Kapitel 2 - Mathematische Grundlagen

Annika Liebgott

February 12, 2024

2.1 - Korrelation und Faltung

2.1 - Korrelation und Faltung

2.1.1 - Korrelation

Kovarianzmatrix

Die Kovarianzmatrix, auch Varianz-Kovarianz-Matrix genannt, zeigt den Zusammenhang von mehr als 2 Zufallsvariablen:

$$\mathbf{C} = \begin{pmatrix} c_{x_1,x_1} & c_{x_1,x_2} & \cdots & c_{x_1,x_n} \\ c_{x_2,x_1} & c_{x_2,x_2} & \cdots & c_{x_2,x_n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{x_n,x_1} & c_{x_n,x_2} & \cdots & c_{x_n,x_n} \end{pmatrix} = \begin{pmatrix} \sigma_{x_1}^2 & c_{x_1,x_2} & \cdots & c_{x_1,x_n} \\ c_{x_2,x_1} & \sigma_{x_2}^2 & \cdots & c_{x_2,x_n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{x_n,x_1} & c_{x_n,x_2} & \cdots & \sigma_{x_n}^2 \end{pmatrix}$$

Viele im Bereich ML wichtige Methoden verwenden Kovarianzmatrizen für Berechnungen (z.B. Optimierungsverfahren, Merkmalstransformationen,...)

Korrelationskoeffizient ρ

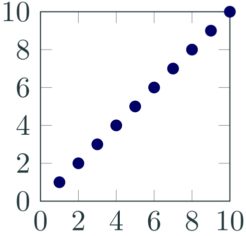
Bedeutung verschiedener Werte für ρ:

- $\rho = 0$: $x(n), y(n)$ unkorreliert
- $\rho \neq 0$: $x(n), y(n)$ korreliert
    - $\rho < 0$: $x(n), y(n)$ negativ korreliert
    - $\rho > 0$: $x(n), y(n)$ positiv korreliert
    - $|\rho| = 1$: $x(n), y(n)$ kohärent (maximal positiv bzw. negativ korreliert)

Korrelationskoeffizient ρ

Bedeutung verschiedener Werte für ρ:

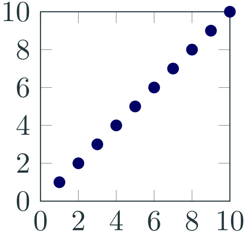
- $\rho = 0$: $x(n), y(n)$ unkorreliert
- $\rho \neq 0$: $x(n), y(n)$ korreliert
    - $\rho < 0$: $x(n), y(n)$ negativ korreliert
    - $\rho > 0$: $x(n), y(n)$ positiv korreliert
    - $|\rho| = 1$: $x(n), y(n)$ kohärent (maximal positiv bzw. negativ korreliert)



Korrelationskoeffizient ρ

Bedeutung verschiedener Werte für ρ:

- $\rho = 0$: $x(n), y(n)$ unkorreliert
- $\rho \neq 0$: $x(n), y(n)$ korreliert
  - $\rho < 0$: $x(n), y(n)$ negativ korreliert
  - $\rho > 0$: $x(n), y(n)$ positiv korreliert
  - $|\rho| = 1$: $x(n), y(n)$ kohärent (maximal positiv bzw. negativ korreliert)

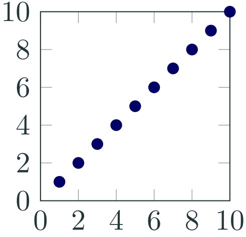


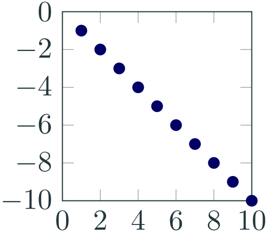
$\rho = 1$

Korrelationskoeffizient ρ

Bedeutung verschiedener Werte für ρ:

- $\rho = 0$: $x(n), y(n)$ unkorreliert
- $\rho \neq 0$: $x(n), y(n)$ korreliert
  - $\rho < 0$: $x(n), y(n)$ negativ korreliert
  - $\rho > 0$: $x(n), y(n)$ positiv korreliert
  - $|\rho| = 1$: $x(n), y(n)$ kohärent (maximal positiv bzw. negativ korreliert)



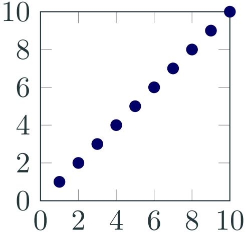


$\rho = 1$

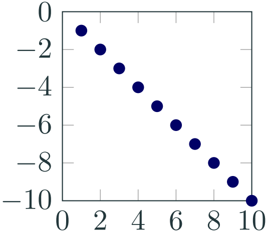
Korrelationskoeffizient ρ

Bedeutung verschiedener Werte für ρ:

- $\rho = 0$: $x(n), y(n)$ unkorreliert
- $\rho \neq 0$: $x(n), y(n)$ korreliert
  - $\rho < 0$: $x(n), y(n)$ negativ korreliert
  - $\rho > 0$: $x(n), y(n)$ positiv korreliert
  - $|\rho| = 1$: $x(n), y(n)$ kohärent (maximal positiv bzw. negativ korreliert)



$\rho = 1$

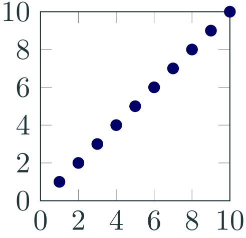


$\rho = -1$

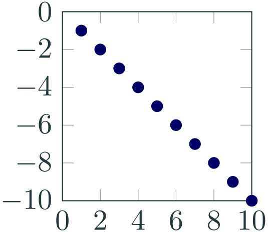
Korrelationskoeffizient ρ

Bedeutung verschiedener Werte für ρ:

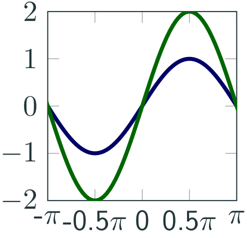
- $\rho = 0$: $x(n), y(n)$ unkorreliert
- $\rho \neq 0$: $x(n), y(n)$ korreliert
  - $\rho < 0$: $x(n), y(n)$ negativ korreliert
  - $\rho > 0$: $x(n), y(n)$ positiv korreliert
  - $|\rho| = 1$: $x(n), y(n)$ kohärent (maximal positiv bzw. negativ korreliert)



$\rho = 1$



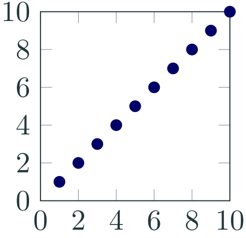
$\rho = -1$



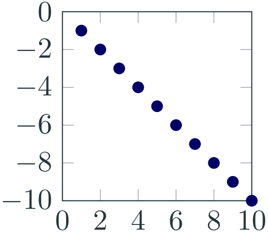
Korrelationskoeffizient ρ

Bedeutung verschiedener Werte für ρ:

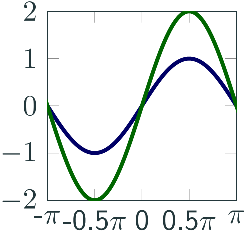
- $\rho = 0$: $x(n), y(n)$ unkorreliert
- $\rho \neq 0$: $x(n), y(n)$ korreliert
    - $\rho < 0$: $x(n), y(n)$ negativ korreliert
    - $\rho > 0$: $x(n), y(n)$ positiv korreliert
    - $|\rho| = 1$: $x(n), y(n)$ kohärent (maximal positiv bzw. negativ korreliert)



$\rho = 1$



$\rho = -1$

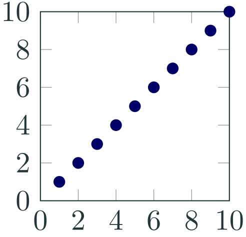


$\rho = 1$

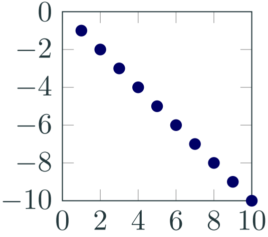
Korrelationskoeffizient ρ

Bedeutung verschiedener Werte für ρ:

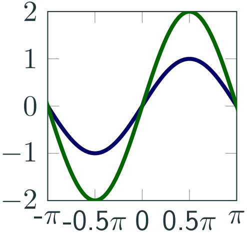
- $\rho = 0$: $x(n), y(n)$ unkorreliert
- $\rho \neq 0$: $x(n), y(n)$ korreliert
    - $\rho < 0$: $x(n), y(n)$ negativ korreliert
    - $\rho > 0$: $x(n), y(n)$ positiv korreliert
    - $|\rho| = 1$: $x(n), y(n)$ kohärent (maximal positiv bzw. negativ korreliert)



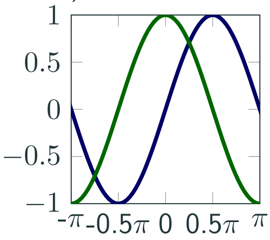
$\rho = 1$



$\rho = -1$



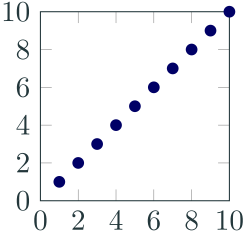
$\rho = 1$



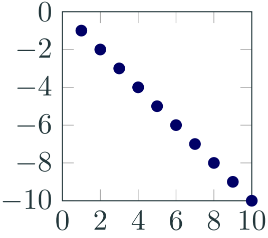
Korrelationskoeffizient ρ

Bedeutung verschiedener Werte für ρ:

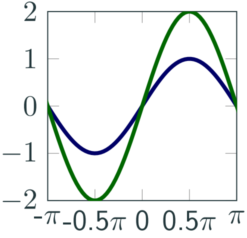
- $\rho = 0$: $x(n), y(n)$ unkorreliert
- $\rho \neq 0$: $x(n), y(n)$ korreliert
    - $\rho < 0$: $x(n), y(n)$ negativ korreliert
    - $\rho > 0$: $x(n), y(n)$ positiv korreliert
    - $|\rho| = 1$: $x(n), y(n)$ kohärent (maximal positiv bzw. negativ korreliert)



