# Maschinelles Lernen 0

## Mathematische Grundlagen

Prof. Dr. Christoph Böhm

Hochschule München

3. Januar 2024

Bevor wir uns mit dem maschinellen Lernen beschäftigen können, brauchen wir eine solide mathematische Basis.

**Bitte um Mithilfe!**

Das Grundlagenmaterial ist bei Weitem nicht vollständig und soll durch Ihre Mithilfe wachsen. Sollten Sie an einer Stelle ein mathematisches Konzept oder eine Notation nicht verstehen, so melden Sie sich bitte und ich werde ggf. das Material erweitern.

# Lineare Algebra

Lineare Algebra

# Lineare Algebra
## Skalare

### Skalar

Ein Skalar ist eine einzelne Zahl.

### Beispiel

- Natürliche Zahlen $n \in \mathbb{N}$
- Rationale Zahlen $q \in \mathbb{Q}$
- Reelle Zahlen $r \in \mathbb{R}$

# Lineare Algebra

## Vektoren

### Vektor

Ein Vektor ist eine (geordnete) Liste von Zahlen.

### Beispiel

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \in \mathbb{R}^n$$

$$\mathbf{x}_i \in \mathbb{R} \qquad \forall i \in \{1, \dots, n\}$$

- Null-Vektor $\mathbf{0} \in \mathbb{R}^n$ with $\mathbf{0}_i = 0$ für alle $i \in \{1, \dots, n\}$

# Lineare Algebra

## Vektoren

### Skalarprodukt

Gegeben zwei Vektoren $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ definieren wir das Skalarprodukt als

$$\mathbf{x} \circ \mathbf{y} = \sum_{i=1}^{n} \mathbf{x}_i \cdot \mathbf{y}_i = \mathbf{x}^T \mathbf{y}$$

### Beispiel

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \circ \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 1 \cdot 3 + 2 \cdot 4 = 3 + 8 = 11$$

# Lineare Algebra

## Vektoren

### Vektornorm

Eine Vektornorm ist eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ mit

- $f(\mathbf{x}) = 0 \Rightarrow \mathbf{x} = \mathbf{0}$
- $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$ (Dreiecksungleichung)
- $f(\alpha \mathbf{x}) = |\alpha| f(\mathbf{x})$ für alle $\alpha \in \mathbb{R}$

### Beispiel

- Eine Vektornorm misst die *Größe* eines Vektors.
- $L_1$-Norm: $||\mathbf{x}||_1 = \sum_i |\mathbf{x}_i|$
- $L_2$-Norm: $||\mathbf{x}||_2 = \sqrt{\sum_i \mathbf{x}_i^2}$, auch euklidische Norm genannt

# Lineare Algebra

## Matrizen

### Matrix

Eine Matrix ist ein 2-dimensionales Feld von Zahlen.

### Beispiel

Eine reele Matrix mit $m$ Zeilen und $n$ Spalten:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \dots & \mathbf{A}_{1n} \\ & & \\ \mathbf{A}_{m1} & \dots & \mathbf{A}_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

$$\mathbf{A}_{ij} \in \mathbb{R} \qquad \forall i \in \{1, \dots, m\}, j \in \{1, \dots, n\}$$

# Lineare Algebra

## Matrizen

### Transponierte Matrix

Die Einträge der transponierten Matrix $\mathbf{A}^T$ einer Matrix $\mathbf{A}$ sind definiert als

$$(\mathbf{A}^T)_{ij} = \mathbf{A}_{ji}.$$

### Beispiel

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \mathbf{A}^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

# Lineare Algebra

## Matrizen

### Matrix-Multiplikation

Das Produkt zweier Matrizen $\mathbf{A} \in \mathbb{R}^{k \times m}$ und $\mathbf{B} \in \mathbb{R}^{m \times n}$ ist eine Matrix $\mathbf{C} = \mathbf{AB}$ mit den Einträgen

$$\mathbf{C}_{ij} = \sum_k \mathbf{A}_{ik}\mathbf{B}_{kj}.$$

### Beispiel

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} g & h \\ i & j \\ k & l \end{bmatrix} = \begin{bmatrix} ag + bi + ck & ah + bj + cl \\ dg + ei + fk & dh + ej + fl \end{bmatrix}$$

# Lineare Algebra

## Matrizen

### Identitätsmatrix

Eine Identitätsmatrix ist eine Matrix $\mathbf{I} \in \mathbb{R}^{n \times n}$ mit

$$\mathbf{I}_{ij} = \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{andernfalls} \end{cases}$$

### Beispiel

Die Matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

ist eine Identitätsmatrix.

