



Mathematische Grundlagen

BSc Psychologie | MSc Psychologie

MSc Klinische Psychologie und Psychotherapie

WiSe 2023/24

Prof. Dr. Dirk Ostwald

(5) Differentialrechnung

Grundlagen der Differentialrechnung

Analytische Optimierung

Differentialrechnung multivariater reellwertiger Funktionen

Selbstkontrollfragen

Grundlagen der Differentialrechnung

Analytische Optimierung

Differentialrechnung multivariater reellwertiger Funktionen

Selbstkontrollfragen

Definition (Ableitung)

Es sei $I \subseteq \mathbb{R}$ ein Intervall und

$$f : I \rightarrow \mathbb{R}, x \mapsto f(x) \quad (1)$$

eine univariate reellwertige Funktion. f heißt in $a \in I$ *differenzierbar*, wenn der Grenzwert

$$f'(a) := \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} \quad (2)$$

existiert. $f'(a)$ heißt dann die *Ableitung von f an der Stelle a* . Ist f differenzierbar für alle $x \in I$, so heißt f *differenzierbar* und die Funktion

$$f' : I \rightarrow \mathbb{R}, x \mapsto f'(x) \quad (3)$$

heißt *Ableitung von f* .

Bemerkungen

- Für $h > 0$ heißt $\frac{f(a+h) - f(a)}{h}$ *Newtonscher Differenzquotient*.
- Der Differenzquotient misst die Änderung $f(a+h) - f(a)$ von f pro Strecke h .
- Für $h \rightarrow 0$ misst der Differenzquotient die Änderungsrate von f in a .
- $f'(a)$ ist eine Zahl, f' ist eine Funktion.
- Wir werden keine Grenzwertbildung zur Berechnung von Ableitungen benötigen.

Definition (Notation für Ableitungen univariater reellwertiger Funktionen)

Es sei f eine univariate reellwertige Funktion. Äquivalente Schreibweisen für die Ableitung von f und die Ableitung von f an einer Stelle x sind

- (1) die *Lagrange-Notation* f' und $f'(x)$,
- (2) die *Newton-Notation* \dot{f} und $\dot{f}(x)$,
- (3) die *Leibniz-Notation* $\frac{df}{dx}$ und $\frac{df(x)}{dx}$ und
- (4) die *Euler-Notation* Df und $Df(x)$,

respektive.

Bemerkungen

- Für univariate reellwertige Funktionen benutzen wir f' und $f'(x)$ als Bezeichner.
- In Berechnungen benutzen wir auch die "Operator-Schreibweise" $\frac{d}{dx} f(x)$.
- Wir verstehen $\frac{d}{dx} f(x)$ als den Auftrag, die Ableitung von f zu berechnen.

Definition (Höhere Ableitungen)

Es sei f eine univariate reellwertige Funktion und

$$f^{(1)} := f' \quad (4)$$

sei die Ableitung von f . Die k -te Ableitung von f ist rekursiv definiert durch

$$f^{(k)} := (f^{(k-1)})' \quad \text{für } k \geq 0, \quad (5)$$

unter der Annahme, dass $f^{(k-1)}$ differenzierbar ist. Insbesondere ist die *zweite Ableitung* von f definiert durch die Ableitung von f' , also

$$f'' := (f')'. \quad (6)$$

Bemerkungen

- Wir schreiben auch $\frac{d^2}{dx^2} f(x)$ für den Auftrag, die zweite Ableitung von f zu bestimmen.
- Die nullte Ableitung $f^{(0)}$ von f ist f selbst.
- Üblicherweise schreibt man für $k < 4$ f' , f'' , f''' statt $f^{(1)}$, $f^{(2)}$, $f^{(3)}$.
- Im Allgemeinen benötigen wir nur f' und f'' .

Theorem (Rechenregeln für Ableitungen)

Für $i = 1, \dots, n$ seien g_i reellwertige univariate differenzierbare Funktionen. Dann gelten folgende Rechenregeln für Ableitungen

(1) Summenregel

$$\text{Für } f(x) := \sum_{i=1}^n g_i(x) \text{ gilt } f'(x) = \sum_{i=1}^n g_i'(x). \quad (7)$$

(2) Produktregel

$$\text{Für } f(x) := g_1(x)g_2(x) \text{ gilt } f'(x) = g_1'(x)g_2(x) + g_1(x)g_2'(x). \quad (8)$$

(3) Quotientenregel

$$\text{Für } f(x) := \frac{g_1(x)}{g_2(x)} \text{ gilt } f'(x) = \frac{g_1'(x)g_2(x) - g_1(x)g_2'(x)}{g_2^2(x)}. \quad (9)$$

(4) Kettenregel

$$\text{Für } f(x) := g_1(g_2(x)) \text{ gilt } f'(x) = g_1'(g_2(x))g_2'(x). \quad (10)$$

Bemerkung

- Für Beweise verweisen wir auf die weiterführende Literatur.