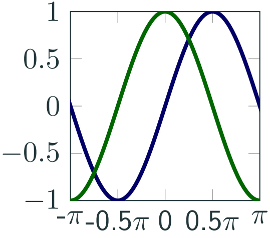
$\rho = 1$



$\rho = -1$

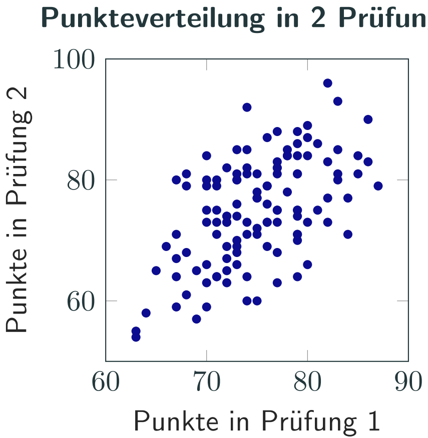


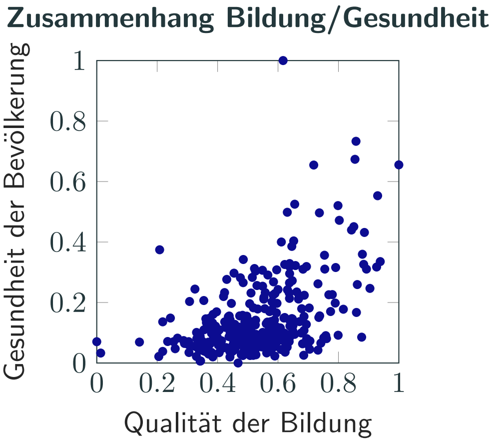
$\rho = 1$



$\rho = 0$

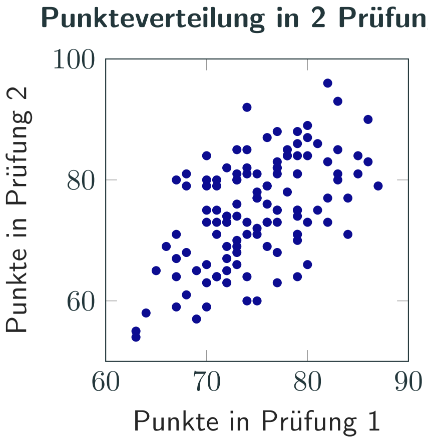
Beispiel 2.1: Korrelationen anhand von Plots abschätzen



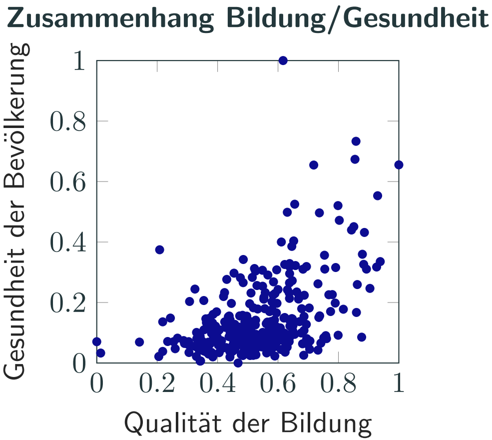


Quelle: Mathworks Sample Data Sets "examgrades" und "cities"

Beispiel 2.1: Korrelationen anhand von Plots abschätzen



$\rho = 0.55$

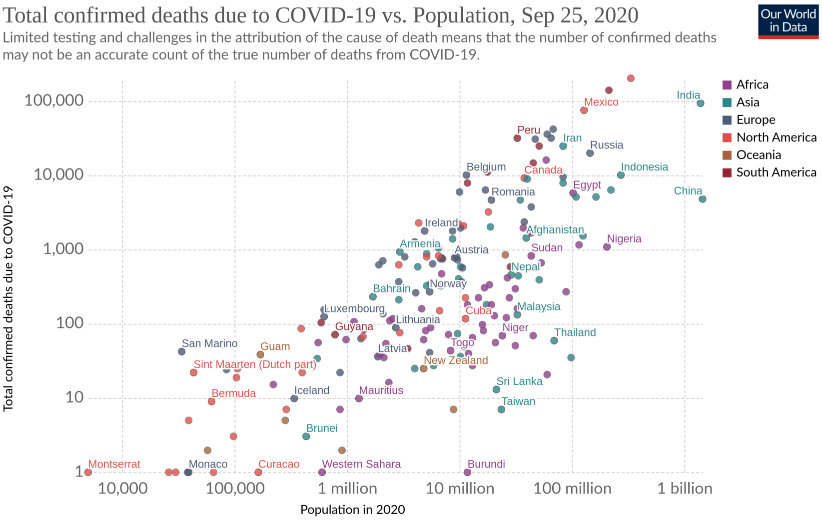


$\rho = 0.49$

Quelle: Mathworks Sample Data Sets “examgrades” und “cities”

Beispiel 2.2: Covid-Statistiken(1)

Frage: besteht eine Korrelation zwischen der Gesamtzahl Todesfälle und der Einwohnerzahl eines Landes?

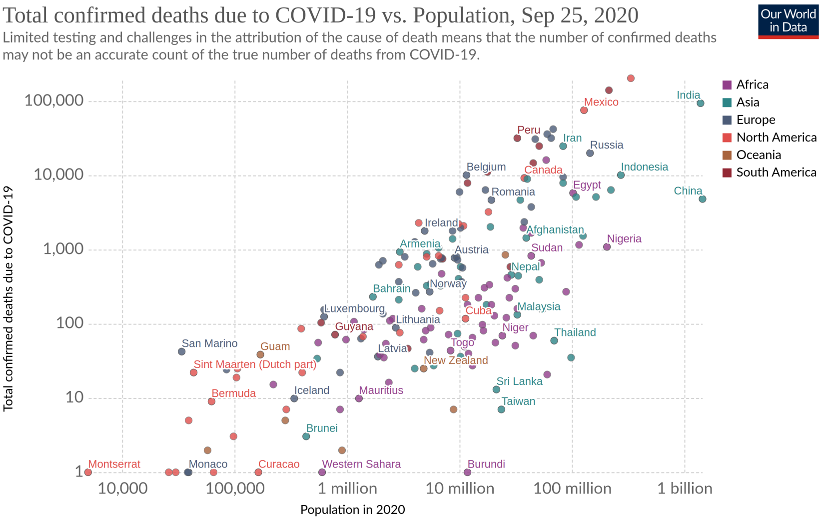


Source: European CDC – Situation Update Worldwide – Last updated 25 September, 10:35 (London time), Population data for 2020 (UNWPP, 2019), Our World In Data
OurWorldInData.org/coronavirus • CC BY

Beispiel 2.2: Covid-Statistiken(1)

Frage: besteht eine Korrelation zwischen der Gesamtzahl Todesfälle und der Einwohnerzahl eines Landes?

1. Intuition: ja, eindeutig!



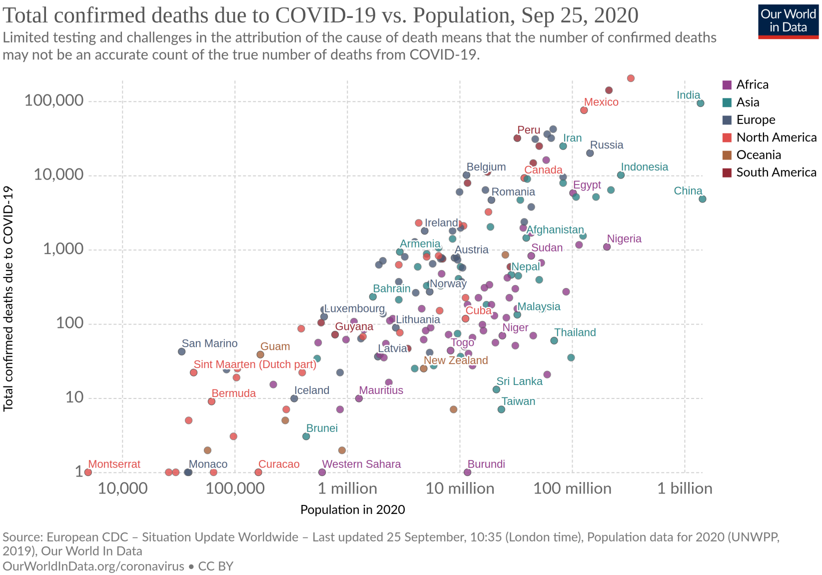
Source: European CDC – Situation Update Worldwide – Last updated 25 September, 10:35 (London time), Population data for 2020 (UNWPP, 2019), Our World In Data
OurWorldInData.org/coronavirus • CC BY

Beispiel 2.2: Covid-Statistiken(1)

Frage: besteht eine Korrelation zwischen der Gesamtzahl Todesfälle und der Einwohnerzahl eines Landes?

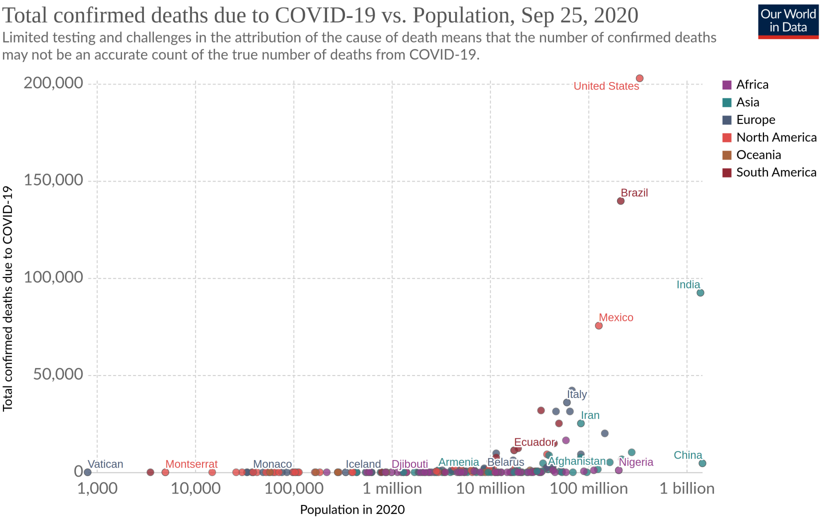
1. Intuition: ja, eindeutig!

Aber: Achsen sind logarithmisch.



Beispiel 2.2: Covid-Statistiken (2)

Frage: besteht eine Korrelation zwischen der Gesamtzahl Todesfälle und der Einwohnerzahl eines Landes?

