Zahlen**: $\sqrt{2} = 1,41421356\dots$

1.4 Die reellen Zahlen \mathbb{R}

Vereinigt man die rationalen und die irrationalen Zahlen, erhält man die reellen Zahlen \mathbb{R} . Jedoch: Wurzelziehen aus negativen Zahlen ist nicht definiert. Zum Beispiel ist $\sqrt{-4}$ nicht definiert. Solche Zahlen sind nicht in \mathbb{R} enthalten.

1.5 Komplexe Zahlen \mathbb{C}

Eine komplexe Zahl - z - ist beschreibbar als reelles Zahlenpaar: $x + iy | x, y \in \mathbb{R}, i = \sqrt{-1}$. Wichtig ist das mit der imaginären Zahl 1.6.3 das ziehen von negativen Wurzeln möglich ist. Man kann für c auch (x,y) schreiben, wobei x der Realteil und y der Imaginärteil ist. So kann die Menge der komplexen Zahlen (\mathbb{C}) auch in Form der reellen Zahlenpaare (\mathbb{R}^2) geometrisch auf der komplexen Ebene (auch Gaußsche Zahlenebene) dargestellt werden, siehe Bild unten.

Analog definiert man Addition wie folgt:

$$\begin{aligned} z_1 + z_2 &= (x_1 + x_2) + (y_1 + y_2) \cdot i \\ z_1 - z_2 &= (x_1 - x_2) + (y_1 - y_2) \cdot i \end{aligned}$$

und Multiplikation so:

$$\begin{aligned} z_1 * z_2 &= (x_1 + x_2) * (y_1 - y_2) \cdot i = x_1 y_1 + x_1 y_2 \cdot i + y_1 x_2 \cdot i + x_2 y_2 \cdot i^2 \\ z_1 * z_2 &= (x_1 y_1 - x_2 y_2) + (x_1 y_2 + y_1 x_2) \cdot i \end{aligned}$$

Ist z eine komplexe Zahl, so ist z^* die komplexe Konjugation. In der Darstellung unten würde der Realteil(z) hierbei gespiegelt. Als Spezialfall gilt: $i^* = -i$, sodass $z^* = x - y \cdot i$. Multipliziert mit dem komplexen Konjugat ergibt sich der Betrag von z .

Den Betrag einer komplexen Zahl ergibt sich durch:

$$|z| = \sqrt{x^2 + y^2}$$

Die Division ist definiert als: $\frac{z_1}{z_2} = \frac{z_1}{z_2} \cdot \frac{z_2^*}{z_2^*}$. Das ist nicht sehr angenehm und kann vermieden werden, wenn es nicht unbedingt sein muss. Sollte kein Weg daran vorbeiführen, so multipliziere die Faktoren des Nenners und des Zählers je für sich und vereinfache so gut es geht mit der Definition von i .

Eine andere Darstellung ist mithilfe von Polarkoordinaten (3.2) möglich:

$$z = a + i * b \mid a, b, r \in \mathbb{R}, \theta \in [0, 2\pi] \Leftrightarrow z = r * (\cos(\theta) + i * \sin(\theta)) \Leftrightarrow r * e^{i\theta}.$$

1.6 Besondere Zahlen

1.6.1 π - Die Kreiszahl

3.1415926535...damit beginnt die irrationale Zahl π . π beschreibt das Verhältnis von Umfang zu Durchmesser eines Kreises. Es gibt viele Formeln, in denen π die tragende Rolle spielt:

Kreisumfang
$U = \pi \cdot d = 2 \cdot \pi \cdot r$
Kreisfläche
$F = \pi \cdot r^2$
Volumen
$V = \frac{4}{3} \cdot \pi \cdot r^3$

1.6.2 e - Euler's Zahl

Die Eulersche Zahl ist die Basis des natürlichen Logarithmus, also $\ln(e) = 1$. Die Eulersche Zahl kann beschrieben werden durch $e = 2,71828...$, aber ähnlich wie für π gibt es keine exakte Lösung. Die Eulersche Zahl wurde nach dem Schweizer Mathematiker und Physiker Leonhard Euler 1707 – 1783 benannt. Unter anderem beinhaltet die natürliche Exponentialfunktion $f(x) = e^x$ die Eulersche Zahl. Für uns ist e vor allem wichtig, weil man anhand von der Eulerschen Formel $e^{j\phi} = \cos(\phi) + j * \sin(\phi)$ die komplexen Zahlen in der Exponentialform darstellen kann, was die Berechnungen erheblich erleichtert.

Außerdem gilt $\partial e^t / \partial t = e^t$ was bedeutet, dass die Position von e gleich der Ableitung, also der Geschwindigkeitsänderung ist. Außerdem ist dieser Zusammenhang besonders schön:

¹8979323846 2643383279 5028841971 6939937510 5820974944 5923078164 0628620899
8628034825 3421170679 8214808651 3282306647 0938446095 5058223172 5359408128
4811174502 8410270193 8521105559 6446229489 5493038196 4428810975 6659334461
2847564823 3786783165 2712019091 4564856692 3460348610 4543266482 1339360726
0249141273 7245870066 0631558817 4881520920 9628292540 9171536436 7892590360
0113305305 4882046652 1384146951 9415116094 3305727036 5759591953 0921861173
8193261179 3105118548 0744623799 6274956735 1885752724 8912279381 8301194912

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = 2,718281828459\dots$$

Damit ist e der Grenzwert (7.5) der (exponentiell) steigenden Zahlenfolge.

1.6.3 i - die imaginäre Zahl

$$i^2 = -1$$

$$i^3 = -i$$

$$i^4 = 1$$

$$i^{-1} = -i.$$

2 Grundlagen & Rechengesetze

Eine Verknüpfung lässt sich definieren als die Art und Weise wie 2 Objekte ein Drittes definieren. Die Verknüpfung wird abstrakt mit ' \circ ' ausgedrückt.

2.1 Kommutativgesetz

Die Reihenfolge der Verknüpfung spielt keine Rolle, e.g. $2 \circ 3 = 3 \circ 2$.

2.2 Assoziativgesetz

Die Verknüpfung dreier Zahlen hängt nicht davon ab, wie wir sie gruppieren, e.g. $(2 \circ 3) \circ 4 = 2 \circ (3 \circ 4)$.