Theorem (Ableitungen elementarer Funktionen)

Für einige elementare Funktionen der Datenanalyse ergeben sich folgende Ableitungen

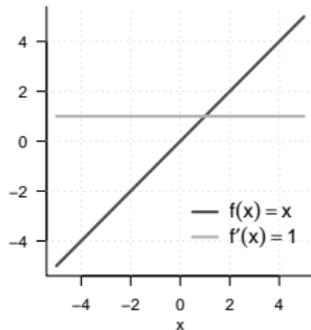
Name	Definition	Ableitung
Polynomfunktionen	$f(x) := \sum_{i=0}^n a_i x^i$	$f'(x) = \sum_{i=1}^n i a_i x^{i-1}$
Konstante Funktion	$f(x) := a$	$f'(x) = 0$
Identitätsfunktion	$f(x) := x$	$f'(x) = 1$
Linear-affine Funktion	$f(x) := ax + b$	$f'(x) = a$
Quadratfunktion	$f(x) := x^2$	$f'(x) = 2x$
Exponentialfunktion	$f(x) := \exp(x)$	$f'(x) = \exp(x)$
Logarithmusfunktion	$f(x) := \ln(x)$	$f'(x) = \frac{1}{x}$

Bemerkung

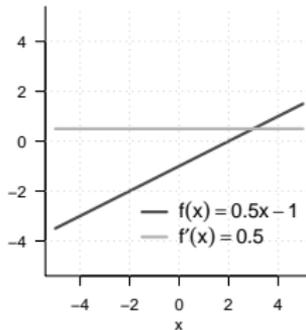
- Für Beweise verweisen wir auf die weiterführende Literatur.

Ableitungen elementarer Funktionen

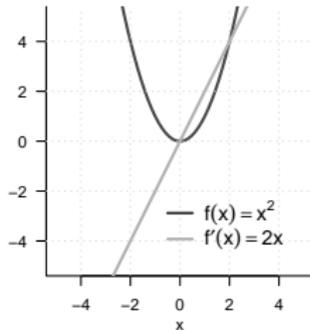
Identitätsfunktion



Linear-affine Funktion

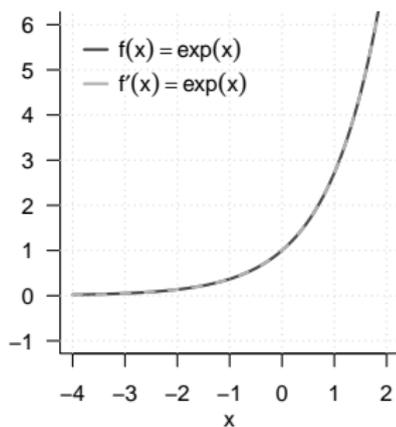


Quadratfunktion

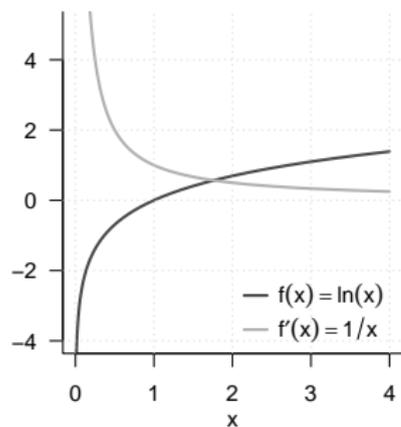


Ableitungen elementarer Funktionen

Exponentialfunktion



Logarithmusfunktion



Grundlagen der Differentialrechnung

Analytische Optimierung

Differentialrechnung multivariater reellwertiger Funktionen

Selbstkontrollfragen

Definition (Extremstellen und Extremwerte)

Es sei $U \subseteq \mathbb{R}$ und $f : U \rightarrow \mathbb{R}$ eine univariate reellwertige Funktion. Dann hat f an der Stelle $x_0 \in U$