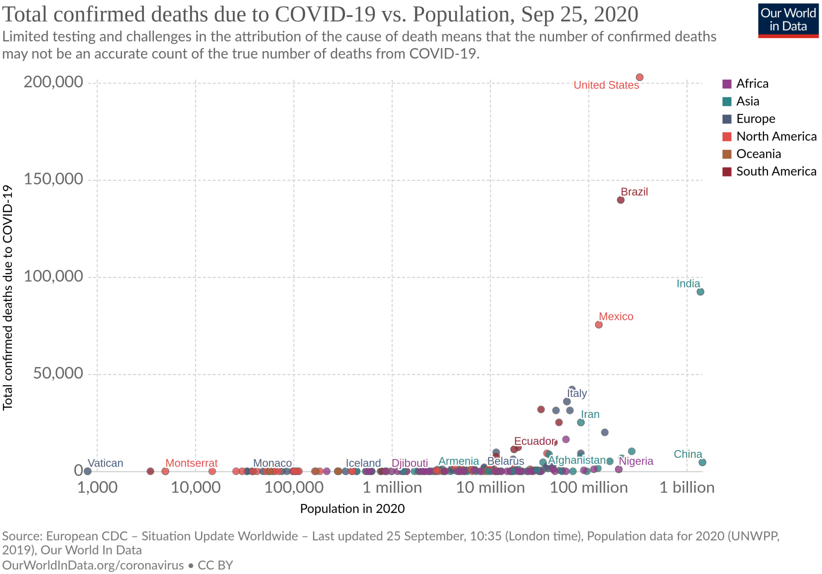


Source: European CDC – Situation Update Worldwide – Last updated 25 September, 10:35 (London time), Population data for 2020 (UNWPP, 2019), Our World In Data
OurWorldInData.org/coronavirus • CC BY

Beispiel 2.2: Covid-Statistiken (2)

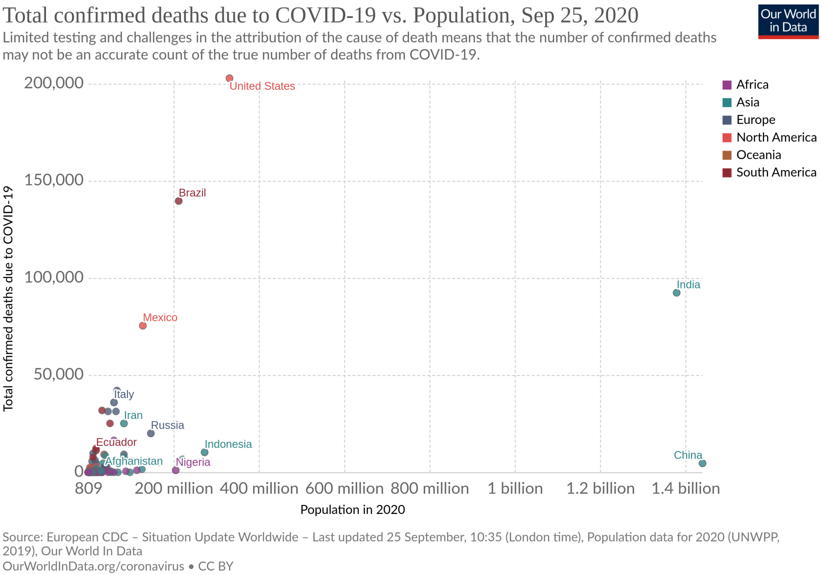
Frage: besteht eine Korrelation zwischen der Gesamtzahl Todesfälle und der Einwohnerzahl eines Landes?

Mit logarithmischer Einwohnerzahl, aber linearen Fallzahlen: Zusammenhang weniger stark ausgeprägt



Beispiel 2.2: Covid-Statistiken (3)

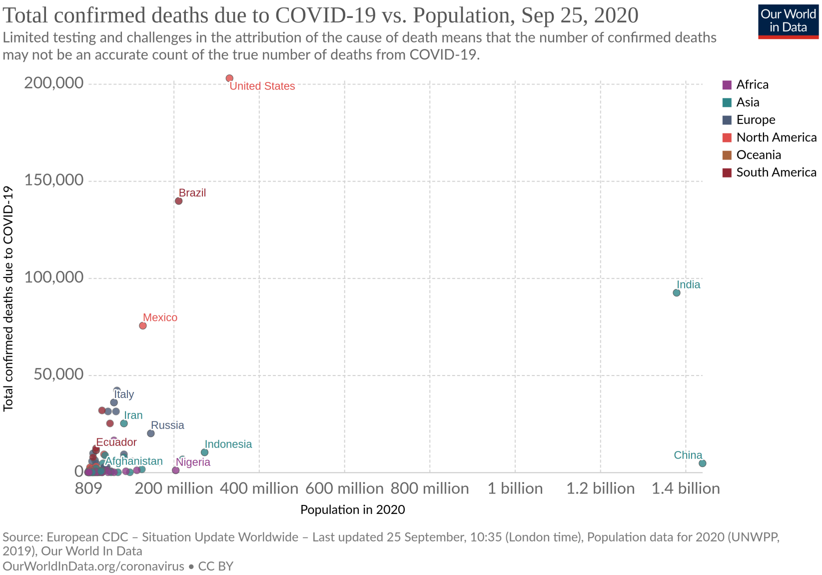
Frage: besteht eine Korrelation zwischen der Gesamtzahl Todesfälle und der Einwohnerzahl eines Landes?



Beispiel 2.2: Covid-Statistiken (3)

Frage: besteht eine Korrelation zwischen der Gesamtzahl Todesfälle und der Einwohnerzahl eines Landes?

Mit linearen Achsen: Zusammenhang fällt auch weniger stark aus

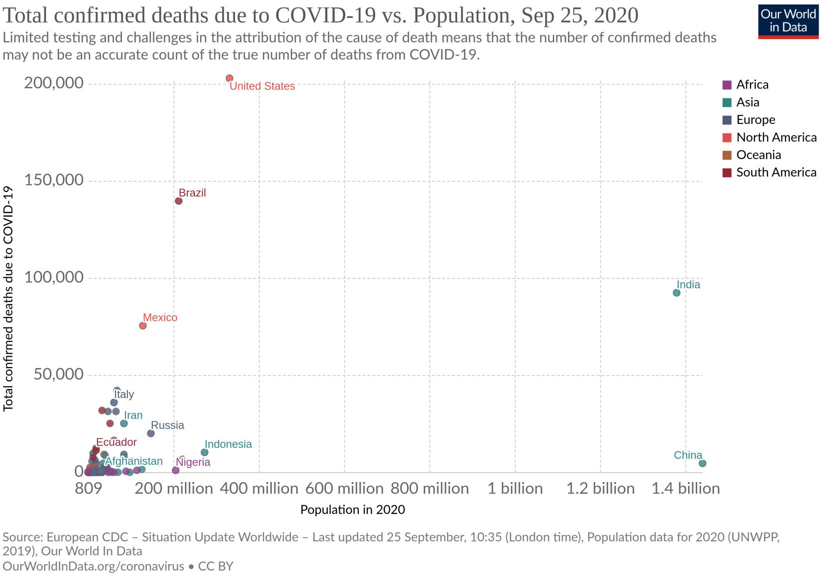


Beispiel 2.2: Covid-Statistiken (3)

Frage: besteht eine Korrelation zwischen der Gesamtzahl Todesfälle und der Einwohnerzahl eines Landes?

Mit linearen Achsen: Zusammenhang fällt auch weniger stark aus

$\Rightarrow \rho$ kein gutes Maß



Verschiedene Korrelationskoeffizienten und ihre Anwendung (1)

Pearson-Korrelationskoeffizient: $\rho = \frac{c_{xy}}{\sigma_x \sigma_y}$, misst den linearen Zusammenhang zweier Signale x, y

Rangkorrelationen: Signale x, y werden in Rangfolgen r_x, r_y konvertiert (d.h. aufsteigend sortiert), anschließend werden die Rangfolgen verglichen.

Unterschiede zu Pearson-Korrelation:

- weniger anfällig gegenüber extremen Ausreißerwerten
- auch geeignet für monotone, aber nichtlineare Zusammenhänge

Verschiedene Korrelationskoeffizienten und ihre Anwendung (2)

Rangkorrelation nach Spearman (ρ_s): r_x, r_y werden gebildet, anschließend wird Pearson-Korrelation zw. r_x und r_y berechnet, statt zw. x und y. Alternativ kann die Berechnung über Rangdifferenz $d_i = r_{x_i} - r_{y_i}$ von Beobachtungsparen (x_i, y_i) erfolgen:

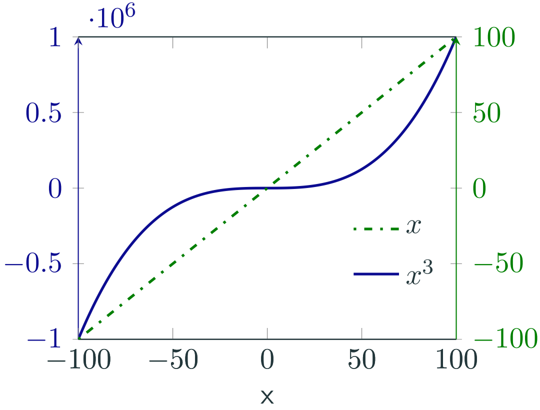
$$\rho_s = \frac{c_{r_x r_y}}{\sigma_{r_x}\sigma_{r_y}} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)}$$

Rangkorrelation nach Kendall (τ): Signalpaare (x_i, y_i) werden nur nach x sortiert, anschließend werden (x_i, y_i) und nachfolgende Wertepaare (x_j, y_j) auf Konkordanz (d.h. $x_i < x_j$ und $y_i < y_j$) bzw. Diskonkordanz (d.h. $x_i < x_j$ und $y_i > y_j$) untersucht. τ basiert auf der Anzahl konkordanter (n_k) und diskonkordanter (n_d) Beobachtungen:

$$\tau = \frac{n_k - n_d}{\frac{N}{2}(N-1)}$$

Verschiedene Korrelationskoeffizienten und ihre Anwendung (3)

Beispiel: Unterschied zwischen ρ, ρ_s und τ für monotone Signale

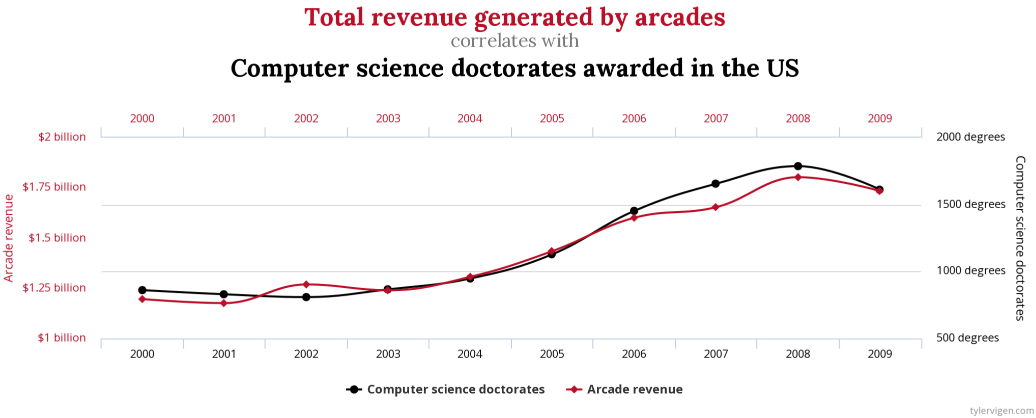


Korrelelation nach Pearson: $\rho = 0.91$

Rangkorrelation nach Spearman: $\rho_s = 1$

Rangkorrelation nach Kendall: $\tau = 1$

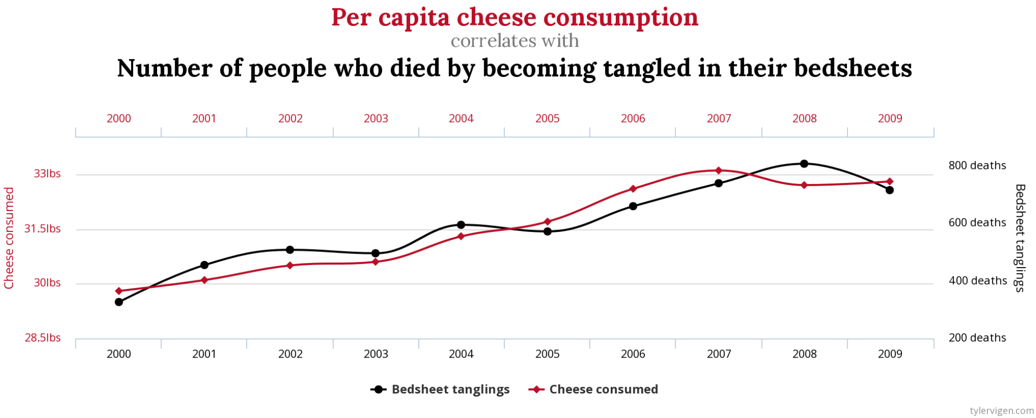
Beispiel 2.3 - Korrelation vs. Kausalität (1)



