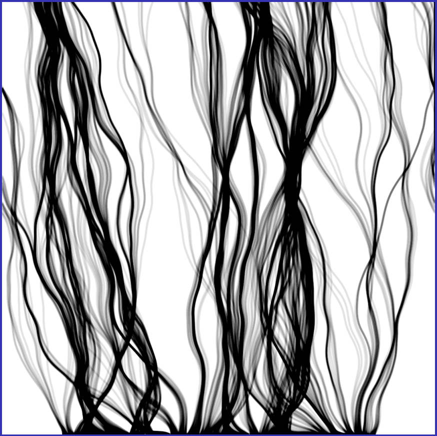
Angewandte Mathematik



Dr. rer. nat. Johannes Riesterer

Was ist Angewandte Mathematik?

Was ist Angewandte Mathematik?

- Algorithmen zum Lösen von Problemen.

Was ist Angewandte Mathematik?

- Algorithmen zum Lösen von Problemen.
- Abschätzungen, wie gut und genau die Algorithmen funktionieren.

Was ist Angewandte Mathematik?

- Algorithmen zum Lösen von Problemen.
- Abschätzungen, wie gut und genau die Algorithmen funktionieren.
- Mathematische Grundlagen, auf denen Algorithmen und Abschätzungen basieren.

Was ist Angewandte Mathematik?

- Algorithmen zum Lösen von Problemen.
- Abschätzungen, wie gut und genau die Algorithmen funktionieren.
- Mathematische Grundlagen, auf denen Algorithmen und Abschätzungen basieren.
- Softwaretechnische Aspekte in Bezug auf Implementierung der Algorithmen.

Kann jeder Mathematik lernen?

Lebenslauf einer von Emmy Noether, einer der bedeutendsten Mathematikerinnen aller Zeiten eingehen (Auszug aus Wikipedia): ... 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.

Kann jeder Mathematik lernen?

Kann jeder Mathematik lernen?

- Oft ist Mathematik ein Motivationsproblem.

Kann jeder Mathematik lernen?

- Oft ist Mathematik ein Motivationsproblem.
- Starten mit Beispielen, die motivieren sollen, wozu der Stoff gut ist.

Kann jeder Mathematik lernen?

- Oft ist Mathematik ein Motivationsproblem.
- Starten mit Beispielen, die motivieren sollen, wozu der Stoff gut ist.
- Unterfüttern Beispiele mit Softwarecode.

Kann jeder Mathematik lernen?

- Oft ist Mathematik ein Motivationsproblem.
- Starten mit Beispielen, die motivieren sollen, wozu der Stoff gut ist.
- Unterfüttern Beispiele mit Softwarecode.
- Rechnen Übungsaufgaben.

Stoff

Stoff

- Mehrdimensionale Differentialrechnung $->$ Backpropagation zum Trainieren Neuronaler Netze.

Stoff

- Mehrdimensionale Differentialrechnung $->$ Backpropagation zum Trainieren Neuronaler Netze.
- Mehrdimensionale Integralrechnung $->$ Signalanalyse mit FFT, Abtast-Theorem von Shanon, Bildverarbeitung

Stoff

- Mehrdimensionale Differentialrechnung $->$ Backpropagation zum Trainieren Neuronaler Netze.
- Mehrdimensionale Integralrechnung $->$ Signalanalyse mit FFT, Abtast-Theorem von Shanon, Bildverarbeitung
- Gewöhnliche Integralgleichungen (ODE) $->$ Vorhersage von zeitlich abhängigen, deterministischen Vorgängen.

Stoff

- Mehrdimensionale Differentialrechnung $->$ Backpropagation zum Trainieren Neuronaler Netze.
- Mehrdimensionale Integralrechnung $->$ Signalanalyse mit FFT, Abtast-Theorem von Shanon, Bildverarbeitung
- Gewöhnliche Integralgleichungen (ODE) $->$ Vorhersage von zeitlich abhängigen, deterministischen Vorgängen.
- Partielle Integralgleichungen $->$ Vorhersage von zeitlich und räumlich abhängigen, deterministischen Vorgängen.

Neuroanles Netz

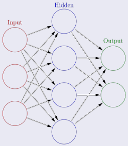


Figure: Quelle: Wikipedia

Motivation

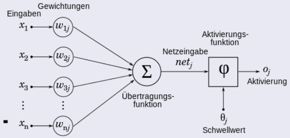


Figure: Quelle: Wikipedia

Neuronales Netz

Ein neuronales Netz lässt sich als Funktion $f(\omega, x)$ mit Gewichten ω und Input x auffassen. Diese ist aus einzelnen, einfachen Teilen zusammengesetzt.

Neuronales Netz

Ein neuronales Netz lässt sich als Funktion $f(\omega, x)$ mit Gewichten ω und Input x auffassen. Diese ist aus einzelnen, einfachen Teilen zusammengesetzt.