2.3 Distributivgesetz

Der Umgang mit Klammern hängt vom Zahlenraum und der Art der Verknüpfung ab. Für Addition sowie Multiplikation für die Zahlenmenge \mathbb{R} seien beide Verknüpfungen distributiv, sodass gilt: $2 \odot (3 \oplus 4) = 2 \odot 3 \oplus 2 \odot 4$.

2.4 Die binomischen Formeln

$$1 \quad (a + b)^2 = a^2 + 2ab + b^2$$

$$2 \quad (a - b)^2 = a^2 - 2ab + b^2$$

$$3 \quad (a + b) \cdot (a - b) = a^2 - b^2$$

2.5 Potenzen

Eine Potenz ist eine abgekürzte Schreibweise der Multiplikation.

- $a^0 = 1$
- $a^1 = a$
- $a^{-1} = \frac{1}{a}$
- $a^{-n} = \frac{1}{a^n}$
- $a^n = \frac{1}{a^{-n}}$
- $a^p * a^q = a^{p+q}$
- $a^p : a^q = a^{p-q}$
- $a^q * b^q = (a * b)^q$
- $a^q : b^q = (a : b)^q$
- $(a^p)^q = a^{p*q}$
- $\frac{a^m}{a^n} = a^{m-n}$
- $\frac{a^n}{b^n} = \left(\frac{a}{b}\right)^n$
- $\left(\frac{a}{b}\right)^{-n} = \left(\frac{b}{a}\right)^n$

2.6 Wurzeln

Die Wurzel einer Zahl ergibt mit sich selbst multipliziert die Wurzel. Im Normalfall handelt es sich um eine quadratische Multiplikation, aber auch höhere Wurzeln - e.g. die Kubische - ist denkbar. Als Potenz geschrieben ist die quadratische Wurzel $\sqrt{x} = x^{\frac{1}{2}}$ und $\sqrt[n]{x} = x^{\frac{1}{n}}$. Z.B.: $\sqrt[3]{125} = 125^{\frac{1}{3}}$. Haben wir eine Potenz mit einer Lösung, dann gilt $x^n = a \Leftrightarrow x = \sqrt[n]{a}$.

Beispiel: $3^4 = 81 \equiv \sqrt[4]{81} = 3$

Randnotiz: Damit ist die n-te Wurzel die inverse Funktion (7.1.3) zu der Funktion: x^n , die abhängig von x ist (quadratische, kubische, etc.).

2.7 Logarithmus

Frage: Mit welcher Zahl muss ich a potenzieren um y zu bekommen? Geschrieben als $\log_a(x) = y$. Beachte, dass das Logarithmieren von Null und negativen Zahlen **nicht definiert** ist!

Außerdem ist der Logarithmus die Umkehrfunktion zur Exponentialfunktion:

$$f(x) = a^x = y, f^{-1}(y) = \log_a(y) = x$$

Also liefert der log. den Exponenten der Exponentialfunktion zur Basis a. Haben wir die Exponentialfunktion $f(x) = e^x$ mit e als Basis, so ist der Logarithmus Naturalis $(\ln) = f^{-1}$. Der Logarithmus hat ein paar eigene Rechenregeln:

- $\log_a(1) = 0$
- $\log_a(a) = 1$
- $\log_a(p * q) = \log_a(p) + \log_a(q)$
- $\log_a(\frac{p}{q}) = \log_a(p) - \log_a(q)$
- $\log_a(p^q) = q * \log_a(p)$
- $\log_a(\sqrt[n]{p}) = \frac{\log_a(p)}{n}$
- $\log_a(p) = \frac{\log_b(p)}{\log_b(a)}$

Logarithmische Skalierungen ist hilfreich, wenn die Daten stark variieren oder wenn die Verhältnisunterschiede zwischen Werten von Bedeutung sind, denn *logarithmische Skalierung deckt einen größeren Bereich ab*. Beispiel:

$$\log_{10}(10) = 1, \log_{10}(100) = 2, \log_{10}(1000) = 3.$$

3 Geometrie

3.1 Trigonometrie

Ein Einheitskreis ist definiert als ein Kreis mit radius 1. Das Interessante hierbei ist, dass sich die x- und y-Koordinate mithilfe von sinus und cosinus bestimmen lassen, abhängig vom Winkel. Man könnte auch sagen:

$$\cos(\theta) = \frac{x}{1}$$

und

$$\sin(\theta) = \frac{y}{1}$$

Ist der *radius* $\neq 1$ spielt der Wert r eine Rolle da die Division mit einem anderen Wert x und y verschiebt. Schließlich gilt auch für den radius (oder die Hypothenuse):

$$\tan(\theta) = \frac{y}{x}.$$

Die gewöhnliche Definition von sinus, cosinus und tangens, die sich auf Dreiecke bezieht, ist sehr ähnlich und hat dementsprechend Gemeinsamkeiten, die sich darin finden lassen, dass wir diese Dreiecke einfach in den Einheitskreis einsetzen (siehe Abbildung 3.1).

$$\begin{aligned}\sin(\theta) &= \frac{y}{z} = \frac{\text{Gegenkathete}}{\text{Hypotenuse}} \\ \cos(\theta) &= \frac{x}{z} = \frac{\text{Ankathete}}{\text{Hypotenuse}} \\ \tan(\theta) &= \frac{y}{x} = \frac{\text{Gegenkathete}}{\text{Ankathete}}\end{aligned}$$

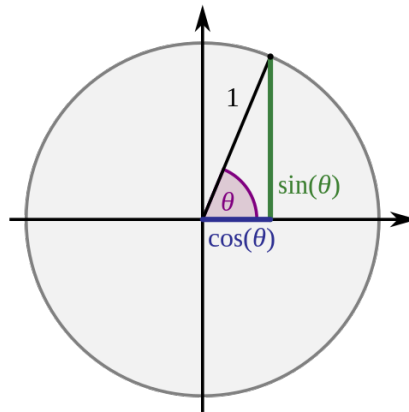


Abbildung 1: Einheitskreis von der Wikipedia Seite. Vom 13. Oktober, 2023.