$\rho = 0.9851$

Quelle: https://www.tylervigen.com/spurious-correlations

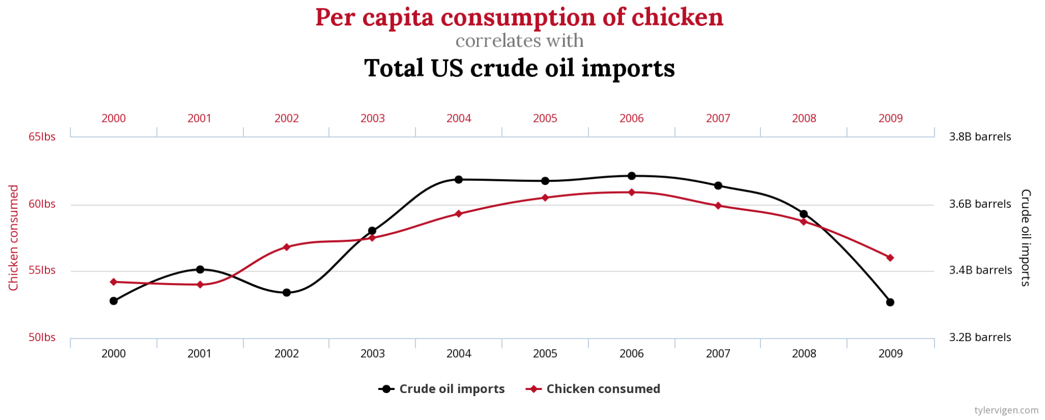
Beispiel 2.3 - Korrelation vs. Kausalität (2)



$\rho = 0.9471$

Quelle: https://www.tylervigen.com/spurious-correlations

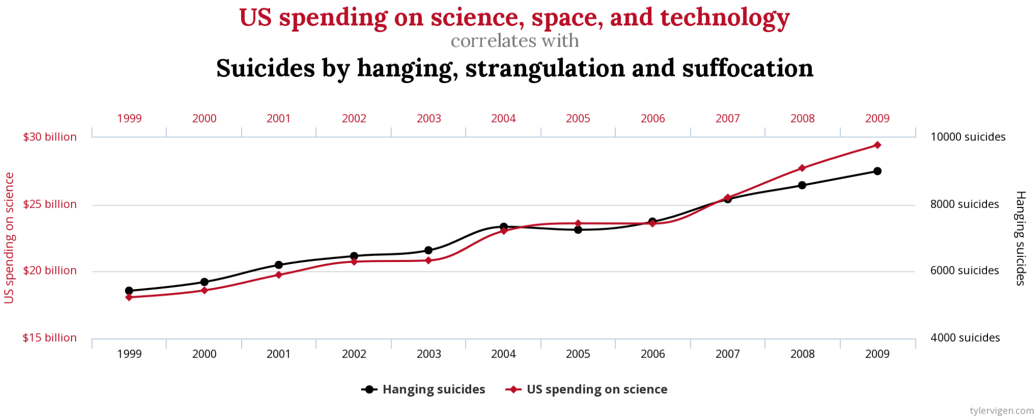
Beispiel 2.3 - Korrelation vs. Kausalität (3)



$\rho = 0.8999$

Quelle: https://www.tylervigen.com/spurious-correlations

Beispiel 2.3 - Korrelation vs. Kausalität (4)



$\rho = 0.9979$

Quelle: https://www.tylervigen.com/spurious-correlations

Der Unterschied zwischen Korrelation und Kausalität

Teilweise werden (z.B. in der Presse) Korrelation und Kausalität fälschlicherweise synonym verwendet. Die Beispiele auf Folien 2.10-2.13 zeigen, wie wichtig die richtige Unterscheidung beider Begriffe ist.

Korrelation: beschreibt den (möglicherweise zufälligen) Zusammenhang zwischen zwei Größen

Kausalität: bedeutet, dass die eine Größe tatsächlich ursächlich von der anderen abhängig ist (Beispiel: höhere Rate an Masernimpfungen führt zu geringerer Anzahl an Masernerkrankungen)

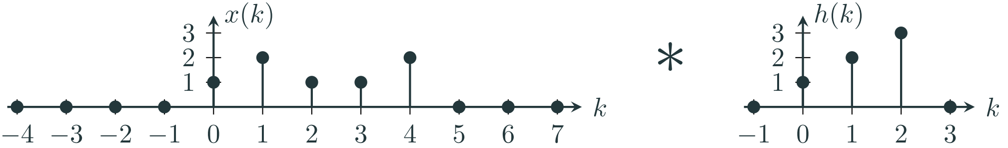
Merke:

Eine hohe Korrelation zweier Größen bedeutet nicht automatisch, dass ein kausaler Zusammenhang zwischen den Variablen vorhanden ist!

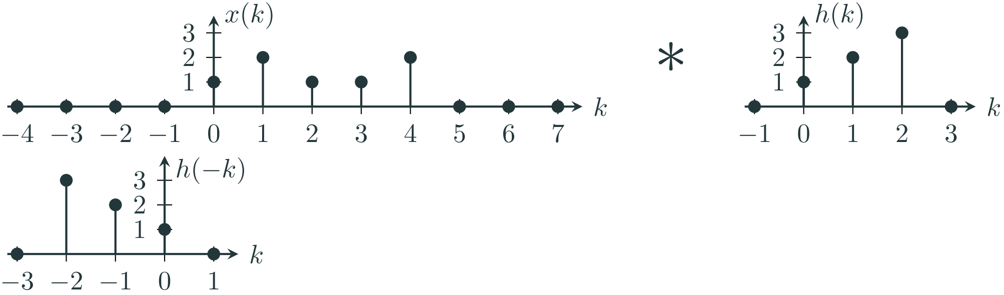
2.1 - Korrelation und Faltung

2.1.2 - Faltung

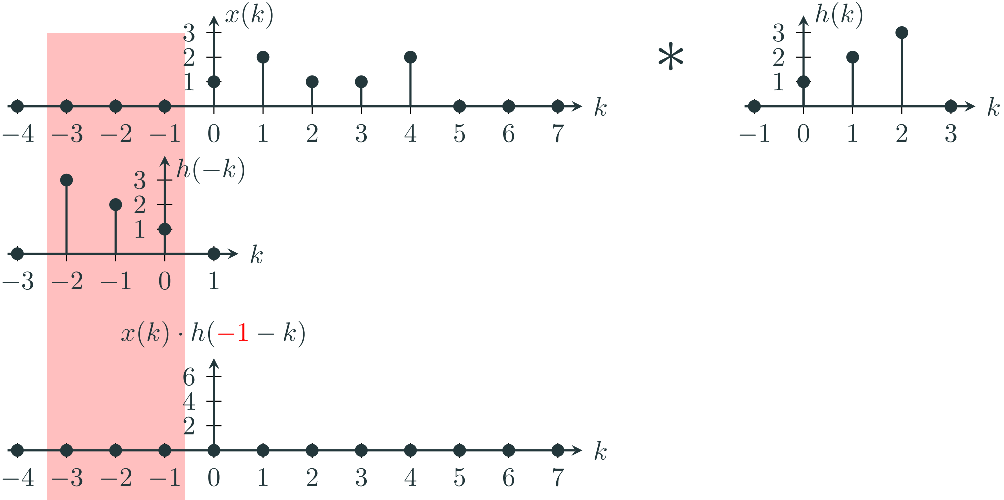
Faltung eindimensional - graphische Veranschaulichung



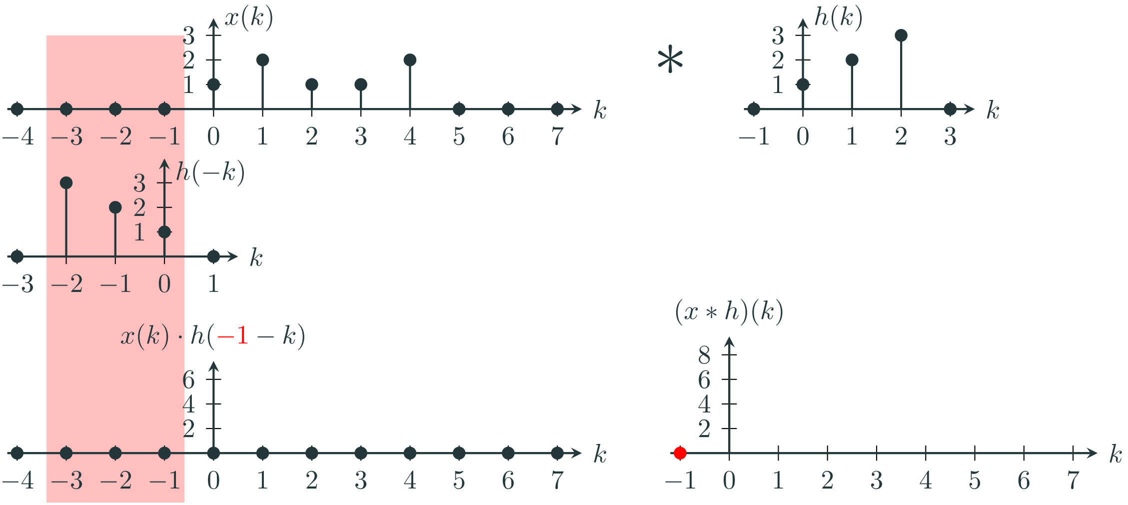
Faltung eindimensional - graphische Veranschaulichung



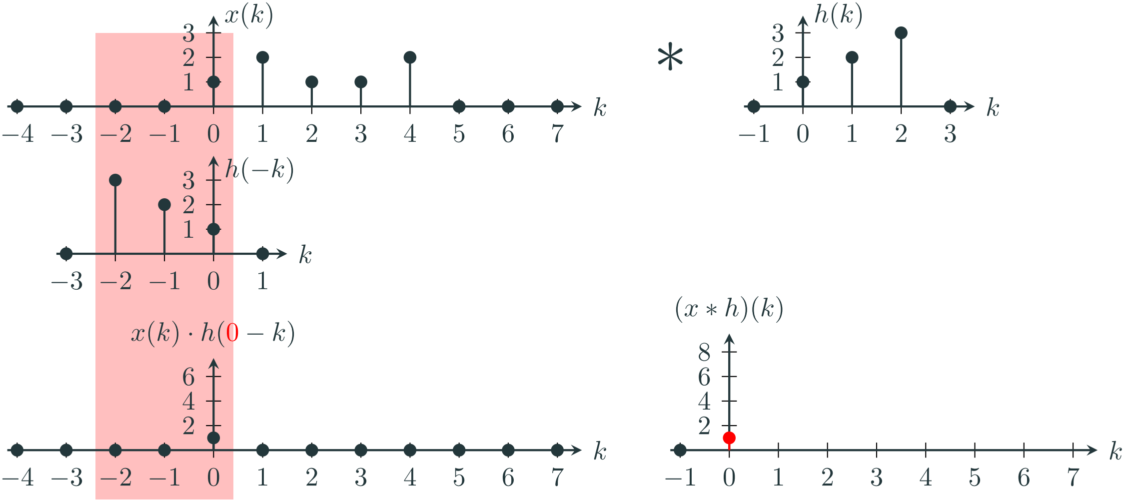
Faltung eindimensional - graphische Veranschaulichung