- ein *lokales Minimum*, wenn es ein Intervall $I :=]a, b[$ gibt mit $x_0 \in]a, b[$ und

$$f(x_0) \leq f(x) \text{ für alle } x \in I \cap U, \quad (11)$$

- ein *globales Minimum*, wenn gilt, dass

$$f(x_0) \leq f(x) \text{ für alle } x \in U, \quad (12)$$

- ein *lokales Maximum*, wenn es ein Intervall $I :=]a, b[$ gibt mit $x_0 \in]a, b[$ und

$$f(x_0) \geq f(x) \text{ für alle } x \in I \cap U, \quad (13)$$

- *globales Maximum*, wenn gilt, dass

$$f(x_0) \geq f(x) \text{ für alle } x \in U. \quad (14)$$

Der Wert $x_0 \in U$ der Definitionsmenge von f heißt entsprechend *lokale* oder *globale Minimalstelle* oder *Maximalstelle*, der Funktionswert $f(x_0) \in \mathbb{R}$ heißt entsprechend *lokales* oder *globales Minimum* oder *Maximum*. Generell heißt der Wert $x_0 \in U$ *Extremstelle* und der Funktionswert $f(x_0) \in \mathbb{R}$ *Extremwert*.

Bemerkungen

- Extremstellen werden auch mit $\arg \min_{x \in I \cap U} f(x)$ oder $\arg \max_{x \in I \cap U} f(x)$ bezeichnet.
- Extremwerte werden auch mit $\min_{x \in I \cap U} f(x)$ oder $\max_{x \in I \cap U} f(x)$ bezeichnet.

Definition (Notwendige Bedingung für Extrema)

f sei eine univariate reellwertige Funktion. Dann gilt

$$x_0 \text{ ist Extremstelle von } f \Rightarrow f'(x_0) = 0. \quad (15)$$

Bemerkungen

- Wenn x_0 eine Extremstelle von f ist, dann ist die erste Ableitung von f in x_0 null.
- Sei zum Beispiel x_0 eine lokale Maximalstelle von f
- Dann gilt, dass f links von x_0 ansteigt und rechts von x_0 abfällt.
- In x_0 steigt f weder an, noch fällt f ab, also ist $f'(x_0) = 0$.

Definition (Hinreichende Bedingungen für lokale Extrema)

f sei eine zweimal differenzierbare univariate reellwertige Funktion.

- Wenn für $x_0 \in U \subseteq \mathbb{R}$

$$f'(x_0) = 0 \text{ und } f''(x_0) > 0 \quad (16)$$

gilt, dann hat f an der Stelle x_0 ein Minimum.

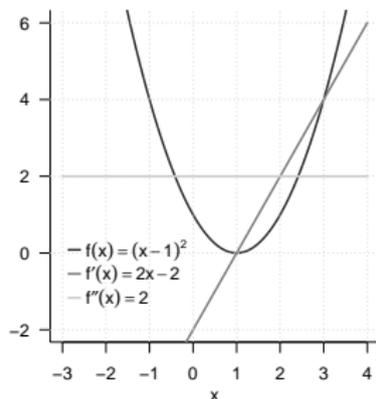
- Wenn für $x_0 \in U \subseteq \mathbb{R}$

$$f'(x_0) = 0 \text{ und } f''(x_0) < 0 \quad (17)$$

gilt, dann hat f an der Stelle x_0 ein Maximum.

Bemerkung

- Eine Intuition vermittelt nachfolgende Abbildung.



Hier ist offenbar $x_0 = 1$ eine lokale Minimalstelle von $f(x) = (x - 1)^2$. Man erkennt:

- Links von x_0 fällt f ab, rechts von x_0 steigt f an.
- In x_0 steigt f weder an, noch fällt f ab, also ist $f'(x_0) = 0$.
- Links und rechts von x_0 und in x_0 ist die Änderung f'' von f' positiv.
- Links von x_0 schwächt sich die Negativität von f' zu 0 ab.
- Rechts von x_0 verstärkt sich die Positivität von f' .

