Maschinelles Lernen 01

Prof. Dr. David Spieler – david.spieler@hm.edu

Hochschule München

8. Oktober 2019

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 $f : \mathbb{R}^n \to \mathbb{R}$ 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 $\mathbf{I}$

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

and

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

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

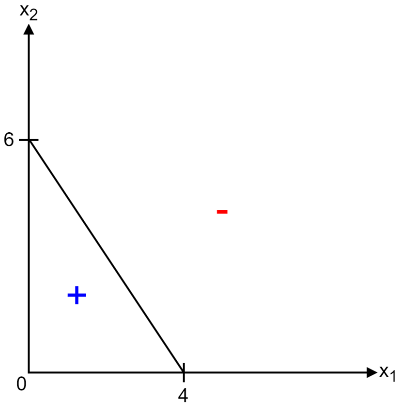

Abbildung 1: Eine Gerade $1 - \frac{1}{4}\mathbf{x}_1 - \frac{1}{6}\mathbf{x}_2 = 0$ als Beispiel einer Hyperebene im 2-dimensionalen Raum. Sie teilt mit $1 - \frac{1}{4}\mathbf{x}_1 - \frac{1}{6}\mathbf{x}_2 > 0$ und $1 - \frac{1}{4}\mathbf{x}_1 - \frac{1}{6}\mathbf{x}_2 < 0$ den Raum in einen positiven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

Multivariate Analysis

Multivariate Analysis

Multivariate Analysis

Kettenregel

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 partialle 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

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

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

Multivariate Analysis

Gradient

Für eine multivariate Funktion $f : \mathbb{R}^n \to \mathbb{R}$ ist der Gradient ∇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

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

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

Einführung

Einführung

Einführung

Was ist maschinelles Lernen?

Beispiele:

- Spracherkennung
- Zeitreihenanalyse
- Künstliche Intelligenz / Bots in Computerspielen
- Gesichtserkennung

Frage:

Was haben alle diese Beispiele gemeinsam?

Einführung

Was ist maschinelles Lernen?

Paradigmenwechsel

Für alle diese Beispiele ist es relativ schwierig, entsprechenden Programmcode manuell zu schreiben. Beim maschinellen Lernen (ML) wird daher ein anderes Paradigma verwendet.

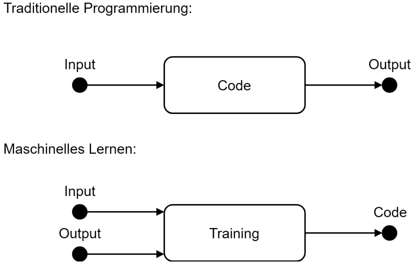

Abbildung 2: Paradigmenwechsel von manuell geschriebenem Code zu trainierten Modellen.

Einführung

Was ist maschinelles Lernen?

Ziel des maschinellen Lernens ist es, Verständnis über Daten zu gewinnen und Vorhersagen bzgl. potentiell neuartiger Daten treffen zu können. Grundsätzlich gibt es drei verschiedene Lernmethoden

- Überwachtes Lernen (Supervised Learning)
- Unüberwachtes Lernen (Unsupervised Learning)
- Bestärkendes Lernen (Reinforcement Learning)

In diesem Kurs werden wir uns mit den ersten beiden Methoden beschäftigen.

Einführung

Überwachtes Lernen

Beim überwachten Lernen versuchen wir eine Funktion

$$f : \mathcal{X} \to \mathcal{Y}$$

zu finden, welche den Zusammenhang zwischen den potentiell mehrdimensionalen Mengen $\mathcal{X}$ und $\mathcal{Y}$ möglichst gut repräsentiert, denn meistens werden wir eine perfekte Abbildung aufgrund von statistischen Effekten nicht erreichen. Dabei gibt es zwei Arten von Fehlern:

- reduzierbar z.B. durch eine bessere Funktion f
- nicht reduzierbar z.B. aufgrund von Messfehlern in den Daten

Einführung

Überwachtes Lernen

Modell

Wir nennen eine Repräsentation von f mathematisch aber auch als Datenstruktur im Computer Modell.