Fassen wir zusammen: Die expliziten Formeln zeigen, dass **Sinus**, **Cosinus** und **Tangens** einen Winkel nutzen um das Verhältnis zweier zueinander liegender Seite zu bestimmen. Da für den Einheitskreis der Divident = 1 ist, bilden sinus und cosinus so den x- und y-Wert ab, gegeben die X- und Y-Achse ist so orientiert. Dieses Bild reflektiert sehr schön, warum folgendes gilt:

$$\begin{aligned}
\sin(0) &= 0 \\
\cos(0) &= 1 \\
\sin(90) &= 1 \\
\cos(90) &= 0 \\
\sin(-\theta) &= -\sin(\theta) \\
\cos(-\theta) &= \cos(\theta)
\end{aligned}$$

Das ist auch der Grund, warum diese Funktionen (cos und sin) periodisch sind. Für Winkel > 360 kann man genauso gut Modulo rechnen und kommt wieder bei 0, bzw. 1 an.

3.2 Polarkoodinaten

In Polarkoordinaten wird ein Punkt durch Angabe des Abstands r zum Koordinatenursprung und durch einen Winkel θ (auch Azimut genannt) bezüglich einer vorgegebenen Achse (z.B. der X-Achse) beschrieben. Das Zahlenpaar (r, θ) wird als Polarkoordinaten des Punktes bezeichnet. Umrechnen lassen sich die Koordinaten in das andere System wie folgt:

Von Polarkoodinaten in kartesische Koordinaten:

$$x = r \cdot \cos(\theta) \text{ and } y = r \cdot \sin(\theta)$$

Von kartesischen Koordinaten in Polarkoordinaten:

Hier müssen wir zwei Dinge bestimmen. Der radius ist mittels Pythagoras gegeben, i.e. $r = \sqrt{x^2 + y^2}$. Ist der radius $= 0$, sind alle möglichen θ möglich, daher gilt als (praktische) Lösung $\theta = 0$.

Außerdem sollte der Wertebereich des Winkels mit einem haloffenen Intervall gegeben sein, damit die Transformation eindeutig ist, i.e. $[0, 2\pi)$. Der

Winkel θ muss nach unterschiedlichen Fällen behandelt werden:

$$x > 0, y \geq 0 : \theta = \arctan\left(\frac{y}{x}\right)$$

$$x > 0, y < 0 : \theta = \arctan\left(\frac{y}{x}\right) + 2\pi$$

$$x < 0 : \theta = \arctan\left(\frac{y}{x}\right) + \pi$$

$$x = 0, y > 0 : \theta = 90 \text{ deg}$$

$$x = 0, y < 0 : \theta = 270$$

An dieser Stelle sei eine Besonderheit genannt. Wir wissen, dass für die Umrechnung $f_{P \rightarrow CK} = r * \sin(\theta) = y, r * \cos(\theta) = x$. Da sich r aus der Formel für den Tangens herauskürzen lässt, erhalten wir $\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}$. Anders formuliert bedeutet das folgendes: Über den Winkel θ lässt sich mit $\tan(\theta)$ das Verhältnis $y : x$ berechnen. Daher können wir den Winkel θ mit der inversen zu berechnen. $f^{-1} \equiv \tan^{-1}\left(\frac{y}{x}\right)$ übersetzt sich zu $\arctan\left(\frac{y}{x}\right)$, woraus folgt: $\arctan(\text{Gegenkathete}/\text{Ankathete}) = \theta$.²

Werden Kreiskoordinaten mit einer dritten Koordinate ergänzt, ergeben sich räumliche Polarkoordinaten, wie z. B. Kugelkoordinaten oder Zylinderkoordinaten.

3.2.1 Komplexe Zahlen mittels Polarkoordinaten

$$z \in \mathbb{C} = a + b \cdot i \equiv z = |z| \cdot e^{i\theta} = |z|(\cos(\theta) + i \cdot \sin(\theta)) = |z| \cdot \cos(\theta) + |z| \cdot i \cdot \sin(\theta)$$

3.3 Bogenmaß und Radian

Meistens werden Winkel in Grad angegeben. Aber ein Winkel von bsp. 45 Grad kann auch in Bogenmaß ($\frac{1}{4}\pi \approx 0,79$) angegeben werden. Es gilt: $360 \text{ deg} = 2\pi$, $1\pi = 180 \text{ deg}$ und $1,5\pi = 270 \text{ deg}$. Zu jedem Mittelpunktswinkel am Einheitskreis gehört ein Kreisbogen auf dem Einheitskreis. Die Länge des Kreisbogens ist ein Maß für die Größe des Winkels. Dieses wird als Bogenmaß bezeichnet und trägt die Einheit *Radian*, abgekürzt *rad*.

$$rad = \frac{\text{deg}}{180 \text{ deg}} * \pi \text{ und } \text{deg} = \frac{rad}{\pi} * 180 \text{ deg}$$

² θ is in radian.

3.4 Pythagoras

$$a^2 + b^2 = c^2$$

4 Statistik

4.1 Durchschnitt

Gibt es eine endliche Datenmenge D , die aus Stichproben der Form $x \sim P_X$ besteht, ist der Durchschnitt der arithmetische Mittelwert:

$$\bar{x} := \frac{1}{|D|} \sum_{x \in D} x$$

Der Durchschnitt ist also ein Schätzwert für den unbekannten Erwartungswert einer Verteilung von dem die Stichproben kommen.