Theorem (Standardverfahren der analytischen Optimierung)

f sei eine univariate reellwertige Funktion. Lokale Extremstellen von f können mit folgendem Standardverfahren der analytischen Optimierung identifiziert werden:

- (1) Berechnen der ersten und zweiten Ableitung von f .*
- (2) Bestimmen der Nullstellen x^* von f' durch Auflösen von $f'(x^*) = 0$ nach x^* . Die Nullstellen von f' sind dann Kandidaten für Extremstellen von f .*
- (3) Evaluation von $f''(x^*)$: Wenn $f''(x^*) > 0$, dann ist x^* lokale Minimumstelle von f ; wenn $f''(x^*) < 0$, dann ist x^* lokale Maximumstelle von f ; wenn $f''(x^*) = 0$, dann ist x^* keine Extremstelle von f .*

Beispiel

Wir betrachten die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := (x - 1)^2. \quad (18)$$

aus obiger Abbildung. Die erste Ableitung von f ergibt sich mit der Kettenregel zu

$$f'(x) = \frac{d}{dx} ((x - 1)^2) = 2(x - 1) \cdot \frac{d}{dx} (x - 1) = 2x - 2. \quad (19)$$

Die zweite Ableitung von f ergibt sich zu

$$f''(x) = \frac{d}{dx} f'(x) = \frac{d}{dx} (2x - 2) = 2 > 0 \text{ für alle } x \in \mathbb{R}. \quad (20)$$

Auflösen von $f'(x^*) = 0$ nach x^* ergibt

$$f'(x^*) = 0 \Leftrightarrow 2x^* - 2 = 0 \Leftrightarrow 2x^* = 2 \Leftrightarrow x^* = 1. \quad (21)$$

$x^* = 1$ ist folglich eine Minimalstelle von f mit zugehörigen Minimalwert $f(1) = 0$.

Grundlagen der Differentialrechnung

Analytische Optimierung

Differentialrechnung multivariater reellwertiger Funktionen

Selbstkontrollfragen

Definition (Multivariate reellwertige Funktion)

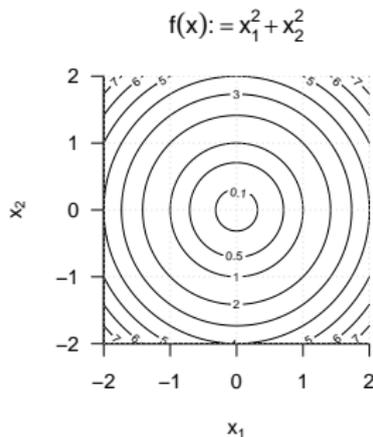
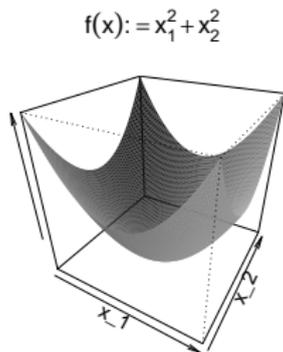
Eine Funktion der Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto f(x) = f(x_1, \dots, x_n) \quad (22)$$

heißt *multivariate reellwertige Funktion*.

Beispiel für $n := 2$

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2 \quad (23)$$



Definition (Partielle Ableitung)

Es sei $D \subseteq \mathbb{R}^n$ eine Menge und

$$f : D \rightarrow \mathbb{R}, x \mapsto f(x) \quad (24)$$

eine multivariate reellwertige Funktion. f heißt in $a \in D$ nach x_i *partiell differenzierbar*, wenn der Grenzwert

$$\frac{\partial}{\partial x_i} f(x) := \lim_{h \rightarrow 0} \frac{f(a + h e_i) - f(a)}{h} \quad (25)$$

existiert. $\frac{\partial}{\partial x_i} f(a)$ heißt dann die *partielle Ableitung von f nach x_i an der Stelle a* . Wenn f für alle $x \in D$, nach x_i partiell differenzierbar ist, dann heißt f *nach x_i partiell differenzierbar* und die Funktion

$$\frac{\partial}{\partial x_i} f : D \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_i} f(x) \quad (26)$$

heißt *partielle Ableitung von f nach x_i* .

f heißt *partiell differenzierbar in $x \in D$* , wenn f für alle $i = 1, \dots, n$ in $x \in D$ nach x_i partiell differenzierbar ist, und f heißt *partiell differenzierbar*, wenn f für alle $i = 1, \dots, n$ in allen $x \in D$ nach x_i partiell differenzierbar ist.