Die Dimensionen von

- $\mathcal{X}$ werden Eingabevariablen, Prädiktoren, unabhängige Variablen oder Features
- $\mathcal{Y}$ werden Ausgabevariablen, Responses oder abhängige Variablen

genannt.

Einführung

Überwachtes Lernen

Grundsätzlich gibt es beim überwachten Lernen zwei grobe Zielsetzungen zwischen denen meist abgewogen werden muss:

- Vorhersage: Gewünscht ist eine möglichst gute Vorhersage $y = f(\mathbf{x})$ wobei die Funktionsweise von f im Extremfall eine Blackbox sein kann.
- Inferenz: Hier steht die Interpretierbarkeit von f im Vordergrund, z.B. Aussagen welche Prädiktoren für welchen Response relevant sind oder auch welcher Zusammenhang (linear, quadratisch, etc.) genau besteht.

Einführung
Überwachtes Lernen

Auch für die Herangehensweise gibt es im Großen und Ganzen zwei Möglichkeiten:

- Parametrische Methoden: Hier wird zunächst eine Annahme bzgl. einer parametrisierten Struktur von f gemacht und diese Parameter werden schließlich mit Hilfe von Daten bestimmt.
- Nicht-parametrische Methoden: Es wird keine Annahme bzgl. der Struktur von f gemacht und es wird versucht f möglichst direkt mit Hilfe von Daten zu definieren.

Einführung

Überwachtes Lernen

Üblicherweise kennen wir die Mengen $\mathcal{X}$ und $\mathcal{Y}$, aber die genaue Abbildung f können wir trotzdem nur anhand von vielen Beispielen

$$\mathcal{D} = \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}) \mid \mathbf{x}^{(i)} \in \mathcal{X}, \mathbf{y}^{(i)} \in \mathcal{Y}, 1 \leq i \leq n\}$$

erahnen.

Trainingsdatensatz

Wir nennen eine solche Menge an Beispielen, die wir für den Lernprozess verwenden Trainingsdatensatz.

Wir sprechen bei der Menge $\mathcal{D}$ auch von gelabelten Daten. Oft muss ein großer (manueller) Aufwand investiert werden, um an solche Daten zu gelangen.

Einführung

Überwachtes Lernen

Üblicherweise ist $\mathcal{X}$ ein d-dimensionaler reellwertiger Vektorraum, im allgemeinen ist also $\mathcal{X} = \mathbb{R}^d$ für ein $d \in \mathbb{N}$.

Beispiele

- $\mathcal{X} = \mathbb{R}$: Temperatur in °C
- $\mathcal{X} = \mathbb{R}^2$: Temperatur in °C und Windgeschwindigkeit in $\frac{\mathrm{m}}{\mathrm{s}}$
- $\mathcal{X} = \mathbb{R}^{16384}$: Graustufenbild 128×128 Pixel (Grauwerte von 0.0 bis 1.0)

Hier wird auch klar, warum wir meist (außer für Beispiele zu Illustrationszwecken) keine einfachen Wertetabelle für f verwenden können.

Einführung

Überwachtes Lernen

Ist $\mathcal{Y}$ eine diskrete Menge, das heißt $\mathcal{Y} = \{C_1, \ldots, C_k\}$ für ein $k \in \mathbb{N}$, dann handelt es sich um ein Klassifikationsproblem. Bei der Klassifikation sind wir an qualitativen Aussagen interessiert. Die einzelnen Objekte $C_1, \ldots, C_k$ werden Klassen oder Kategorien genannt.

Einführung

Überwachtes Lernen

Beispiel: Binäre Klassifikation mit $|\mathcal{Y}| = 2$

Temperaturklassifikation nach menschlichem Empfinden:

$$f : \mathbb{R} \to \{\text{angenehm}, \text{unangenehm}\}$$

$$f(x) = \begin{cases} \text{angenehm} & \text{falls } x \in [18.0, 25.0] \\ \text{unangenehm} & \text{andernfalls.} \end{cases}$$

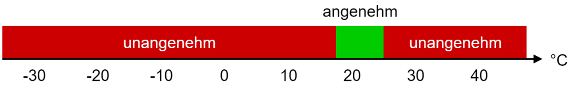

Abbildung 3: Temperaturklassifikation.