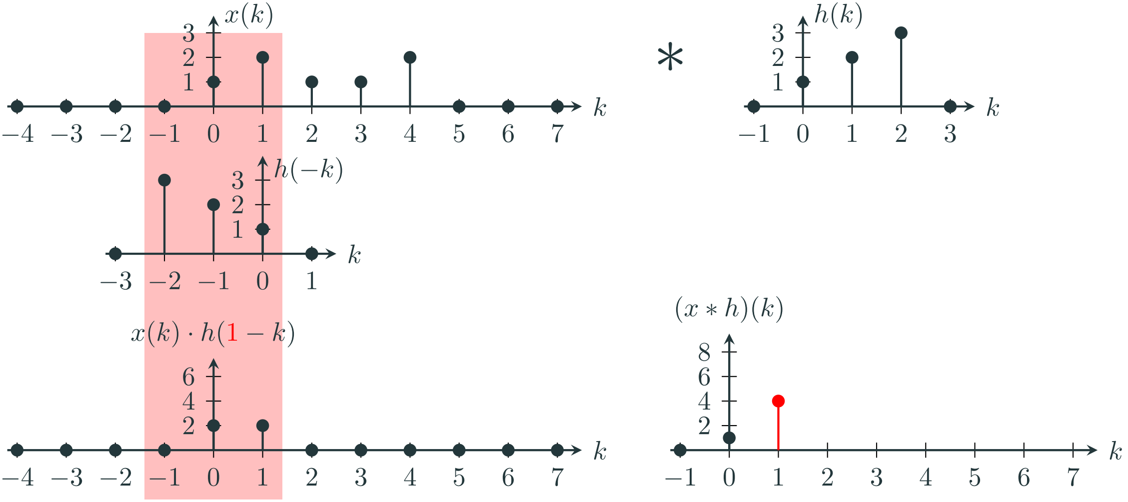
Faltung eindimensional - graphische Veranschaulichung



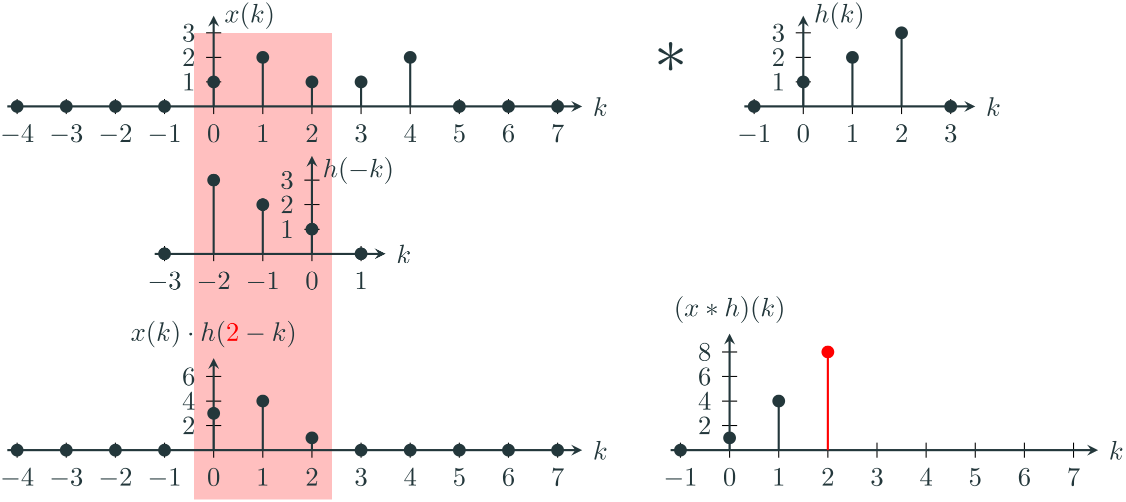
Faltung eindimensional - graphische Veranschaulichung



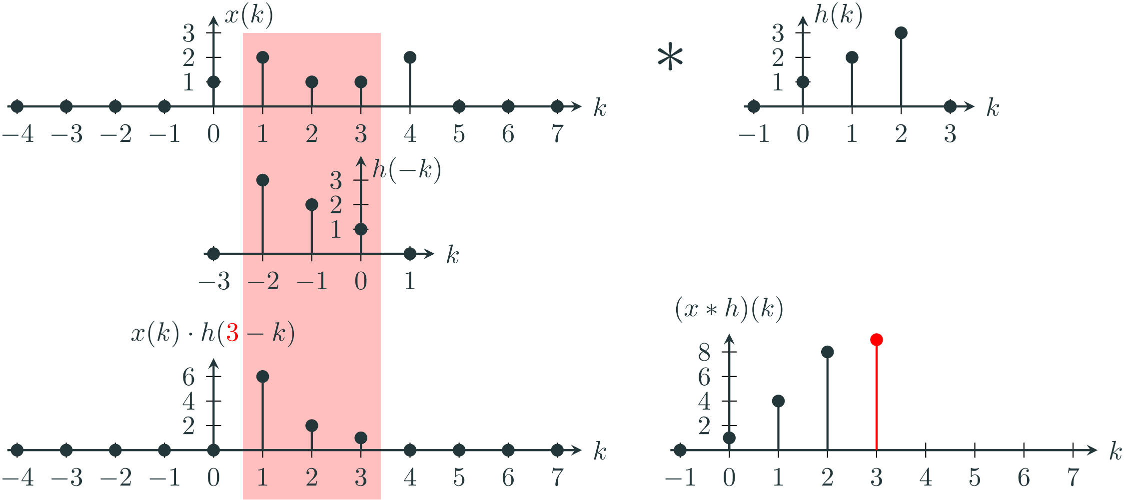
Faltung eindimensional - graphische Veranschaulichung



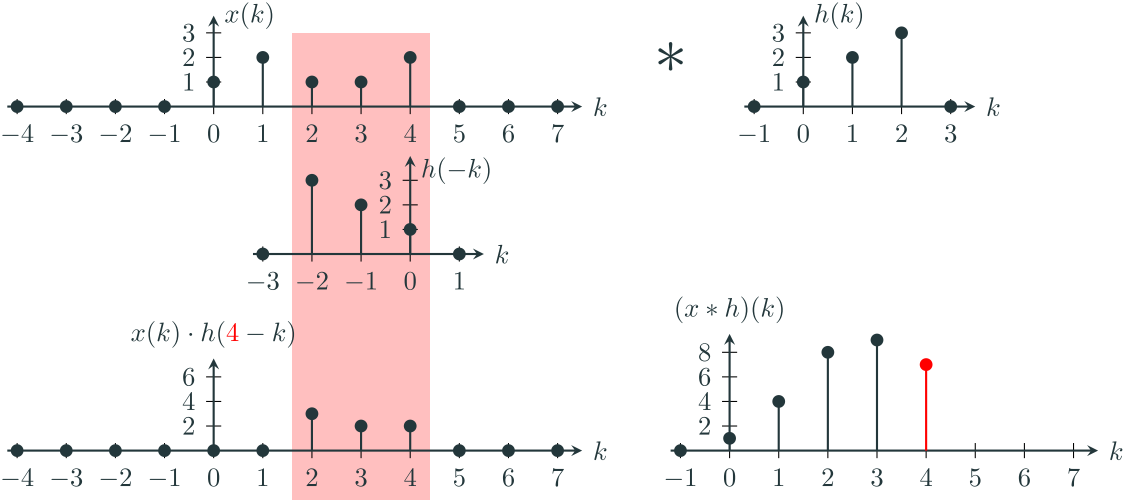
Faltung eindimensional - graphische Veranschaulichung



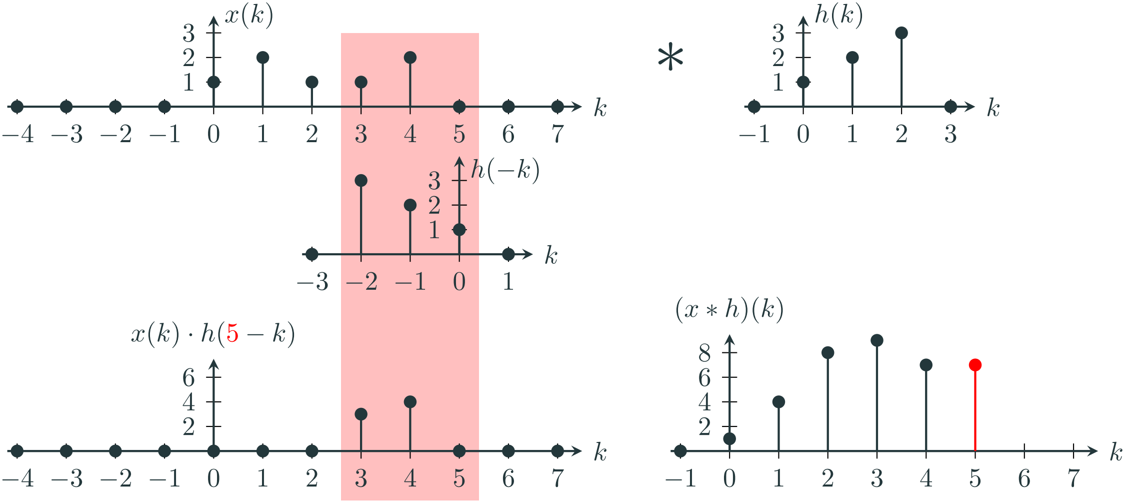
Faltung eindimensional - graphische Veranschaulichung