Bemerkungen

- $e_i \in \mathbb{R}^n$ bezeichnet den i -ten *kanonischen Einheitsvektor*.
- Es gilt also $e_{ij} = 1$ für $i = j$ und $e_{ij} = 0$ für $i \neq j$ mit $j = 1, \dots, n$.
- $\frac{f(x+he_i)-f(x)}{h}$ misst die Änderung $f(x+he_i) - f(x)$ von f pro Strecke h in Richtung e_i .
- Für $h \rightarrow 0$ misst der Differenzquotient die Änderungsrate von f in x in Richtung e_i .
- $\frac{\partial}{\partial x_i} f(x)$ ist eine Zahl, $\frac{\partial}{\partial x_i} f$ ist eine Funktion.
- Praktisch berechnet man $\frac{\partial}{\partial x_i} f$ als die (einfache) Ableitung

$$\frac{d}{dx_i} \tilde{f}_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n}(x_i) \quad (27)$$

der univariaten reellwertigen Funktion

$$\tilde{f} : \mathbb{R} \rightarrow \mathbb{R}, x_i \mapsto \tilde{f}_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n}(x_i) := f(x_1, \dots, x_i, \dots, x_n). \quad (28)$$

- Man betrachtet alle x_j mit $j \neq i$ also als Konstanten.
- Wir werden also keine Grenzwertbildung zur Berechnung von partiellen Ableitungen benötigen.

Differentialrechnung multivariater reellwertiger Funktionen

Beispiel (1)

Wir betrachten die Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2. \quad (29)$$

Weil die Definitionsmenge dieser Funktion zweidimensional ist, kann man zwei partielle Ableitungen berechnen

$$\frac{\partial}{\partial x_1} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_1} f(x) \quad \text{und} \quad \frac{\partial}{\partial x_2} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_2} f(x). \quad (30)$$

Um die erste dieser partiellen Ableitungen zu berechnen, betrachtet man die Funktion

$$f_{x_2} : \mathbb{R} \rightarrow \mathbb{R}, x_1 \mapsto f_{x_2}(x_1) := x_1^2 + x_2^2, \quad (31)$$

wobei x_2 hier die Rolle einer Konstanten einnimmt. Um explizit zu machen, dass x_2 kein Argument der Funktion ist, die Funktion aber weiterhin von x_2 abhängt haben wir die Subskriptnotation $f_{x_2}(x_1)$ verwendet. Um nun die partielle Ableitung zu berechnen, berechnen wir die (einfache) Ableitung von f_{x_2} ,

$$f'_{x_2}(x) = 2x_1. \quad (32)$$

Es ergibt sich also

$$\frac{\partial}{\partial x_1} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_1} f(x) = \frac{\partial}{\partial x_1} (x_1^2 + x_2^2) = f'_{x_2}(x) = 2x_1. \quad (33)$$

Analog gilt mit der entsprechenden Formulierung von f_{x_1} , dass

$$\frac{\partial}{\partial x_2} f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{\partial}{\partial x_2} f(x) = \frac{\partial}{\partial x_2} (x_1^2 + x_2^2) = f'_{x_1}(x) = 2x_2. \quad (34)$$

Definition (Zweite partielle Ableitungen)

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine multivariate reellwertige Funktion und $\frac{\partial}{\partial x_i} f$ sei die partielle Ableitung von f nach x_i . Dann ist die zweite partielle Ableitung von f nach x_i und x_j definiert als

$$\frac{\partial^2}{\partial x_j \partial x_i} f(x) := \frac{\partial}{\partial x_j} \left(\frac{\partial}{\partial x_i} f \right). \quad (35)$$

Bemerkungen

- Wie die zweite Ableitung ist auch die zweite partielle Ableitung rekursiv definiert.
- Zu jeder partiellen Ableitung $\frac{\partial}{\partial x_i} f$ gibt es n zweite partiellen Ableitungen $\frac{\partial^2}{\partial x_j \partial x_i} f, j = 1, \dots, n$.

Theorem (Satz von Schwarz)

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine partiell differenzierbare multivariate reellwertige Funktion. Dann gilt

$$\frac{\partial^2}{\partial x_j \partial x_i} f(x) = \frac{\partial^2}{\partial x_i \partial x_j} f(x) \text{ für alle } 1 \leq i, j \leq n. \quad (36)$$

Bemerkungen

- Für einen Beweis verweisen wir auf die weiterführende Literatur.
- Das Theorem besagt, dass die Reihenfolge des partiellen Ableitens irrelevant ist
- Das Theorem erleichtert die Berechnung von zweiten partiellen Ableitungen.
- Das Theorem hilft, Fehler bei der Berechnung zweiter partieller Ableitungen aufzudecken.