Einführung

Überwachtes Lernen

Natürlich kann es wie in der Definition beschrieben auch mehrere Klassen geben.

Beispiel: Mehrklassen-Klassifikation mit $|\mathcal{Y}| = 5$

Temperaturklassifikation nach menschlichem Empfinden:

$$f : \mathbb{R} \rightarrow \{\text{frostig}, \text{kalt}, \text{angenehm}, \text{warm}, \text{heiß}\}$$

$$f(x) = \begin{cases} \text{frostig} & \text{falls } x \in (-\infty, 4.0) \\ \text{kalt} & \text{falls } x \in [4.0, 18.0) \\ \text{angenehm} & \text{falls } x \in [18.0, 25.0) \\ \text{warm} & \text{falls } x \in [25.0, 35.0) \\ \text{heiß} & \text{falls } x \in [35.0, \infty) \end{cases}$$

Einführung
Überwachtes Lernen

Ist $\mathcal{Y}$ eine kontinuierliche Menge, das heißt $\mathcal{Y} \subseteq \mathbb{R}$, dann handelt es sich um ein Regressionsproblem. Bei der Regression sind wir an quantitativen Aussagen interessiert.

Einführung

Überwachtes Lernen

Ein Beispiel einer Regression ist ein *linearer Zusammenhang* zwischen der Temperatur und der Anzahl der verkauften Eiskugeln in einer Eisdiele.

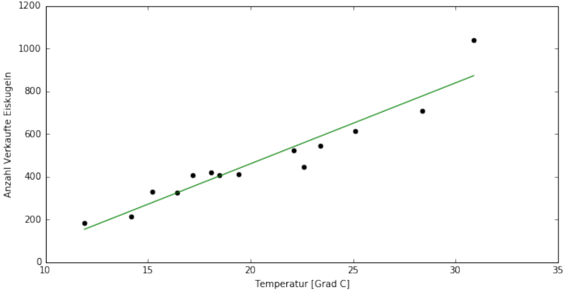

Abbildung 4: Linearer Zusammenhang zwischen der Temperatur und der Anzahl der verkauften Einkugeln, $f : \mathbb{R} \to \mathbb{R}$ mit $f(x) = -320 + 4x$.

Einführung
Überwachtes Lernen

Die Ausgabemenge $\mathcal{Y}$ kann prinzipiell auch mehrdimensional sein.

Beispiele

- $\mathcal{Y} = \{\text{gut}, \text{schlecht}\} \times \{\text{günstig}, \text{normal}, \text{teuer}\}$
- $\mathcal{Y} = \mathbb{R}^2$: Anzahl verkaufte Eiskugeln, Anzahl verkaufte Pizzen

Einführung

Unüberwachtes Lernen

Beim unüberwachten Lernen versucht man ohne Zuhilfenahme von gelabelten Daten einen Mehrtwert zu erhalten. Das Ziel ist daher ausgehend von einer Menge von Daten

$$\mathcal{D} = \{\mathbf{x}^{(i)} \mid \mathbf{x}^{(i)} \in \mathcal{X}, 1 \leq i \leq n\}$$

mehr über die Beschaffenheit von $\mathcal{X}$ herauszubekommen, um dieses Wissen dann direkt oder indirekt anwenden zu können.

Einführung

Unüberwachtes Lernen

Beispiele

- Lernen der Verteilung von $\mathcal{X}$ z.B. bei Sprachmodellen (Welche Wörter folgen auf ein bestimmtes Wort oder einen Satz).
- Dimensionsreduktion zur Verbesserung von überwachten Lernverfahren, z.B. $\mathbf{X} = \mathbb{R}^{10}$ statt $\mathbf{X} = \mathbb{R}^{100}$ für $f : \mathcal{X} \to \mathcal{Y}$
- Finden von Ähnlichkeitsstrukturen durch Clustering

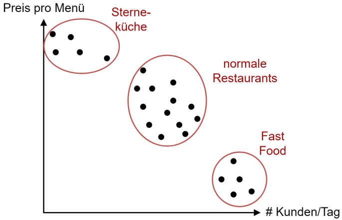

Abbildung 5: Clustering von Restaurants.