4.2 Modus

Das am häufigsten vorkommende Element einer Datenmenge:

$$Modus(D) = 1, 2, 3, 6, 7, 8, 3, 2, 2 = 2$$

4.3 Median

Jenes Element einer Datenmenge, zu dem jeweils 50% größer und 50% kleiner sind:

$$Median(D) = 1, 2, 3, 4, 5, 6, 7 = 4$$

4.4 Erwartungswert

Der Erwartungswert ist der förmliche Begriff für den Durchschnitt einer Variablen, mit dem Unterschied, dass der Erwartungswert prinzipiell ins Unendliche geht. Ausgedrückt wird das mit:

$$E[X] := \mu_X := \int_{x \in \Omega} x dP_X(x)$$

4.5 Varianz

Die Varianz misst die erwartete Abweichung einer Variablen vom Erwartungswert. Mit der euklidischen Distanz als Metrik (hoch 2) ergibt sich folgende Formel:

$$Var(X) = E[(X - E[X])^2]$$

4.6 Standardabweichung

Die Standardabweichung ist die Wurzel der Varianz. Sie ist stets angegeben in der selber Größe wie das zu messende Merkmal.

$$\sigma = \sqrt{\text{Var}(X)}$$

4.7 Korrelation

4.8 Pearson's chi-square test

5 Wahrscheinlichkeit

5.1 Abhängige und unabhängige Variable

5.2 Kolmogorow' Axiome

Kolmogorow's drei Axiome - so heißen die Grundsätze einer Theorie - sind die bekannteste Beschreibung der grundlegenden Eigenschaften der Wahrscheinlichkeitsrechnung. Seien $\Omega = \omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6$ die Ergebnismenge eines Zufallsexperiments, A und B Teilmengen von Ω und P eine Funktion, die jedem A eine reelle Zahl zwischen 0 und 1 zuordnet. P(A) wird Wahrscheinlichkeit genannt, falls folgende drei Bedingungen (Axiome) erfüllt werden:

1. $P(A) \geq 0$ Diese Bedingung besagt, dass jede Wahrscheinlichkeit für das Eintreffen einer Teilmenge von Ω (Ereignis) nicht negativ ist. Man nennt diese Eigenschaft daher auch: Nichtnegativität
2. $P(\Omega) = 1$ Das zweite Axiom bringt eine weitere Eingrenzung des Wertebereichs von der Funktion P. Mit Axiom 1 und 2 darf P(A) mit beliebigem A minimal Wert 0 und maximal Wert 1 annehmen.
3. $P(A \cup B) = P(A) + P(B)$ Dies bedeutet also, dass für kein Ergebnis beide Ereignisse erfüllt werden. A und B nennt man in diesem Fall auch disjunkt.

5.3 Wahrscheinlichkeitskonzepte

Was ist gemeinsame, Grenz- und bedingte Wahrscheinlichkeit? Die Wahrscheinlichkeit gemeinsamer Ereignisse bedeutet, dass für die zwei Variablen ein Wert gegeben ist, also dass beispielsweise $P(X = \text{male}, Y = \text{rugby}) = \frac{100}{500}$. Man kommt auch zu diesem Ergebnis, indem man die bedingte Wahrscheinlichkeit mit der Grenzwahrscheinlichkeit multipliziert, also $P(X, Y) = P(x) \cdot P(y|x) = P(y) \cdot P(x|y) = \frac{125}{500} \cdot \frac{100}{125} = \frac{270}{500} \cdot \frac{100}{270}$. Die Grenzwahrscheinlichkeit ist die Wahrscheinlichkeit einer Kategorie anzugehören, also dass $X=x$ und $Y=y$, oder andersherum. Ein Beispiel: $P(X = \text{female}) = \frac{230}{500} = 0,46$ oder $P(Y = \text{other}) = \frac{180}{500} = 0,36$. Schließlich die bedingte Wahrscheinlichkeit, also der Fall dass für eine Variable schon ein Wert gesetzt ist der den Wert der Zielvariable einschränkt bzw. spezifiziert, wir schreiben ganz generell: $P(X|Y) = \frac{P(X,Y)}{P(Y)}$.

5.4 Wahrscheinlichkeitsverteilungen

Die Wahrscheinlichkeitsverteilung wird in zwei Arten unterteilt, die diskrete und die stetige Zufallsvariable. Diese sind dann jeweils noch mehrmals in verschiedene Kategorien unterteilt. Da es sich bei den Wahrscheinlichkeitsverteilungen um Funktionen handelt, gibt es immer einen Funktionswert und einen x-Wert. Die **Diskrete Zufallsvariable** zeichnet sich dadurch aus, dass sie eine begrenzte, abzählbare Anzahl an möglichen Ausprägungen hat. Beispiele dafür sind der Münz- oder Würfelwurf. Beide haben nur eine begrenzte Anzahl an möglichen Ausprägungen, der Münzwurf hat zum Beispiel zwei und der Würfelwurf hat dafür 6 Ausprägungen. Die **kontinuierliche oder stetige Zufallsvariable** dagegen hat eine unbegrenzte Anzahl an möglichen Ausprägungen. Als Beispiel kann man dafür die Haarlänge nehmen. Theoretisch könnte man sagen, dass es von keinen Haaren, bis zu den weltweit längsten Haaren eine begrenzte Anzahl an Zentimetern gibt. Jedoch, wenn man die Länge in immer genaueren Einheiten angeben würdest, hätte man unendlich viele verschiedene Haarlängen auf der Welt, zumal es keine festgelegte Grenze für das Haarwachstum gibt. Die **Wahrscheinlichkeits Dichtefunktion** ist eine Wahrscheinlichkeitsfunktion, die kontinuierlich ist. Die Summe der Wahrscheinlichkeiten aller Ergebnisse ist in einem Zufallsexperiment immer gleich 1. In einer stetigen Zufallsverteilung muss die 1 auf unendlich viele Ausprägungen verteilt werden. Das führt dazu, dass die Wahrscheinlichkeit für eine einzelne Ausprägung praktisch gegen 0 geht. Ebendarum lässt sich in der Dichtefunktion nicht die Wahrscheinlichkeit einer einzelnen Ausprägung ableiten. Um aber trotzdem an ein Ergebnis zu gelangen, kannst du über mehrere Ausprägungen hinweg *integrieren* und erhältst so die Wahrscheinlichkeit für diese Menge an Ausprägungen (Verteilungsfunktion). Genannt werden hier der Erwartungswert E , die Varianz V (Standardabweichung ist \sqrt{V}), die Funktion der Wahrscheinlichkeitsverteilung, aber nicht die Verteilungsfunktion.