Beispiel (1) (fortgeführt)

Wir wollen die partiellen Ableitungen zweiter Ordnung der Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2. \quad (37)$$

berechnen. Mit den Ergebnissen für die partiellen Ableitungen erster Ordnung dieser Funktion ergibt sich

$$\begin{aligned} \frac{\partial^2}{\partial x_1 x_1} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_1} (2x_1) = 2 \\ \frac{\partial^2}{\partial x_1 x_2} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_1} (2x_2) = 0 \\ \frac{\partial^2}{\partial x_2 x_1} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_2} (2x_1) = 0 \\ \frac{\partial^2}{\partial x_2 x_2} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_2} (2x_2) = 2 \end{aligned} \quad (38)$$

Offenbar gilt

$$\frac{\partial^2}{\partial x_1 x_2} f(x) = \frac{\partial^2}{\partial x_2 x_1} f(x). \quad (39)$$

Beispiel (2)

Wir wollen die partiellen Ableitungen erster und zweiter Ordnung der Funktion

$$f : \mathbb{R}_{>0}^3 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_1x_2 + x_2\sqrt{x_3}. \quad (40)$$

berechnen.

Mit den Rechenregeln für Ableitungen ergibt sich für die partiellen Ableitungen erster Ordnung

$$\begin{aligned} \frac{\partial}{\partial x_1} f(x) &= \frac{\partial}{\partial x_1} (x_1^2 + x_1x_2 + x_2\sqrt{x_3}) = 2x_1 + x_2, \\ \frac{\partial}{\partial x_2} f(x) &= \frac{\partial}{\partial x_2} (x_1^2 + x_1x_2 + x_2\sqrt{x_3}) = x_1 + \sqrt{x_3}, \\ \frac{\partial}{\partial x_3} f(x) &= \frac{\partial}{\partial x_3} (x_1^2 + x_1x_2 + x_2\sqrt{x_3}) = \frac{x_2}{2\sqrt{x_3}}. \end{aligned} \quad (41)$$

Beispiel (2) (fortgeführt)

Für die zweiten partiellen Ableitungen hinsichtlich x_1 ergibt sich

$$\begin{aligned}\frac{\partial^2}{\partial x_1 \partial x_1} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_1} (2x_1 + x_2) = 2, \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_2} (2x_1 + x_2) = 1, \\ \frac{\partial^2}{\partial x_3 \partial x_1} f(x) &= \frac{\partial}{\partial x_3} \left(\frac{\partial}{\partial x_1} f(x) \right) = \frac{\partial}{\partial x_3} (2x_1 + x_2) = 0.\end{aligned}\tag{42}$$

Für die zweiten partiellen Ableitungen hinsichtlich x_2 ergibt sich

$$\begin{aligned}\frac{\partial^2}{\partial x_1 \partial x_2} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_1} (x_1 + \sqrt{x_3}) = 1, \\ \frac{\partial^2}{\partial x_2 \partial x_2} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_2} (x_1 + \sqrt{x_3}) = 0, \\ \frac{\partial^2}{\partial x_3 \partial x_2} f(x) &= \frac{\partial}{\partial x_3} \left(\frac{\partial}{\partial x_2} f(x) \right) = \frac{\partial}{\partial x_3} (x_1 + \sqrt{x_3}) = \frac{1}{2\sqrt{x_3}}.\end{aligned}\tag{43}$$

Beispiel (2) (fortgeführt)

Für die zweiten partiellen Ableitungen hinsichtlich x_3 ergibt sich

$$\begin{aligned}\frac{\partial^2}{\partial x_1 \partial x_3} f(x) &= \frac{\partial}{\partial x_1} \left(\frac{\partial}{\partial x_3} f(x) \right) = \frac{\partial}{\partial x_1} \left(\frac{x_2}{2} \sqrt{x_3} \right) = 0, \\ \frac{\partial^2}{\partial x_2 \partial x_3} f(x) &= \frac{\partial}{\partial x_2} \left(\frac{\partial}{\partial x_3} f(x) \right) = \frac{\partial}{\partial x_2} \left(\frac{x_2}{2\sqrt{x_3}} \right) = \frac{1}{2\sqrt{x_3}}, \\ \frac{\partial^2}{\partial x_3 \partial x_3} f(x) &= \frac{\partial}{\partial x_3} \left(\frac{\partial}{\partial x_3} f(x) \right) = \frac{\partial}{\partial x_3} \left(x_2 \frac{1}{2} x_3^{-\frac{1}{2}} \right) = -\frac{1}{4} x_2 x_3^{-\frac{3}{2}}.\end{aligned}\tag{44}$$