Einführung

Datenvisualisierung

Wenn man ein Projekt mit maschinellen Lernmethoden beginnt, ist es ratsam, sich zunächst einen Überblick über die Daten zu verschaffen. Meist gelingt dies am besten, wenn man die Daten geeignet visualisiert. Im Folgenden finden Sie einige Beispiele verschiedener Diagrammtypen.

Einführung

Datenvisualisierung

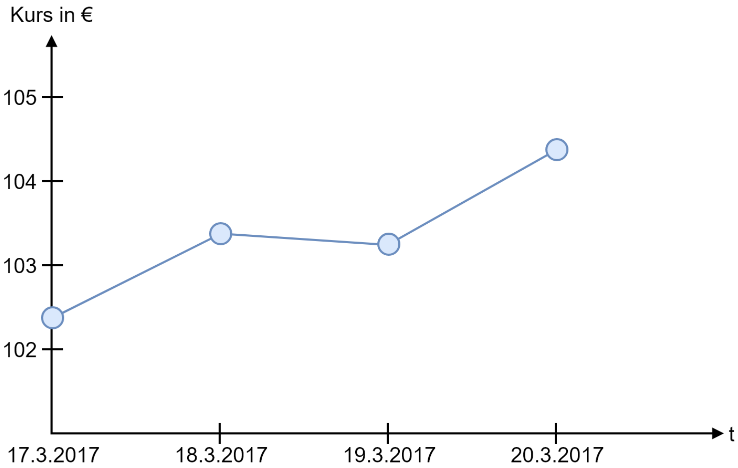

Abbildung 6: Beispiel eines Liniendiagramms.

Einführung

Datenvisualisierung

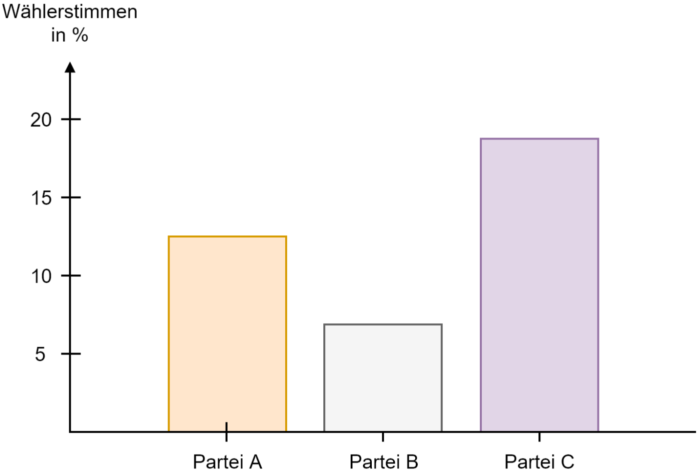

Abbildung 7: Beispiel eines Balkendiagramms.

Einführung
Datenvisualisierung

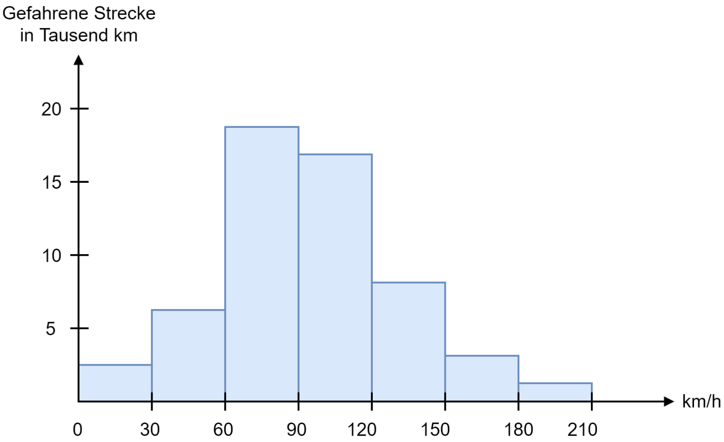

Abbildung 8: Beispiel eines Histogramms – eines speziellen Balkendiagramms.