5.4.1 Normalverteilung

Die Normalverteilung wird auch Gaußsche Glockenkurve genannt. Die beiden Parameter (μ und σ) geben Mittelwert sowie Standardabweichung der Normalverteilung an. Der zentrale Grenzwertsatz besagt, dass unter bestimmten allgemeinen Voraussetzungen die Summe aus n unabhängigen, identisch verteilten Zufallsvariablen wiederum normalverteilt ist. Als Beispiel sei nehmen wir den Wurf von n fairen Würfeln: Wenn man nur einen Würfel wirft, so ist jede Augenzahl gleich wahrscheinlich. Wirft man hingegen n -viele Würfel, so wird die mittlere Augenzahl durch die Normalverteilung beschrieben. Daher ist die Normalverteilung die Wichtigste, da natürliche Phänomene mit ausreichend großem n sich ihr annähern. Die Formel zur Berechnung der Verteilung lautet $f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$. Es gilt: $E = \mu$ und $V = \sigma^2$, bzw. σ (Standardabweichung). Die gesamte Fläche, die von der Kurve der Normalverteilung eingeschlossen wird ist stets 1. Ist $\mu = 0$ und $\sigma = 1$ spricht man von der Standardnormalverteilung, die durch eine vereinfachte Gleichung (da $\mu = 0$ wegfällt

und $\sigma = 1$) beschrieben wird: $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Der Vorfaktor stellt sicher, dass die gesamte Fläche unter der Kurve (und damit auch das Integral von $-\infty$ bis ∞) eine Fläche von genau 1 hat. Die $\frac{1}{2}$ im Exponenten der e-Funktion gibt der Normalverteilung eine Einheitsvarianz. Jede Normalverteilung ist eine Variante der Standardnormalverteilung mit gestreckter Standardabweichung ($\frac{1}{\sigma}$) und *z-transformiertem* $\frac{x-\mu_x}{\sigma}$. Geschrieben wird die Normalverteilung für gewöhnlich so: $\mathcal{N}(\mu, \sigma^2)$.

5.4.2 Exponentialverteilung

Die Exponentialverteilung ist eine kontinuierliche Verteilung, die zur Modellierung der Dauer zufälliger Zeitintervalle genutzt wird. Der Parameter λ steht für die Zahl der erwarteten *Ereignisse* pro Zeitintervall. Als Beispiele nehmen wir hier die Länge eines Telefongesprächs oder der radioaktive Zerfall. Die Verteilung lässt keine negativen Werte zu, da negative Zeiten sinnlos sind. Sie wird in der Statistik häufig mit $\exp(\lambda)$ abgekürzt. Die Dichtefunktion ist folgendermaßen definiert:

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}.$$

Der Erwartungswert E ist definiert als $\frac{1}{\lambda}$, die Varianz V als $\frac{1}{\lambda^2}$. Der Modus (der Wert, bei dem die Wahrscheinlichkeit am größten ist) liegt bei dieser Dichtefunktion bei $x_{mod} = 0$. Möchte man die Wahrscheinlichkeit des Eintretens eines Ereignisses berechnen, so nutzt man dafür idealerweise die Verteilungsfunktion $F(x)$, die das Integral bis zu einem Wert x bildet. So entsteht eine akkumulierte Wahrscheinlichkeit $P(X \leq x)$. Oft ist die tatsächliche Verteilung keine Exponentialverteilung, jedoch ist die Exponentialverteilung einfach zu handhaben und wird zur Vereinfachung angewandt. Sie ist anwendbar, wenn ein Poisson-Prozess vorliegt, also die Poissonschen Annahmen erfüllt sind. Die Exponentialverteilung ist ein Teil der viel größeren und allgemeineren Exponentialfamilie, einer Klasse von Wahrscheinlichkeitsmaßen, die sich durch eine leichte Handhabbarkeit auszeichnen.

5.4.3 Gleichverteilung

Der französische Mathematiker Pierre Simon de Laplace (1749 bis 1827) untersuchte als einer der Ersten intensiv Zufallsexperimente, bei denen angenommen werden kann, dass jedes seiner Ergebnisse mit der gleichen Wahrscheinlichkeit eintritt. Zufallsexperimente mit Gleichverteilung heißen Laplace-Experimente. Die Gleichverteilung ist ein Sonderfall unter den Wahrscheinlichkeitsverteilungen, denn sie existiert sowohl als *stetige* als auch als *diskrete* Verteilung. Hier seien einmal kurz die Formeln für deren Berechnung genannt. Zuerst für den Fall einer diskreten Verteilung: $f(x) = \frac{1}{n}$, $E(x) = \frac{n+1}{2}$ und $V(x) = \frac{1}{n} \sum_{i=0}^n (x_i - \mu)^2$.

Und für eine stetige Verteilung: $f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$. a und b seien

die Grenzen eines Intervalls, die x beinhalten. Da für alle x die gleiche Wahrscheinlichkeit gilt, ist diese von den Grenzen des Intervalls abhängig. $E(x) = \frac{a+b}{2}$ und $V(x) = \frac{1}{12} \cdot (b-a)^2$.

5.4.4 Poissonverteilung

Die Poisson-Verteilung ist eine diskrete Wahrscheinlichkeitsverteilung, welche die Verteilung von Zählgrößen beschreibt. Oder mit anderen Worten: Wie oft tritt ein bestimmtes, zählbares Ereignis ein, wenn man es sehr oft wiederholt? Der Parameter gibt hierbei die mittlere Ereignisrate an. Die Wahrscheinlichkeit für die Zufallsvariable X der Poisson-Verteilung wird durch folgende Formel berechnet: $\frac{\lambda^x}{x!} \cdot e^{-\lambda}$, $x \in N_0$. Es besteht ein Zusammenhang zwischen Exponentialverteilung und Poissonverteilung. Beide betrachten denselben Sachverhalt aus verschiedenen Perspektiven. Die Exponentialverteilung gibt an, wie die Wahrscheinlichkeit der Dauer verschiedener Vorgänge verteilt ist. Die Poissonverteilung zählt, wie oft die gezählten Ereignisse in einem festgelegten Intervall auftreten. Ausgehend von der Exponentialverteilung soll ermittelt werden, wie die Wahrscheinlichkeit ist, dass genau n Ereignisse in einem Zeitintervall von t auftreten. Wie sich zeigen wird, ist das Ergebnis die Poissonverteilung. Da der Binomialkoeffizient bei größeren Werten nur unter erhöhtem Rechenaufwand zu berechnen ist, kann man die Poisson-Verteilung benutzen, um die Binomialverteilung anzunähern. Man benutzt die Poisson-Verteilung im allgemeinen zu Annäherung der Binomialverteilung, wenn n groß ist und p klein. Als Erwartungswert $E = \mu$ der Poisson-Verteilung verwenden wir $\mu = \lambda = n \cdot p$, der mit der Varianz identisch ist. Allgemein approximiert die Poisson-Verteilung die Binomialverteilung sehr gut für Werte von $n \geq 100$ und $\lambda \leq 10$. Neben den Geschwindigkeitsvorteilen bei der Berechnung, hat die Poisson-Verteilung noch den Vorteil, dass sie unendlich abzählbar ist, sich also ins positiv Unendliche infortsetzt.