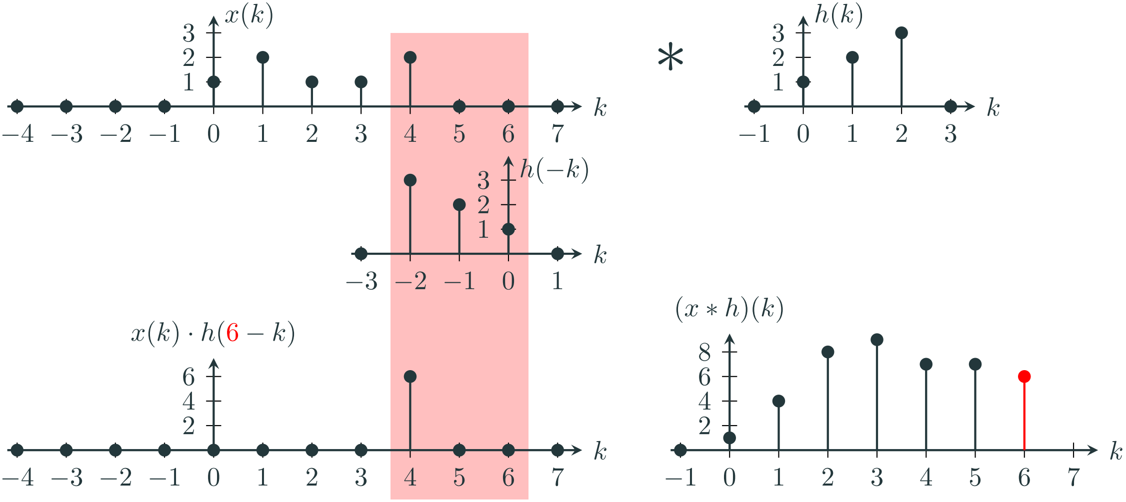
Faltung eindimensional - graphische Veranschaulichung



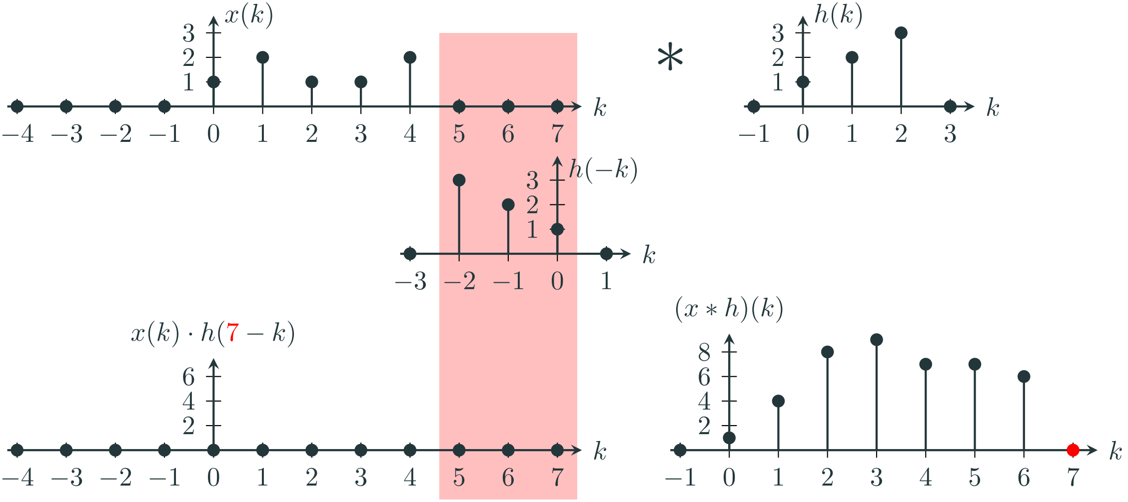
Faltung eindimensional - graphische Veranschaulichung



Faltung eindimensional - graphische Veranschaulichung



Faltung eindimensional - graphische Veranschaulichung



Faltung in der Bildverarbeitung

Grundgedanke: Verschiedene Operationen zur Bildtransformation werden werden durch Faltung des Bildes mit bestimmten Kerneln realisiert.

Beispiele:

- Kantendetektion
- Schärfung
- Glättung

Das resultierende Bild I^* entsteht durch Multiplizieren des Originalbildes I mit einer Kernel-Matrix K, die schrittweise über das Bild geschoben wird.

$$I^*(x, y) = I(x, y) * K(x, y) = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} I(i, j) \cdot K(x - i, y - j)$$

Beispiel 2.4: Faltungsoperationen auf Bildern

$$* \frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix} =$$

$\Rightarrow$ Blurring bzw. Kantenglättung

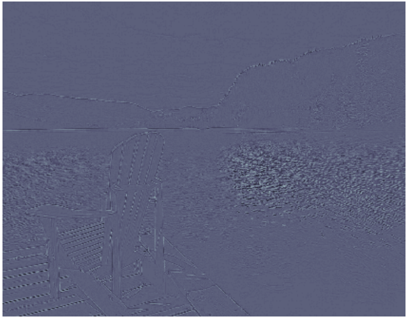
Beispiel 2.4: Faltungsoperationen auf Bildern

$$* \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} =$$

$\Rightarrow$ Schärfung des Bildes

Beispiel 2.4: Faltungsoperationen auf Bildern

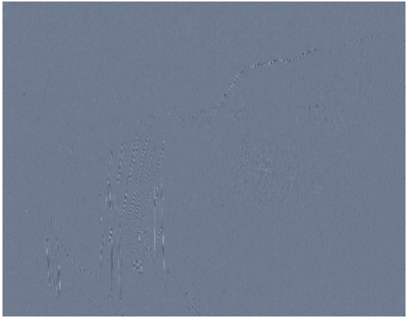
$$* \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} =$$



$\Rightarrow$ Kantendetektion

Beispiel 2.4: Faltungsoperationen auf Bildern

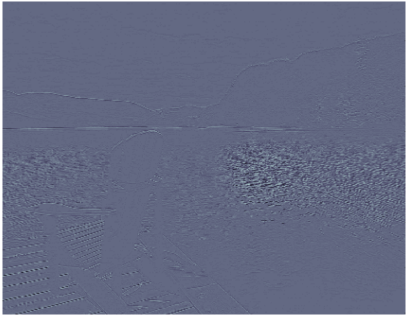
$$* \begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix} =$$



$\Rightarrow$ Detektion vertikaler Strukturen

Beispiel 2.4: Faltungsoperationen auf Bildern

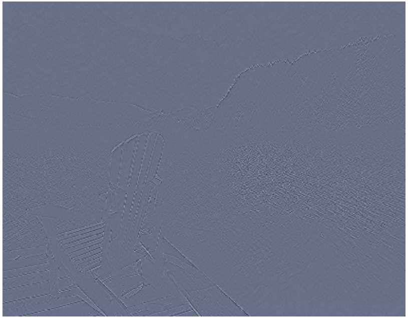
$$* \begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix} =$$



$\Rightarrow$ Detektion horizontaler Strukturen

Beispiel 2.4: Faltungsoperationen auf Bildern

$$* \begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & -1 \end{bmatrix} =$$



$\Rightarrow$ Detektion diagonaler Strukturen

Faltung in der Bildverarbeitung: graphische Veranschaulichung

10	0	5	1
1	8	7	2
5	0	2	6
9	0	1	7

$*$

1	2	4
1	0	1
2	1	2

Faltung in der Bildverarbeitung: graphische Veranschaulichung

$$\begin{array}{|c|c|c|c|}\hline 10 & 0 & 5 & 1 \\ \hline 1 & 8 & 7 & 2 \\ \hline 5 & 0 & 2 & 6 \\ \hline 9 & 0 & 1 & 7 \\ \hline \end{array} \quad * \quad \begin{array}{|c|c|c|}\hline 1 & 2 & 4 \\ \hline 1 & 0 & 1 \\ \hline 2 & 1 & 2 \\ \hline \end{array} \quad \Rightarrow \quad \begin{array}{|c|c|c|}\hline 2 & 1 & 2 \\ \hline 1 & 0 & 1 \\ \hline 1 & 2 & 4 \\ \hline \end{array}$$

Faltung in der Bildverarbeitung: graphische Veranschaulichung

$$
\begin{array}{|c|c|c|c|}
\hline
10 & 0 & 5 & 1 \\ \hline
1 & 8 & 7 & 2 \\ \hline
5 & 0 & 2 & 6 \\ \hline
9 & 0 & 1 & 7 \\ \hline
\end{array}
\quad * \quad
\begin{array}{|c|c|c|}
\hline
1 & 2 & 4 \\ \hline
1 & 0 & 1 \\ \hline
2 & 1 & 2 \\ \hline
\end{array}
\quad \Rightarrow \quad
\begin{array}{|c|c|c|}
\hline
2 & 1 & 2 \\ \hline
1 & 0 & 1 \\ \hline
1 & 2 & 4 \\ \hline
\end{array}
\quad \Rightarrow \quad
\begin{array}{|c|c|c|}
\hline
2 & 1 & 2 \\ \hline
1 & 0 & 1 \\ \hline
4 & 2 & 1 \\ \hline
\end{array}
$$

Faltung in der Bildverarbeitung: graphische Veranschaulichung

$$\begin{array}{|c|c|c|c|}\hline 10 & 0 & 5 & 1 \\ \hline 1 & 8 & 7 & 2 \\ \hline 5 & 0 & 2 & 6 \\ \hline 9 & 0 & 1 & 7 \\ \hline \end{array} \quad * \quad \begin{array}{|c|c|c|}\hline 1 & 2 & 4 \\ \hline 1 & 0 & 1 \\ \hline 2 & 1 & 2 \\ \hline \end{array} \quad = \quad ?$$

Faltung in der Bildverarbeitung: graphische Veranschaulichung

$$\begin{array}{|c|c|c|c|}\hline 10 & 0 & 5 & 1 \\ \hline 1 & 8 & 7 & 2 \\ \hline 5 & 0 & 2 & 6 \\ \hline 9 & 0 & 1 & 7 \\ \hline \end{array} \quad * \quad \begin{array}{|c|c|c|}\hline 1 & 2 & 4 \\ \hline 1 & 0 & 1 \\ \hline 2 & 1 & 2 \\ \hline \end{array} \quad = \quad \begin{array}{|c|c|c|c|}\hline 10 & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|}\hline 2 & 1 & 2 \\ \hline 1 & 0 & 1 \\ \hline 4 & 2 & 1 \\ \hline \end{array}$$

$$0 \cdot 10 + 1 \cdot 0 + 2 \cdot 1 + 1 \cdot 8 = 10$$

Faltung in der Bildverarbeitung: graphische Veranschaulichung

$$\begin{array}{|c|c|c|c|}\hline \boxed{10} & 0 & 5 & 1 \\ \hline 1 & 8 & 7 & 2 \\ \hline 5 & 0 & 2 & 6 \\ \hline 9 & 0 & 1 & 7 \\ \hline \end{array} \quad * \quad \begin{array}{|c|c|c|}\hline 1 & 2 & 4 \\ \hline 1 & 0 & 1 \\ \hline 2 & 1 & 2 \\ \hline \end{array} \quad = \quad \begin{array}{|c|c|c|c|}\hline 10 & 42 & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|}\hline 2 & 1 & 2 \\ \hline 1 & 0 & 1 \\ \hline 4 & 2 & 1 \\ \hline \end{array}$$