Einführung

Datenvisualisierung

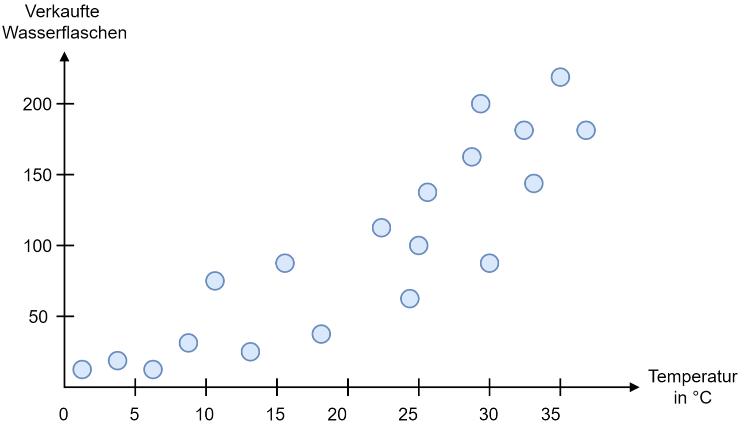

Abbildung 9: Beispiel eines Streudiagramms.

Einführung
Datenvorverarbeitung

Bevor tatsächlich ein ML Modell erstellt und trainiert wird müssen die entsprechenden Daten vorverarbeitet werden. Dazu gehören grundsätzlich drei Schritte

1. Auswahl
2. Aufbereitung
3. Transformation

der Daten. Oftmals muss auch aufgrund neuer Erkenntnisse zwischen den Schritten hin und her gewechselt werden.

Einführung
Datenvorverarbeitung

Auswahl: Nicht immer sind mehr Daten auch wirklich besser, d.h. es sollte darauf geachtet werden, dass nur für den Anwendungszweck relevante Daten verwendet werden, um die Rechen- und Speicheranforderungen im Rahmen zu halten. Auch die Leistung des Systems könnte u.U. unter zu vielen bzw. den falschen Daten leiden – natürlich auch unter zu wenig.

Einführung

Datenvorverarbeitung

Fragestellungen, die bzgl. der Auswahl helfen:

- Auf welche Daten hat man Zugriff?
- Welche Daten kann man mit welchem Aufwand erstellen bzw. simulieren?
- Auf welchen Teil der Daten kann/sollte man verzichten?

Starthilfe

Im Rahmen von Wettbewerben und Benchmarks werden immer wieder Datensätze veröffentlicht, die zum Lernen von ML Techniken verwendet werden können. Ein Beispiel ist `https://www.kaggle.com/datasets`.

Einführung

Datenvorverarbeitung

Aufbereitung:

- Definition eines geeigneten Format (Tabellen, Big Data Formate wie Parquet, CSV, Bilder, etc.) und Umwandlung der Daten
- Bereinigung, d.h. Entfernung von *unvollständigen* oder *ungültigen* Daten oder aufgrund von rechtlichen Bestimmungen (Datenschutz)
- Unterauswahl der Daten (lange Laufzeit, großer Speicheraufwand). Hier muss auf eine *repräsentative* Auswahl (Zeit, Ort, Gruppen, etc.) geachtet werden, um keinen systematischen Fehler einzuführen.

Einführung
Datenvorverarbeitung

Transformation:

- Skalierung: Features in den geeigneten Wertebereich für ML Methode bringen, z.B. auf Wertebereiche $[0, 1]$ oder $[-1, 1]$. Auch eine Normierung auf Mittelwert 0 und Standardabweichung 1 kann notwendig sein.
- Zerlegung in sinnvolle Features, z.B. Extraktion der Zeit und des Fehlercodes aus Logfile-Eintägen
- Aggregation mehrerer Features, z.B. Gesamtzahl der Aktienverkäufe an einem Tag statt jede Einzeltransaktion