5.4.5 t-Verteilung oder Student-t-Verteilung

Die Normalverteilung wird bei vielen statistischen Verfahren eingesetzt. Allerdings unterschätzt die Normalverteilung bei kleinen Stichprobenumfängen die Standardabweichung. Dieser Effekt kann aber ausgeglichen werden, indem man bei manchen statischen Verfahren statt der Normalverteilung die t-Verteilung einsetzt. Die t-Verteilung ist eine der Normalverteilung verwandte Verteilung. Die t-Verteilung erhält man, wenn man den Mittelwert einer normalverteilten Population in Situationen schätzt, in denen der Stichprobenumfang klein³ ist und die Standardabweichung der Population unbekannt ist. Diese Verteilung zeichnet sich dadurch aus, dass Sie breitere Enden als die Normalverteilung hat. Für steigende Stichprobenumfänge nähern sich die beiden Verteilungen an und sind schließlich identisch. Die t-Verteilung ist folgendermaßen definiert:

³Eine gute Faustregel lautet, dass Sie bei einer Stichprobengröße von mindestens 30 die z-Verteilung anstelle einer t-Verteilung nutzen können.

$T = \frac{Z}{\sqrt{\frac{1}{n} * \chi^2}}$, $E = 0$ und $V = \frac{n}{n-2}$, wobei Z normalverteilt ist und χ^2 unabhängig und Chi^2 verteilt ist.

5.4.6 Chi^2 Verteilung

Die Chi^2 Verteilung ist eine kontinuierliche Verteilung, die häufig zur Testung statistischer Unabhängigkeit oder zur Gültigkeit einer Hypothese (goodness of fit) genommen wird, genannt sei hier der *Pearson's chi-square test*, 4.8. Nur wenige weltliche Dinge sind mit der Chi^2 Verteilung gut beschrieben. Es gibt einen Parameter, der die Freiheitsgrade n festlegt⁴. Diese Freiheitsgraden bestimmen die Verteilung insofern, als dass $\chi_n^2 = Z_1^2 + \dots + Z_n^2$ gilt, also dass n unabhängige, quadrierte und standardnormalverteilte Zufallsvariablen ihr ungefähr äquivalent sind, d. h. $Z_k \sim \mathcal{N}(0, 1) \forall k = 1, \dots, n$ und $\chi^2 \sim \chi_n^2$. Außerdem gilt: $E_{Chi^2} = n$ und $V_{Chi^2} = 2 \cdot k$.

5.4.7 Binominalverteilung

Prozesse bei denen nur 2 mögliche Ausgänge denkbar sind (e.g. ein Münzwurf) lassen sich mit der **Binominalverteilung** beschreiben. Voraussetzung dafür ist, dass das Experiment aus gleichen und von einander unabhängigen Versuchen besteht. Die Parameter n und k deuten schon darauf hin, dass es sich um eine diskrete Wahrscheinlichkeitsverteilung handelt, die Fragen nach k Erfolgen bei n Ereignissen beantwortet. Es gilt:

Variabel	Formel
$P(X = k)$	$\binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$
E	$n \cdot p$
V	$n \cdot p \cdot q$
σ	$\sqrt{n \cdot p \cdot q}$
$\binom{n}{k}$	$\frac{n!}{k!(n-k)!}$

(1)

Der Binominalkoeffizient beschreibt die Anzahl der Möglichkeiten, wie k Objekte in einer Gruppe aus n ohne Wiederholung angeordnet werden können. Die Binomialverteilung ist linksschief, wenn $p \geq 0,5$ ⁵, rechtsschief wenn $p \leq 0,5$ und bei $p = 0,5$ symmetrisch. Wenn n hinreichend groß ist, kann die Normalverteilung als Annäherung zur Binomialverteilung verwendet werden, da die Schiefe mit zunehmenden n kleiner wird.

5.4.8 Bernoulliverteilung

Die Bernoulli Verteilung ist eine diskrete Verteilung, deren Zufallsvariable X nur zwei Werte annimmt: 0 = Misserfolg / Niete bzw. 1 = Erfolg / Treffer. Sie

⁴Würde man zufällige unabhängige Stichproben von n normalverteilten Größen nehmen, summierten sich diese Stichproben zu einer einer $Chi^2 - V$ mit n Freiheitsgraden.

⁵Greather than aber nicht gleich! Symbol fehlt.

entsteht, wenn man ein Bernoulli Experiment (welches nur 2 mögliche Ausgänge hat) genau 1 Mal ausführt. Die Bernoulli Verteilung ist daher ein Spezialfall der Binomialverteilung für $n=1$.

$$\text{Es gilt: } E_{\text{bernoulli}} = p, V_{\text{bernoulli}} = p \cdot (1-p) \text{ und } f_{\text{bernoulli}}(x) = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \\ 0 & \text{sonst.} \end{cases}$$

5.5 Bayes

6 Algebra

Ausgehend von dem Wissen was ein Vektor, ein Vektorraum und Basisvektoren sind, definieren wir eine lineare Transformation als eine Abbildung, welche den Nullvektor unverändert und die Linien aller Dimensionen im Vektorraum in den gleichen Abständen zueinander lässt¹. Für die Basisvektoren bedeutet das, dass ein Vektor v , für den gilt **G1.1**: $v = 3 \cdot \hat{i} + 4 \cdot \hat{j}$ nach einer willkürlichen Transformation - wir können hier sagen nach Anwendung der Matrix A auf v - nach wie vor durch den Ausdruck **G1.1** gegeben ist. Sei A eine quadratische Matrix in \mathbb{R}^n , x und b Vektoren im selben Raum, dann ist A eine lineare Transformation auf x , geschrieben **G1.2**: $A * x = b$.