Neuronales Netz trainieren

Gegeben Datensatz D. Finde Gewichte Omega, so dass Lossfunktion

$$L_D : \Omega \subset \mathbb{R}^n \to \mathbb{R}$$

minimal wird. Zum Beispiel $L_D(\omega) := \sum_{(x,y)\in D}(f(\omega, x) - y)^2$.

Backpropagation

Standard verfahren ist das sogenannte Backpropagation-Verfahren. Im Wesentlichen Ableitung hochdimensionaler, zusammengesetzter Funktionen berechnen.

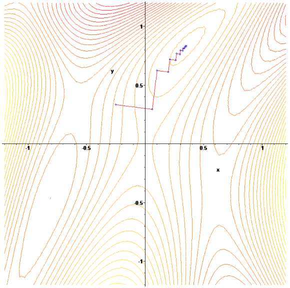


Figure: Quelle: Wikipedia

Differenzierbarkeit reeller Funktionen

Eine reelle Funktion $f : (a, b) \to \mathbb{R}$ heißt differenzierbar in $x \in (a, b)$, falls der Grenzwert $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ existiert. In diesem Fall heißt dieser Grenzwert die Ableitung (Steigung) von f in x und wird mit $f'(x)$ bezeichnet.

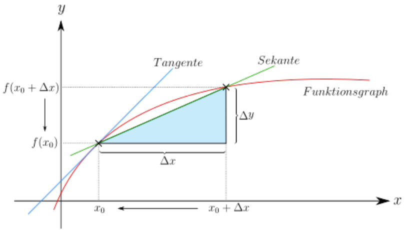


Figure: Quelle: Wikipedia: https://de.wikipedia.org/wiki/Datei:Differencial_quotient_of_a_function.svg

Mittelwertsatz einer Veränderlichen

Sei $f : [a, b] \to \mathbb{R}$ stetig und differenzierbar für alle $x \in (a, b)$.
Dann gibt es $\xi \in (a, b)$ mit $f'(\xi) = \frac{f(b)-f(a)}{b-a}$.

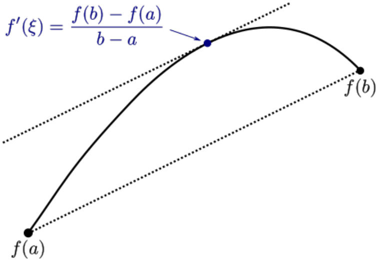


Figure: Quelle: Wikipedia:
https://commons.wikimedia.org/wiki/File:Mittelwertsatz3.svg

Taylorapproximation einer Veränderlichen

Jede reelle Funktion f, deren $p+1$-ten Ableitungen existieren und stetig sind lässt sich mit Hilfe der Taylorreihe

$$f(x) = f(a) + \frac{1}{2!}f^{'}(x-a) + \frac{1}{3!}f^{''}(a)(x-a)^2 + \cdots$$
$$+ \frac{1}{p!}f^{(p)}(a)(x-a)^{p-1} + R_{p+1}(x,a)$$

und dem Restglied $R_p(x,a) := \frac{1}{(p+1)!}f^{(p+1)}(\xi)(x-a)^p$ mit einem $\xi \in (x,a)$ darstellen.

Cauchy-Schwarzsche Ungleichung

Für zwei Vektoren $v, w \in \mathbb{R}^n$ gilt:

$$\frac{\langle v, w \rangle}{||v|| \cdot ||w||} = \cos(\varphi)$$

wobei φ der Innenwinkel zwischen v und w ist.

Äquivalenz von Normen

Die Normen $||v|| := \sqrt{\sum_{i=1}^{n} v_i^2}$ und $||v||_\infty := \max\{v_1, \cdots, v_n\}$ sind Äquivalent. Sie lassen sich mit Konstanten $k_1||v|| < ||v||_\infty < k_2||v||$ gegeneinander abschätzen.

Symmetrische Matrizen

Für eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ ist Äquivalent:

- A hat positive Eigenwerte.
- Für alle $v \in \mathbb{R}^n \neq 0$ gilt $v^T A v > 0$.
- alle Unterdeterminanten sind positiv. Speziell für $n = 2$ und $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ bedeutet dies $a > 0$ und $ad - b^2 > 0$.

Konvergenz

Eine Folge (a_n) in $\mathbb{R}^n$ heißt konvergent gegen den Grenzwert $a \in \mathbb{R}^n$, wenn gilt:

$$\forall \varepsilon > 0 \; \exists \; N \in \mathbb{N} \; \forall \; n > N : \; d(a, a_n) < \varepsilon$$

in Worten: Es gibt für jedes beliebige (noch so kleine) ε einen Index N derart, dass für alle Indizes $n > N$, alle weiteren Folgenglieder, gilt: der Abstand $d(a, a_n)$ ist kleiner als ε.