Weiterhin erkennt man, dass die Reihenfolge der partiellen Ableitungen irrelevant ist, denn es gilt

$$\begin{aligned}\frac{\partial^2}{\partial x_1 \partial x_2} f(x) &= \frac{\partial^2}{\partial x_2 \partial x_1} f(x) = 1, \\ \frac{\partial^2}{\partial x_1 \partial x_3} f(x) &= \frac{\partial^2}{\partial x_3 \partial x_1} f(x) = 0, \\ \frac{\partial^2}{\partial x_2 \partial x_3} f(x) &= \frac{\partial^2}{\partial x_3 \partial x_2} f(x) = \frac{1}{2\sqrt{x_3}}.\end{aligned}\tag{45}$$

Definition (Gradient)

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine multivariate reellwertige Funktion. Dann ist der *Gradient* $\nabla f(x)$ von f an der Stelle $x \in \mathbb{R}^n$ definiert als

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{pmatrix} \in \mathbb{R}^n. \quad (46)$$

Bemerkungen

- $\nabla f(x)$ fasst die partiellen Ableitungen von f an der Stelle $x \in \mathbb{R}^n$ in einem Vektor zusammen.
- Gradienten sind multivariate vektorwertige Abbildungen der Form $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n, x \mapsto \nabla f(x)$.
- $-\nabla f(x)$ zeigt die Richtung des steilsten Abstiegs von f in \mathbb{R}^n an. Diese Einsicht ist aber nicht trivial.
- Für $n = 1$ gilt $\nabla f(x) = f'(x)$.

Der Erfolg neuronaler Netze (aka *AI* oder *Deep Learning*) basiert auf effizienten Gradientenberechnungen. Eine Einführung dazu geben Ostwald & Usée (2021) *An Induction Proof of the Backpropagation Algorithm in Matrix Notation*. <https://arxiv.org/abs/2107.09384>

Beispiele

Für die in Beispiel (1) betrachtete Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ gilt

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \in \mathbb{R}^2. \quad (47)$$

Für die in Beispiel (2) betrachtete Funktion $f : \mathbb{R}_{>0}^3 \rightarrow \mathbb{R}$ gilt

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \frac{\partial}{\partial x_3} f(x) \end{pmatrix} = \begin{pmatrix} 2x_1 + x_2 \\ x_1 + \sqrt{x_3} \\ \frac{x_2}{2\sqrt{x_3}} \end{pmatrix} \in \mathbb{R}^3. \quad (48)$$

Differentialrechnung multivariater reellwertiger Funktionen

Beispiel (1) (fortgeführt)

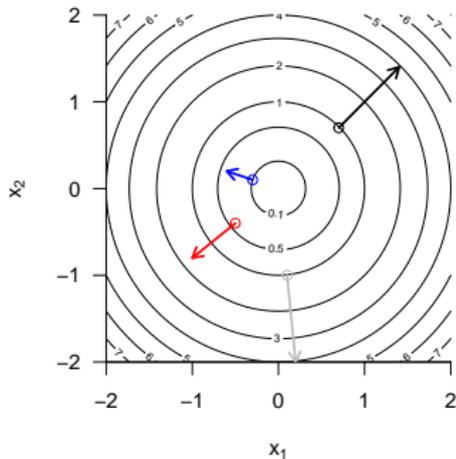
Gradienten von $f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := x_1^2 + x_2^2$ bei

$$x = \begin{pmatrix} 0.7 \\ 0.7 \end{pmatrix}$$

$$x = \begin{pmatrix} -0.3 \\ 0.1 \end{pmatrix}$$

$$x = \begin{pmatrix} -0.5 \\ -0.4 \end{pmatrix}$$

$$x = \begin{pmatrix} 0.1 \\ -1.0 \end{pmatrix}$$



Definition (Hesse-Matrix)

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine multivariate reellwertige Funktion. Dann ist die *Hesse-Matrix* $\nabla^2 f(x)$ von f an der Stelle $x \in \mathbb{R}^n$ definiert als

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \frac{\partial^2}{\partial x_n \partial x_2} f(x) & \cdots & \frac{\partial^2}{\partial x_n \partial x_n} f(x) \end{pmatrix} \in \mathbb{R}^{n \times n}. \quad (49)$$

Bemerkung

- $\nabla^2 f(x)$ fasst die partiellen Ableitungen zweiter Ordnung von f in einer Matrix zusammen.
- Hesse-Matrizen sind multivariate matrixwertige Abbildungen der Form $\nabla^2 f : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}, x \mapsto \nabla^2 f(x)$.
- Für $n = 1$ gilt $\nabla^2 f(x) = f''(x)$.
- Mit $\frac{\partial^2}{\partial x_i \partial x_j} f(x) = \frac{\partial^2}{\partial x_j \partial x_i} f(x)$ für $1 \leq i, j \leq n$ folgt, dass $(\nabla^2 f(x))^T = \nabla^2 f(x)$.