$$1 \cdot 10 + 0 \cdot 0 + 1 \cdot 5 + 4 \cdot 1 + 2 \cdot 8 + 1 \cdot 7 = 42$$

Faltung in der Bildverarbeitung: graphische Veranschaulichung

$$\begin{array}{|c|c|c|c|}\hline 10 & 0 & 5 & 1 \\ \hline 1 & 8 & 7 & 2 \\ \hline 5 & 0 & 2 & 6 \\ \hline 9 & 0 & 1 & 7 \\ \hline \end{array} \quad * \quad \begin{array}{|c|c|c|}\hline 1 & 2 & 4 \\ \hline 1 & 0 & 1 \\ \hline 2 & 1 & 2 \\ \hline \end{array} \quad = \quad \begin{array}{|c|c|c|c|}\hline 10 & 42 & 49 & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|}\hline 2 & 1 & 2 \\ \hline 1 & 0 & 1 \\ \hline 4 & 2 & 1 \\ \hline \end{array}$$

$$1 \cdot 0 + 0 \cdot 5 + 1 \cdot 1 + 4 \cdot 8 + 2 \cdot 7 + 1 \cdot 2 = 49$$

Faltung in der Bildverarbeitung: graphische Veranschaulichung

10	0	5	1
1	8	7	2
5	0	2	6
9	0	1	7

$*$

1	2	4
1	0	1
2	1	2

$=$

10	42	49	37

2	1	2
1	0	1
4	2	1

$$1 \cdot 5 + 0 \cdot 1 + 4 \cdot 7 + 2 \cdot 2 = 37$$

Faltung in der Bildverarbeitung: graphische Veranschaulichung

10	0	5	1
1	8	7	2
5	0	2	6
9	0	1	7

$*$

1	2	4
1	0	1
2	1	2

$=$

10	42	49	37
28			

2	1	2
1	0	1
4	2	1

$$1 \cdot 10 + 2 \cdot 0 + 0 \cdot 1 + 1 \cdot 8 + 2 \cdot 5 + 0 \cdot 1 = 28$$

Faltung in der Bildverarbeitung: graphische Veranschaulichung

$$\begin{array}{|c|c|c|c|}\hline 10 & 0 & 5 & 1 \\ \hline 1 & 8 & 7 & 2 \\ \hline 5 & 0 & 2 & 6 \\ \hline 9 & 0 & 1 & 7 \\ \hline \end{array} \quad * \quad \begin{array}{|c|c|c|}\hline 1 & 2 & 4 \\ \hline 1 & 0 & 1 \\ \hline 2 & 1 & 2 \\ \hline \end{array} \quad = \quad \begin{array}{|c|c|c|c|}\hline 10 & 42 & 49 & 37 \\ \hline 28 & 60 & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

$$\begin{array}{|c|c|c|}\hline 2 & 1 & 2 \\ \hline 1 & 0 & 1 \\ \hline 4 & 2 & 1 \\ \hline \end{array}$$

$$2 \cdot 10 + 1 \cdot 0 + 2 \cdot 5 + 1 \cdot 1 + 0 \cdot 8 + 1 \cdot 7 + 4 \cdot 5 + 2 \cdot 0 + 1 \cdot 2 = 60$$

Faltung in der Bildverarbeitung: graphische Veranschaulichung

$$
\begin{array}{|c|c|c|c|}
\hline
10 & 0 & 5 & 1 \\ \hline
1 & 8 & 7 & 2 \\ \hline
5 & 0 & 2 & 6 \\ \hline
9 & 0 & 1 & 7 \\ \hline
\end{array}
\quad * \quad
\begin{array}{|c|c|c|}
\hline
1 & 2 & 4 \\ \hline
1 & 0 & 1 \\ \hline
2 & 1 & 2 \\ \hline
\end{array}
\quad = \quad
\begin{array}{|c|c|c|c|}
\hline
10 & 42 & 49 & 37 \\ \hline
28 & 60 & 27 & 38 \\ \hline
35 & 68 & 42 & 36 \\ \hline
5 & 24 & 21 & 11 \\ \hline
\end{array}
$$

$$
\begin{array}{|c|c|c|}
\hline
2 & 1 & 2 \\ \hline
1 & 0 & 1 \\ \hline
4 & 2 & 1 \\ \hline
\end{array}
$$

$$2 \cdot 2 + 1 \cdot 6 + 1 \cdot 1 + 0 \cdot 7 = 11$$

Faltungen im maschinellen Lernen

Im maschinellen Lernen sind Faltungen insbesondere im Deep Learning von Bedeutung-

Faltende Neuronale Netze (engl. Convolutional Neural Networks, CNNs) werden vor allem für Lernaufgaben mit Bilddaten eingesetzt. CNNs werden dabei genutzt, um wichtige Merkmale aus den Input-Bildern zu extrahieren.

$\Rightarrow$ Kapitel 7

2.2 - Stochastik

2.2 - Stochastik

2.2.1 - Wahrscheinlichkeiten

2.2 - Stochastik

2.2.2 - Zufallsvariablen und Wahrscheinlichkeitsverteilungen

Zufallsvariablen (1)

Def.: Zufallsvariable $X \mathrel{\widehat{=}}$ Mapping von Ereignissen ξ auf Zahlen $X(\xi)$

$$\Rightarrow \xi \in \Omega \longrightarrow X(\xi) \in \mathbb{R}$$

Darstellung von Zufallsverteilungen

Für einzelne Zufallszahlen: $X \sim \mathcal{N}(\mu, \sigma^2)$

Für Zufallsvektoren: $\underline{X} \sim \mathcal{N}(\underline{\mu}, \mathbf{C})$

σ^2: Varianz, $\mathbf{C}$: Kovarianzmatrix

Zufallsvariablen (2)

Diskrete Zufallsvariablen:

- begrenzte Menge an diskreten möglichen Ereignissen
- exakte Wahrscheinlichkeit bestimmbar
- Summe der Wahrscheinlichkeiten aller möglichen Ereignisse $= 1$

Kontinuierliche Zufallsvariablen:

- Reelle Zahlen $\Rightarrow$ theoretisch ∞ mögliche Ereignisse (beliebig viele Nachkommastellen möglich)
- Berechnung der exakten Wahrscheinlichkeit unmöglich
- Bestimmung der Wahrscheinlichkeit über Fläche unter einer Funktion $f_X(x)$ innerhalb eines Intervalls

Wahrscheinlichkeitsdichtefunktion

Eine funktion f ist die Wahrscheinlichkeitsdichtefunktion (engl. probability density function, pdf) von X wenn gilt:

- $f(x) \geq 0 \qquad \forall x \in \mathbb{R}$
- f ist integrierbar
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$
- $P(X \leq a) = \int_{-\infty}^{a} f(x)dx \qquad \forall a \in \mathbb{R}$

Warum können keine exakten Wahrscheinlichkeiten aus f bestimmt werden?

Die Wahrscheinlichkeit, dass sich ein Wert im Intervall [a,b] befindet, ist
$P([a,b]) = \int_a^b f(x)dx$

Exakter Wert $\Rightarrow a = b \Rightarrow P([a,a]) = \int_a^a f(x)dx = F(a) - F(a) = 0$

Verteilungsfunktion

Die Verteilungsfunktion (engl. cumulative distribution function, cdf) einer Zufallsvariable X ist definiert als:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

Eigenschaften:

- F ist monoton steigend
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$

Bedingte Wahrscheinlichkeitsdichtefunktion/Verteilungsfunktion

Für eine Zufallsvariable X und ein Ereignis A gilt:

- $F_X(x|A) = P(X \leq x|A)$ ist ihre bedingte Verteilungsfunktion
- $f_X(x|A) = \frac{dF_X(x|A)}{dx}$ ist ihre bedingte Wahrscheinlichkeitsdichtefunktion

Bayes-Theorem:

$$F_X(x|A) = P(A|X \leq x)\frac{F_X(x)}{P(A)}$$

$$f_X(x|A) = P(A|X \leq x)\frac{f_X(x)}{P(A)}$$

Totale Wahrscheinlichkeit:

$$F_X(x) = \sum_i F_X(x|A_i)P(A_i)$$

$$f_X(x) = \sum_i f_X(x|A_i)P(A_i)$$

Wichtige Verteilungen

Gaußverteilung/Normalverteilung

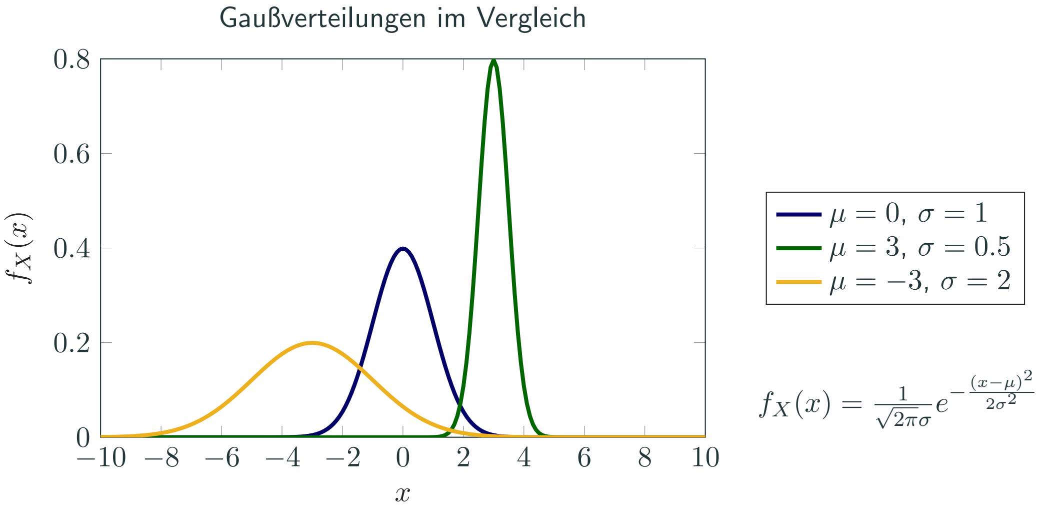
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Standardnormalverteilung

Häufig eingesetzt: Mittelwert 0 und Standardabweichung 1 $\Rightarrow X \sim \mathcal{N}(0, 1)$

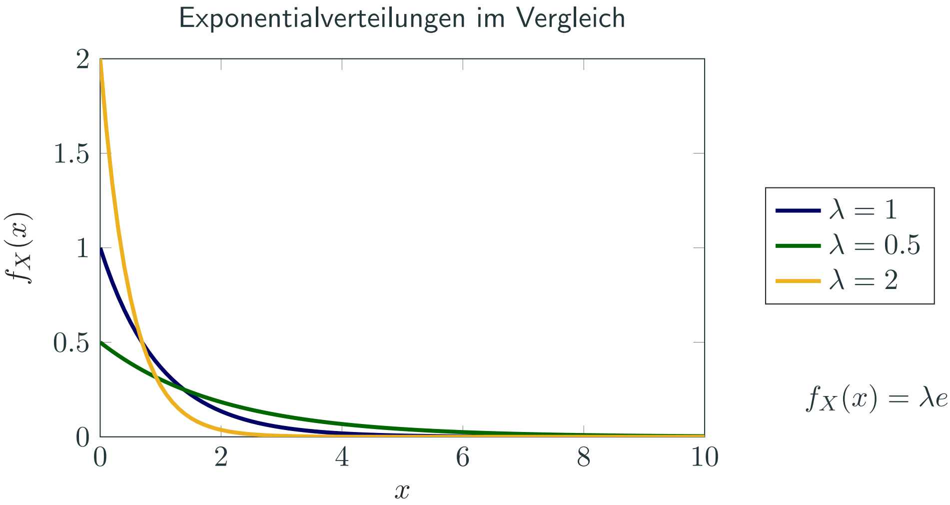
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}x^2}$$