Im folgenden sind die 8 Axiome der Linearen Algebra:

1. $u + (v + w) = (u + v) + w$
2. $v + w = w + v$
3. Es gibt den Nullvektor, sodass $v + 0 = v$
4. Für jeden Vektor v gibt es ein inverses, sodass gilt $v + (-v) = 0$
5. $a * (b * v) = (a * b) * v$
6. $1 * v = v$
7. $a * (v + w) = a * v + a * w$
8. $(a + b) * v = a * v + b * v$

6.1 Determinante

Zur Veranschaulichung empfiehlt es sich die \det als Faktor vorzustellen, der Flächeninhalt/Volumen des Raumes, der von den Basisvektoren aufgespannt wird, skaliert. Ein Wert von Null bedeutet, dass der Raum in seinen Dimensionen reduziert wird und damit ist die Matrix weder invertierbar noch hat sie linear unabhängige Spalten. Gilt $\det(A) \neq 0$, dann ist A invertierbar. Für **G1.2** bedeutet das, dass A^{-1} eine eindeutige Lösung besitzt. Ist $\det(A) \neq 0$, so besitzt A entweder keine oder unendliche viele Lösungen.

¹Der Begriff der Linearität liegt hier vor.

7 Analysis

7.1 Funktionen/Abbildungen

7.1.1 Definitionsbereich und Wertebereich

Die Definitionsmenge ist die Menge aller x-Werte, die in die Funktion eingesetzt werden dürfen. Die Wertemenge ist die Menge aller y-Werte, die die Funktion unter Beachtung ihrer Definitionsmenge annehmen kann. Wird jedes Element des Wertebereichs 'getroffen', so heißt die Funktion **surjektiv**. Bilden keine zwei Elemente des Definitionsbereichs auf dasselbe im Wertebereich ab, so heißt die Funktion **injektiv**. Ist eine Abbildung sowohl surjektiv als auch injektiv, so heißt sie **bijektiv**.

7.1.2 Spezielle Funktionen

Eine **lineare Funktion** ist einer Funktion der allgemeinen Form $f(x) = m \cdot x + b$.

Quadratische Funktionen haben die allgemeine Form $f(x) = a \cdot x^2 + b \cdot x + c$. Dabei ist $a \neq 0$ und b und c Konstanten. Da x^2 der höchste Term ist (die höchste Ordnung) ist er der entscheidende Faktor für das Aussehen der Funktion⁶.

Eine Funktion f mit der Funktionsgleichung $f(x) = x^n | n \in \mathbb{Z} / \{0\}$ heißt **Potenzfunktion**. Wichtig ist hier zwischen den Exponenten zu unterscheiden, da hier grade/ungrade und positive/negative Exponenten die Potenzfunktion verschieden aussehen lassen. So nennt man Potenzfunktionen mit positivem Exponentem größer als 1 **Parabel**, und jene mit negativem Exponentem **Hyperbel**.

Die Sigmoid Funktion: Es gibt mehrere Sigmoidsche Funktionen. Zum Beispiel die **logistische Sigmoid**, die **Arctangent** und die **hyperbolische tangent Funktion**. Charakteristisch ist eine s-förmige Kurve, die du auch als logistische Kurve kennst. Sigmoidsche Funktionen mappen alle x-Werte zwischen 0 und 1 ($S_{log} \rightarrow$ Wahrscheinlichkeit), zwischen -1 und 1 ($S_{arctangent}$) oder noch größere Intervalle. Damit kann sie großartig als Aktivierungsfunktion in einem NN genutzt werden, um die letzte Schicht (output) zu erstellen. Außerdem ist bemerkenswert, dass die logistische Sigmoid für plus und minus unendlich gegen ihre spezifischen Limits convergiert. Daher hat sie stets Gradienten ungleich 0. Die **logistische Sigmoid** wird häufig nur **Sigmoid** genannt und wird so geschrieben:

$$S(x) = \frac{1}{1+\exp(-x)} \equiv \frac{\exp(x)}{\exp(x)+1}$$

Die **hyperbolische tangent Funktion**, die alle Werte in den Wertebereich $[-1,1]$ mapped, hat folgende Form:

⁶<https://www.schlauerlernen.de/quadratische-funktion/>

$$S_{tan}(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$$

Zuletzt genannt sei die **Arctangent Funktion**:

$$S_{arc}(x) = \arctan(x)$$

Bekannt ist sie aus der Trigonometrie(3.1), da sie die inverse Funktion (7.1.3) zur tangent Funktion ist. Ihr Wertebereich ist $[-2\pi, 2\pi]$. Die beiden letzteren Funktionen nähern sich viel schneller ihrem Grenzwert als für die klassische Sigmoid Funktion. Die **rectified linear Unit: ReLU** ist ebenfalls sehr bekannt aus KI. Sie wird geschrieben als:

$$ReLU(x) = \max(0, x)$$

Das coole an ihr, dass ihre Ableitung stets 1 ist für $x > 0$ und dass sie nicht vom *vanishing gradient* betroffen ist, der für große NN eine schnelle Konvergenz gegen 0 in der Backpropagation hat. In diesem Aspekt ist ReLU der Sigmoid klar überlegen, vorausgesetzt man kompensiert für negative x (deren Gradient = 0 ist), indem ein kleiner Term dazugerechnet wird.

7.1.3 Inversibilität

Eine Funktion heißt umkehrbar eindeutige (**eineindeutige**) Funktion, wenn nicht nur jedem Argument eindeutig ein Funktionswert zugeordnet ist, sondern auch umgekehrt zu jedem Funktionswert genau ein Argument gehört. Um die Umkehrfunktion einer Abbildung zu erhalten, löst man die Funktionsgleichung nach x um und vertauscht dann die Variablen. Zum Beispiel: $f(x) = x^2 + 1 = y \rightarrow f^{-1} = x = \sqrt{y - 1}$, vertauschen führt zu $f^{-1} = \sqrt{x - 1} = y \mid \forall x \neq 1$. Allerdings, ist diese Funktion teilweise nicht definiert und daher nicht invertierbar. Das gilt nur solange wir die negativen Zahlen zulassen. Nehmen wir R^+ , ist $f(x)$ invertierbar.