Beispiel

Für die in Beispiel (1) betrachtete Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ gilt

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2}{\partial x_1 x_1} f(x) & \frac{\partial^2}{\partial x_1 x_2} f(x) \\ \frac{\partial^2}{\partial x_2 x_1} f(x) & \frac{\partial^2}{\partial x_2 x_2} f(x) \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \in \mathbb{R}^{2 \times 2}$$

Für die in Beispiel (2) betrachtete Funktion $f : \mathbb{R}_{>0}^3 \rightarrow \mathbb{R}$ gilt

$$\begin{aligned} \nabla^2 f(x) &:= \begin{pmatrix} \frac{\partial^2}{\partial x_1 x_1} f(x) & \frac{\partial^2}{\partial x_1 x_2} f(x) & \frac{\partial^2}{\partial x_1 x_3} f(x) \\ \frac{\partial^2}{\partial x_2 x_1} f(x) & \frac{\partial^2}{\partial x_2 x_2} f(x) & \frac{\partial^2}{\partial x_2 x_3} f(x) \\ \frac{\partial^2}{\partial x_3 x_1} f(x) & \frac{\partial^2}{\partial x_3 x_2} f(x) & \frac{\partial^2}{\partial x_3 x_3} f(x) \end{pmatrix} \\ &:= \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & \frac{1}{2\sqrt{3}} \\ 0 & \frac{1}{2\sqrt{3}} & -\frac{1}{4}x_2x_3^{-3/2} \end{pmatrix}. \end{aligned}$$

Grundlagen der Differentialrechnung

Analytische Optimierung

Differentialrechnung multivariater reellwertiger Funktionen

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition des Begriffs der Ableitung $f'(a)$ einer Funktion f an einer Stelle a wieder.
2. Geben Sie die Definition des Begriffs der Ableitung f' einer Funktion f .
3. Erläutern Sie die Symbole $f'(x)$, $\dot{f}(x)$, $\frac{df(x)}{dx}$, und $\frac{d}{dx}f(x)$.
4. Geben Sie die Definition des Begriffs der zweiten Ableitung f'' einer Funktion f wieder.
5. Geben Sie die Summenregel für Ableitungen wieder.
6. Geben Sie die Produktregel für Ableitungen wieder.
7. Geben Sie die Quotientenregel für Ableitungen wieder.
8. Geben Sie die Kettenregel für Ableitungen wieder.
9. Bestimmen Sie die erste Ableitung der Funktion $f(x) := 3x^2 + \exp(-x^2) - x \ln(x)$.
10. Bestimmen Sie die erste Ableitung der Funktion $f(x) := \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2$ für $\mu \in \mathbb{R}$.
11. Geben Sie die Definition der Begriffe des globalen und lokalen Maximums/Minimums einer univariaten reellwertigen Funktion wieder.
12. Geben Sie die notwendige Bedingung für ein Extremum einer Funktion wieder.
13. Geben Sie die hinreichende Bedingung für ein lokales Extremum einer Funktion wieder.
14. Geben Sie das Standardverfahren der analytischen Optimierung wieder.
15. Bestimmen Sie einen Extremwert von $f(x) := \exp\left(-\frac{1}{2}(x - \mu)^2\right)$ für $\mu \in \mathbb{R}$.
16. Berechnen Sie die partiellen Ableitungen der Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) := \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right)$.
17. Berechnen Sie die zweiten partiellen Ableitungen obiger Funktion f .
18. Geben Sie den Satz von Schwarz wieder.
19. Geben Sie die Definition des Gradienten einer multivariaten reellwertigen Funktion wieder.
20. Geben Sie den Gradienten obiger Funktion f an und werten Sie ihn in $x = (1, 2)^T$ aus.
21. Geben Sie die Definition der Hesse-Matrix einer multivariaten reellwertigen Funktion wieder.
22. Geben Sie die Hesse-Matrix obiger Funktion f an und werten Sie sie in $x = (1, 2)^T$ aus.