Normalverteilung



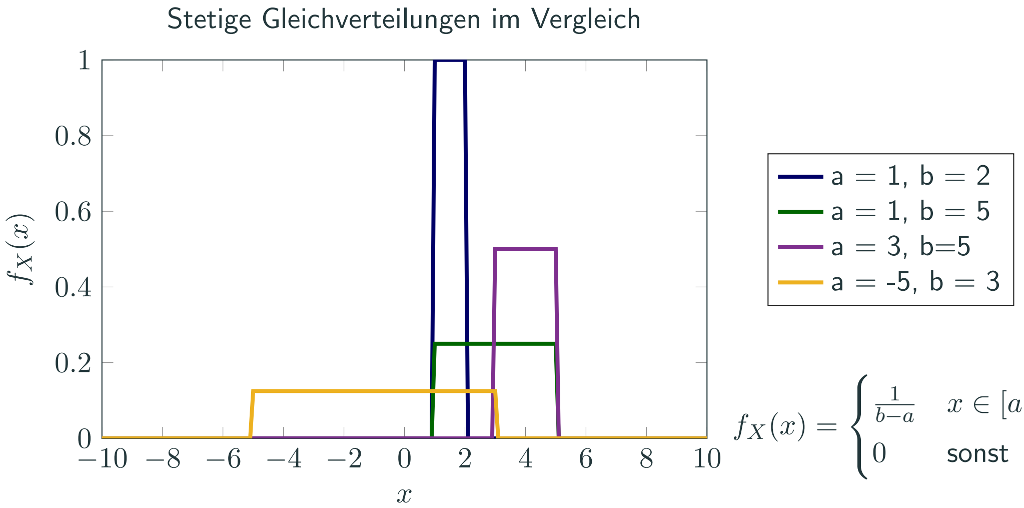
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Exponentialerteilung



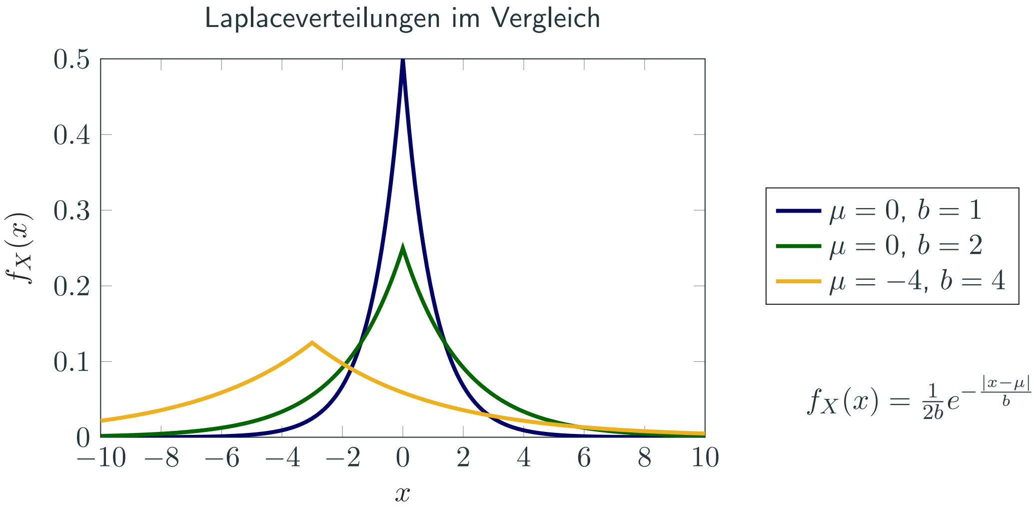
$$f_X(x) = \lambda e^{-\lambda x}$$

Stetige Gleichverteilung



$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{sonst} \end{cases}$$

Laplaceverteilung



$$f_X(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

2.3 - Grundlagen der Optimierung

2.3 - Grundlagen der Optimierung

2.3.1 - Was versteht man unter Optimierung?

Optimierungsprobleme im Alltag

Optimierungsprobleme begnen uns überall:

- Lernen in der Prüfungsphase - wie kann ich mit möglichst wenig Aufwand möglichst gute Noten erzielen?
- Finanzen - wie kann ich meine Ausgaben so optimieren, dass ich möglichst viel für meinen Urlaub sparen kann?
- Obst- und Gemüsegarten - wie nutze ich den Platz im Garten so aus, dass ich möglichst viel Ertrag bekomme?
- Altersvorsorge - wie verteile ich mein Geld optimal, um mit möglichst großer Sicherheit eine möglichst hohe Rente zu kassieren?
- Produktionsplanung - wie können möglichst große Stückzahlen zu möglichst geringen Kosten produziert werden?
- ...

Optimierung im maschinellen Lernen

Optimierung im mathematischen Sinne:

Optimierungsverfahren werden genutzt, um innerhalb eines gegebenen Suchraums $\mathcal{X}$ die Parameter zu finden, die zur bestmöglichen Lösung eines Problems führen.

Realisiert wird dieses Vorhaben durch Minimierung (oder seltener Maximierung) einer Zielfunktion $f(\underline{x})$, in der Regel unter Berücksichtigung einer oder mehrerer Nebenbedingungen $g_i(\underline{x})$.

Optimierung allgemein:

$$\underline{x}_{\text{opt}} = \arg\min f(\underline{x}), \quad \underline{x} \in \mathcal{X}$$
$$\text{s.t.} g_i(\underline{x})$$

2.3 - Grundlagen der Optimierung

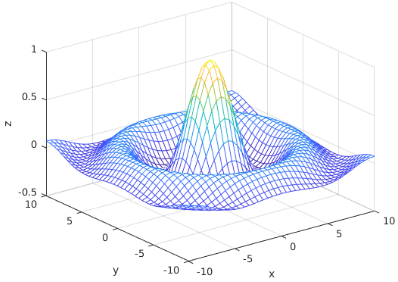
2.3.2 - Beispiele für Optimierungsverfahren

Grid-Search

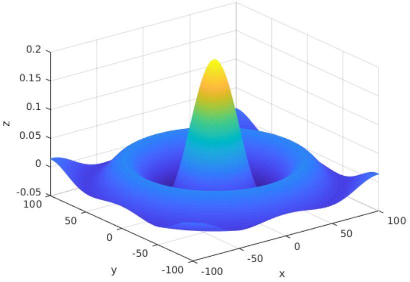
Idee: Systematische Suche über mögliche Parameterkombinationen

Vorteil: optimale Lösung wird auf jeden Fall gefunden

Nachteil: wird schnell sehr Zeit- und Rechenintensiv



1.681 Kombinationsmöglichkeiten



160.801 Kombinationsmöglichkeiten

Lineare Optimierung (engl. “Linear Programming”)

Idee: Minimierung bzw. Maximierung einer linearen Funktion unter bestimmten Nebenbedingungen (durch Lösen eines Ungleichungssystems)

$$\begin{aligned} \max \quad & \underline{c}^{\mathrm{T}} \underline{x} \\ \text{s.t.} \quad & \mathbf{A}\underline{x} \leq \underline{b} \\ \text{and} \quad & \underline{x} \geq 0 \end{aligned}$$

$\underline{c}$ und $\underline{b}$: Vektoren bekannter Koeffizienten, $\mathbf{A}$: Matrix bekannter Koeffizienten, $\underline{x}$ zu bestimmende Parameter

Hinweis

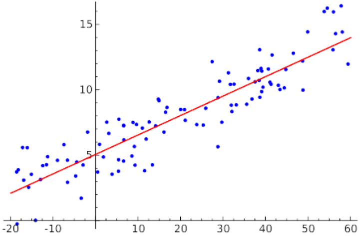
“Programming” oder “Program” im Bezug auf Optimierungsverfahren sind historische Begriffe, die nichts mit Programmierung zu tun haben, wie wir sie heutzutage kennen!

Methode der kleinsten Fehlerquadrate (Least Squares)

Idee: Die Parameter wählen, für die der quadratische Fehler am geringsten ausfällt

Am Beispiel Regression:

Gegeben seien N Werte für x_i sowie die dazugehörigen Messwerte y_i. Ziel: Fitting einer optimalen Funktion durch die Messwerte Minimierung einer Kostenfunktion L in Abhängigkeit der Parameter $\underline{\theta}$.



$$\underline{\theta_{\text{opt}}} = \arg\min \underline{\theta} L(\underline{\theta}) = \sum_{i=1}^{N} (y_i - f(\underline{x}_i, \underline{\theta}))^2 \qquad (y_i - f(\underline{x}_i, \underline{\theta}))^2 : l_2\text{-Loss}$$

Least-Squares-Probleme können sowohl lineare als auch nichtlinear gelöst werden.

Gradient Descent

Idee: Iterative Annäherung an Minimum durch Berechnung des Gradienten von $f(x)$.

Eine bekannte Veranschaulichung: Bergwanderer im Nebel

Vorgehen:

1. Starte an zufälligem Punkt auf der Funktion (Startpunkt des Wanderers)
2. Definiere die Änderung, die in jeder Iteration erfolgen soll (Wie groß sind die Schritte, die der Wanderer geht?)
3. Berechne den Gradienten in jede mögliche Richtung und wähle den mit der größten negativen Änderung (Wo geht es am steilsten bergab?)
4. Gehe einen Schritt in die gewählte Richtung und berechne erneut alle Gradienten
5. Wiederhole den Vorgang so lange, bis nur noch positive Gradienten existieren (Wanderer ist im Tal angekommen)

Gradient Descent ist heute eines der wichtigsten ML-Optimierungsverfahren

Beispiele für weitere Methoden

Kreuzentropie-Methode: Ein Monte-Carlo-Verfahren, bei dem Beispiele aus einer Wahrscheinlichkeitsverteilung gezogen werden und versucht wird, die Kreuzentropie zwischen dieser Verteilung und der Zielverteilung zu minimieren, um in der nächsten Iteration Beispiele aus einer besser passenden Verteilung ziehen zu können.

Schwarmintelligenz-Methoden: Orientieren sich an Modellen, die das Verhalten realer Schwärme (z.B. Bienen, Ameisen) nachbilden. Beispiel Ameisenalgorithmen: Suche nach der optimalen, kürzesten Route zum Futter $\Rightarrow$ Minimierungsproblem, das durch Zusammenspiel vieler einzelner Ameisen gelöst wird (Agentenbasiertes Modell)

Genetische Algorithmen: Basieren auf Evolutionsmodellen, d.h. das Ziel einer Optimierung besteht darin, diejenigen Paramter auszuwählen, die in der nächsten Iteration die bestmöglichen "Nachkommen" zeugen