# Lineare Algebra

## Matrizen

### Beispiel

Gegeben eine Matrix $\mathbf{A}$ und ein Vektor $\mathbf{x}$, so gilt für eine Identitätsmatrix $\mathbf{I}$

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

und

$$\mathbf{xI} = \mathbf{Ix} = \mathbf{x}.$$

# Lineare Algebra

## Matrizen

### Lineares Gleichungssystem

Ein lineares Gleichungssystem (LGS) ist definiert durch

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

mit $\mathbf{A} \in \mathbb{R}^{n \times n}$ und $\mathbf{b}, \mathbf{x} \in \mathbb{R}^n$.

### Beispiel

Ein solchen LGS repräsentiert die Menge an Gleichungen:

(1) $\mathbf{A}_{11}\mathbf{x}_1 + \mathbf{A}_{12}\mathbf{x}_2 + \cdots + \mathbf{A}_{1n}\mathbf{x}_n = \mathbf{b}_1$

(2) $\mathbf{A}_{21}\mathbf{x}_1 + \mathbf{A}_{22}\mathbf{x}_2 + \cdots + \mathbf{A}_{2n}\mathbf{x}_n = \mathbf{b}_2$

$\ldots$

(n) $\mathbf{A}_{n1}\mathbf{x}_1 + \mathbf{A}_{n2}\mathbf{x}_2 + \cdots + \mathbf{A}_{nn}\mathbf{x}_n = \mathbf{b}_n$

# Lineare Algebra

## Matrizen

### Lösungen eines LGS

Ein LGS

$$\mathbf{Ax} = \mathbf{b}$$

kann

- keine
- eine eindeutige ($\mathbf{A}$ ist invertierbar)
- viele

Lösung(en) besitzen.

# Lineare Algebra

## Matrizen

### Inverse Matrix

Die inverse Matrix $\mathbf{A}^{-1} \in \mathbb{R}^{n \times n}$ einer Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ ist definiert durch

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

### Beispiel

- Nicht jede Matrix ist invertierbar (linear abhängige Zeilen/Spalten – niedriger Rang)
- Theoretisch kann man LGS mit Hilfe der Inversen berechnen ($\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$) aber dies kann zu numerischen Problemen führen.

# Lineare Algebra

## Hyperebenen

Eine Hyperebene im $d$-dimensionalen Raum $\mathbb{R}^d$ ist ein $d-1$-dimensionaler Unterraum, welcher nicht unbedingt den Ursprung $\mathbf{0} \in \mathbb{R}^d$ beinhalten muss. Beispiele:

- $d = 1$: eine Hyperebene ist ein Skalar
- $d = 2$: eine Hyperebene ist eine Gerade
- $d = 3$: eine Hyperebene ist eine Ebene
- Hyperebenen in höheren Dimensionen sind schwierig vorzustellen

# Lineare Algebra
## Hyperebenen

Formal ist eine Hyperebene im $d$-dimensionalen Raum gegeben durch die Menge der Punkte $\mathbf{x} \in \mathbb{R}^d$ , welche die lineare Gleichung

$$\mathbf{w}_0 + \mathbf{w}_1\mathbf{x}_1 + \mathbf{w}_2\mathbf{x}_2 + \cdots + \mathbf{w}_d\mathbf{x}_d = 0$$

in Kurzform

$$\mathbf{w}_0 + \mathbf{w}^T\mathbf{x} = 0$$

erfüllen.

# Lineare Algebra
## Hyperebenen

Angenommen ein Punkt $\mathbf{x} \in \mathbb{R}^d$ erfüllt diese Gleichung nicht, dann liegt er auch nicht auf der Hyperebene, sonder auf einer der beiden Seiten. Formal gilt $\mathbf{w}_0 + \mathbf{w}^T\mathbf{x} = 0$, es muss daher

$$\mathbf{w}_0 + \mathbf{w}^T\mathbf{x} > 0$$

oder

$$\mathbf{w}_0 + \mathbf{w}^T\mathbf{x} < 0$$

gelten. Das bedeutet, das Vorzeichen

$$\mathrm{sgn}(\mathbf{w}_0 + \mathbf{w}^T\mathbf{x}) \in \{-1, 0, 1\}$$

entscheidet, auf welcher der beiden Seiten der Hyperbene der Punkt $\mathbf{x} \in \mathbb{R}^d$ liegt bzw. ob er auf ihr liegt.