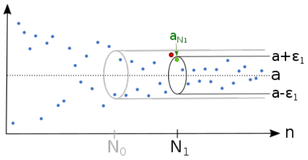


Figure: Quelle: Wikipedia:
https://commons.wikimedia.org/wiki/File:Epsilonschlauch_klein.svg

Konvergenz

Sei $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$ eine Funktion und $a \in X$. Wir nennen $L_a \in \mathbb{R}^m$ Grenzwert von f bezüglich der Annäherung von x an a, falls für jede konvergente Folge $x_n \to a$ die Folge $f(x_n)$ nach L_a konvergiert. In diesem Fall bezeichnen wir

$$\lim_{x \to a} f(x) = L_a \, .$$

Dies ist gleichbedeutend damit, dass für jedes $\epsilon > 0$ ein $\delta > 0$ existiert, so dass $d(f(x), L_a) < \epsilon$ gilt für jedes x mit $d(x, a) < \delta$.

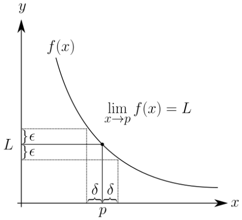


Figure: Quelle: Wikipedia:
https://de.wikipedia.org/wiki/Datei:Limes_Definition_Vektorgrafik.svg

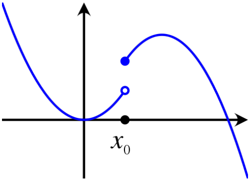


Figure: Quelle: Wikipedia:
https://commons.wikimedia.org/wiki/File:Upper_semi.svg

Stetigkeit

Eine reellwertige Funktion $f : U \to \mathbb{R}$ heißt stetig, wenn für alle $y \in U$ der Grenzwert $\lim_{x \to y} f(x) = L_y$ existiert.

Richtungsableitung

Sei $f : U \to \mathbb{R}$ eine Funktion. Für einen Vektor $h \in \mathbb{R}^n$ und einen Punkt $a \in U$ heißt der Grenzwert (falls er existiert)

$$\partial_h f(a) := \lim_{t \to 0} \frac{f(a+th) - f(a)}{t}$$

Richtungsableitung von f am Punkt a in Richtung h. Sie misst die Änderung der Funktion in Richtung h.
Speziell nennen wir für die Standard Basisvektoren e_i

$$\frac{\partial f(a)}{\partial x_n} := \partial_{e_i} f(a) := \lim_{t \to 0} \frac{f(a+te_i) - f(a)}{t}$$

die partielle Ableitung von f in a nach x_i.

Partielle Differenzierbarkeit

Eine Funktion $f : U \to \mathbb{R}$ heißt partiell differenzierbar im Punkt $a \in U$, falls alle partiellen Ableitungen

$$\frac{\partial f(a)}{\partial x_1}, \dots, \frac{\partial f(a)}{\partial x_n}$$

existieren.

Differenzierbarkeit

Eine Funktion $f : U \to \mathbb{R}$ heißt differenzierbar im Punkt $a \in U$, falls alle partiellen Ableitungen

$$\frac{\partial f(a)}{\partial x_1}, \cdots, \frac{\partial f(a)}{\partial x_n}$$

existieren und stetig sind. Mann nennt in diesem Fall die $1 \times n$-Matrix

$$df(a) := \left(\frac{\partial f(a)}{\partial x_1}, \cdots, \frac{\partial f(a)}{\partial x_n} \right)$$

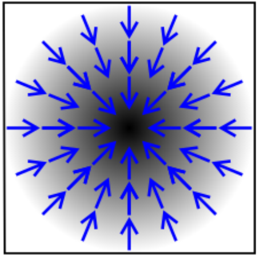
das Differential von f im Punkt a.

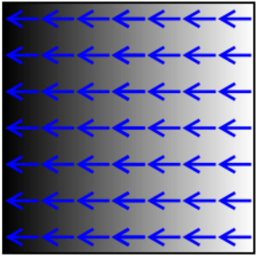
Gradient

Der Vektor

$$\nabla f(a) := \begin{pmatrix} \frac{\partial f(a)}{\partial x_1} \\ \vdots \\ \frac{\partial f(a)}{\partial x_n} \end{pmatrix}$$

wird als Gradient bezeichnet. Es ist $df(a) \cdot h = \langle \nabla f(a), h \rangle$.





Gradient

Sei $f : U \to \mathbb{R}$ differenzierbare Funktion, $a \in U$ und $v := \operatorname{argmax}_{||h||=1}\{\partial_h f(a)\}$. Dann gilt

$$||\nabla f(a)||v = \nabla f(a) \ .$$

Beweis

Mit der CSU Ungleichung folgt für beliebiges h

$$\partial_h f(a) = df(a)h = \langle \nabla f(a), h \rangle = ||\nabla f(a)|| \cdot ||h|| \cdot \cos(\varphi)$$

wobei φ den Innenwinkel zwischen $\nabla f(a)$ und h bezeichnet. Für $||h|| = 1$ wird somit $\partial_h f(a)$ maximal, wenn $\varphi = 0$ und somit $h = \nabla f(a)$ ist.