7.1.4 Nullstellen

Nullstellen sind diejenigen Punkte einer Funktion, an denen der y-Wert den Wert 0 hat. Um sie auszurechnen setzen wir die Funktionsgleichung gleich 0 und lösen sie. Für Ableitungen ist das Verfahren dasselbe, doch hat es dann eine etwas andere Bedeutung, nämlich ist die (erste) Ableitung als Steigungsgraph der Funktion anzusehen, und diejenigen Punkte die im Ableitungsgraphen Nullstellen sind, sind in der **Stammfunktion** maxima oder minima. Die zweite Ableitung mit ihren Nullstellen gibt die **Wendepunkte** der Stammfunktion wieder.

7.2 Ableitungen

Was beschreibt eine Ableitung? Die Ableitung beschreibt das Änderungsverhalten von Funktionen. Am interessantesten sind die erste und zweite Ableitung, da man hier graphisch leicht Nullstellen mit Maxima und Minima gleichsetzen kann. Die Nullstellen in der zweiten Ableitung geben Nullstellen der ersten Ableitung und damit Wende- oder Sattelpunkte der Abbildung wieder. Ableitungen werden mit $f'(x) = \frac{df(x)}{dx} = \frac{d}{dx}f(x) = \frac{dy}{dx} = y'$ und mit ∂ für d (für partielle Ableitungen) kenntlichgemacht.

7.2.1 Spezielle Ableitungen

- $f(x) = c \Rightarrow f'(x) = 0$
- $f(x) = e^x \Rightarrow f'(x) = e^x$
- $f(x) = \ln(x) \Rightarrow f'(x) = \frac{1}{x}$
- $f(x) = \sin(x) \Rightarrow f'(x) = \cos(x)$
- $f(x) = \cos(x) \Rightarrow f'(x) = -\sin(x)$
- $f(x) = \tan(x) \Rightarrow f'(x) = \frac{1}{\cos^2(x)}$

7.2.2 Ableitungsregeln

- Potenzregel: $f(x) = x^n \Rightarrow f'(x) = n * x^{n-1}$
- Faktorregel: $f(x) = c * g(x) \Rightarrow f'(x) = c * g'(x)$
- Summenregel/Differenzregel: $f(x) = h(x) + g(x) \Rightarrow f'(x) = h'(x) + g'(x)$
- Produktregel: $f(x) = h(x) * g(x) \Rightarrow f'(x) = h'(x) * g(x) + h(x) * g'(x)$
- Kettenregel: $f(x) = h(g(x)) \Rightarrow f'(x) = h'(g(x)) * g'(x)$
- Quotientenregel: $f(x) = \frac{g(x)}{h(x)} \Rightarrow f'(x) = \frac{g'(x)*h(x)-g(x)*h'(x)}{h(x)^2}$
- Linearität: $f'(a * x) = a * f'(x)$ und $f'(x + z) = f'(x) * f'(z)$

7.2.3 Partielle Ableitungen ∂

Die Ableitung einer Funktion mit mehreren Argumenten nach einem dieser Argumente heißt partielle Ableitung. Das Argument, nach dem nicht abgeleitet wird, verhält sich wie eine Konstante.

$f_x = \frac{\partial f}{\partial x}, f_y = \frac{\partial f}{\partial y}, f_{xy} = \frac{\partial^2 f}{\partial xy}, f_x = \frac{\partial^2 f}{\partial x^2}, f_{yx} = \frac{\partial^2 f}{\partial yx}$

7.3 Polynome

7.4 Integrale

7.5 Der Grenzwert

7.6 Differentialgleichungen

Eine Differentialgleichung ist eine Gleichung, in der eine Funktion und auch Ableitungen von dieser Funktion auftauchen können. Die Lösung dieser Art von Gleichung ist eine **Funktion, keine Zahl!** Die explizite Darstellung einer DGL erhalten wir, wenn wir die DGL auf die höchste vorkommende Ableitung umstellen können. Falls das nicht möglich ist, kann die DGL nur in impliziter Darstellung geschrieben werden. Also explizit: $y \hat{=}$ Ausdruck, implizit: Ausdruck $+ y \hat{=}$ Ausdruck. Wir unterscheiden zwischen gewöhnlichen DGL, bei denen die gesuchte Funktion von einer Variable abhängt, bzw. es tauchen nur Ableitungen nach einer Variablen auf, und die partielle DGL. Hier hängt die gesuchte Funktion von mehreren Variablen ab und es tauchen partielle Ableitungen der Funktion auf. Die Ordnung einer DGL ist stets n , die höchste enthaltene Ableitung. Alle anderen Ableitungen niedrigerer Ordnung sind Funktionsargumente. Außerdem unterscheidet man zwischen linearen und nicht-linearen DGL, wobei erstere als Linearkombination explizit geschrieben werden kann, $y^n + a_{n-1}(x)y^{n-1} + \dots + a_2(x)y^2 + a_1(x)y = b(x)$. Ist das nicht der Fall, ist die DGL nicht-linear. In vielen Büchern und Skripten taucht die Typisierung autonome DGL auf. Eine DGL heißt autonom, wenn die Variable x nicht explizit in der DGL auftaucht (also lediglich versteckt als Funktionsargument in der Funktion y und deren Ableitungen).

1.

7.7 Divergenz

7.7.1 Kullback-Leibler

$$D_{KL} = \mathbf{E} = [\log(\frac{p(x)}{q(x)})] = \int_a^b \log(p(x)/q(x)) dx$$

7.7.2 Jensen-Shannon

7.8 Lösungsstrategien

Wir beschreiben hier ein paar generelle Techniken und Tricks, die beim Lösen von Gleichungen hilfreich sein können.

1. Binomische Formeln nutzen

2. Faktorisieren: $2x^3 + 4x^2 - 2x - 4 = 0 \equiv 2(x+1)(x-1)(x+2)$

7.9 Regression

7.9.1 Lineare Regression

7.9.2 Polynomiale Regression

7.9.3 Logistische Regression

8 Modellierung