# Lineare Algebra

## Hyperebenen

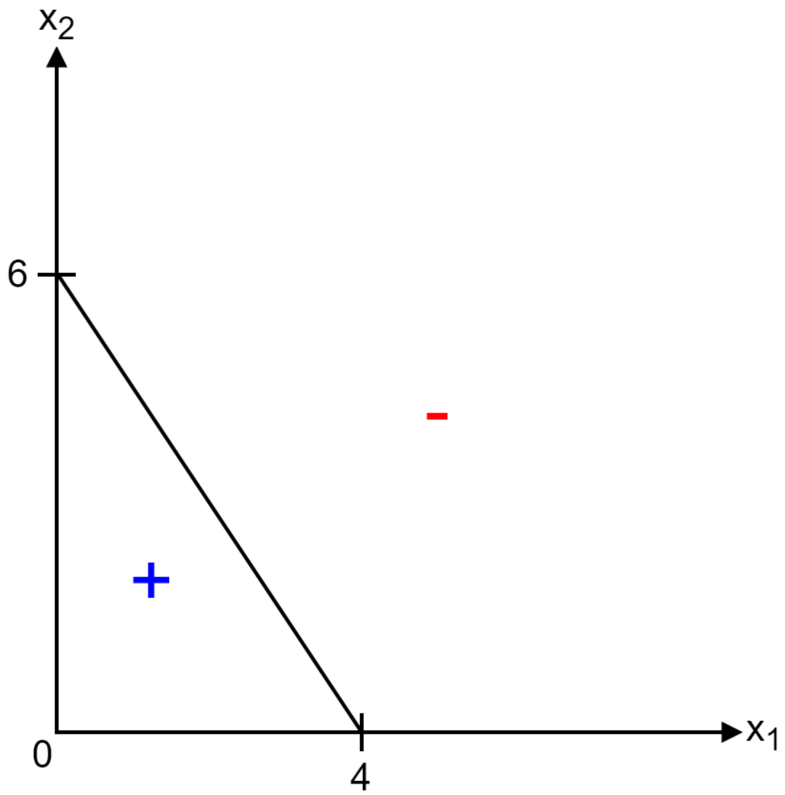


Abbildung 1: Eine Gerade $1 - \frac{1}{4}\mathbf{x}_1 - \frac{1}{6}\mathbf{x}_2 = 0$ als Beispiel einer Hyperebene im 2-dimensionalen Raum. Sie teilt mit $1 - \frac{1}{4}\mathbf{x}_1 - \frac{1}{6}\mathbf{x}_2 > 0$ und $1 - \frac{1}{4}\mathbf{x}_1 - \frac{1}{6}\mathbf{x}_2 < 0$ den Raum in einen positiven und einen negativen Halbraum.

# Multivariate Analysis

Multivariate Analysis

# Multivariate Analysis

## Kettenregel ('Nachdifferenzieren')

Wenn Variable $z$ von $y$ abhängt und $y$ von $x$, dann gilt

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$$

## Beispiel

Für $f(\mathbf{x}) = g(h(x)) = \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)^2$ und daher $g(x) = \frac{1}{2}x^2$ und $h(\mathbf{x}) = \mathbf{x}_1 - \mathbf{x}_2$ gilt

$$\frac{\partial f}{\partial \mathbf{x}_2} = \frac{\partial g}{\partial h}\frac{\partial h}{\partial \mathbf{x}_2} = h(\mathbf{x})(-1) = -(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{x}_2 - \mathbf{x}_1.$$

# Multivariate Analysis

## Partielle Ableitung

Für eine multivariate Funktion $f : \mathbb{R}^n \to \mathbb{R}$ ist die partielle Ableitung $\frac{\partial f}{\partial \mathbf{x}_i}$ bzgl. $\mathbf{x}_i$ definiert als

$$\frac{\partial f}{\partial \mathbf{x}_i}(a_1, \ldots, a_n) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h}$$

## Beispiel

$$f(\mathbf{x}) = 2\mathbf{x}_1^3 - 5\mathbf{x}_2^2 + 3$$

$$\frac{\partial f}{\partial \mathbf{x}_1} = 6\mathbf{x}_1^2, \quad \frac{\partial f}{\partial \mathbf{x}_2} = -10\mathbf{x}_2$$

# Multivariate Analysis

## Gradient

Für eine multivariate Funktion $f : \mathbb{R}^n \to \mathbb{R}$ ist der Gradient $\nabla f$ definiert als der Vektor der partiellen Ableitungen

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{x}_1} \\ \vdots \\ \frac{\partial f}{\partial \mathbf{x}_n} \end{bmatrix}$$

## Beispiel

$$f(\mathbf{x}) = 2\mathbf{x}_1^3 - 5\mathbf{x}_2^2 + 3$$

$$\nabla f = \begin{bmatrix} 6\mathbf{x}_1^2 \\ -10\mathbf{x}_2 \end{bmatrix}$$