

# Machine Learning 1

## Ergänzungsfach Informatik

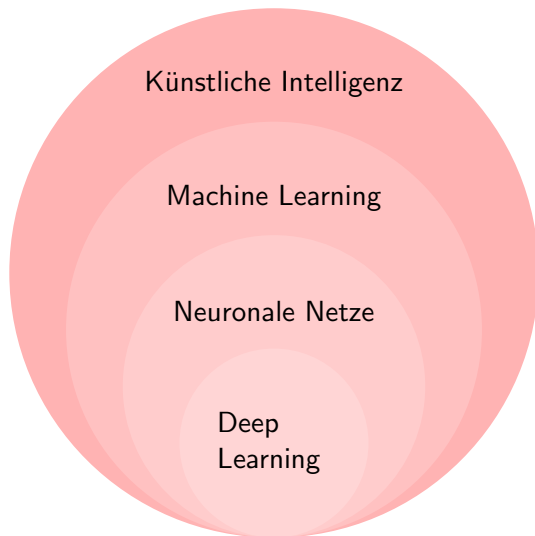
Oliver Probst (pro@kwse.ch)

KSWE

10. November 2021

KANTONSSCHULE  
WETTINGEN  
GYMNASIUM  
FACHMITTELSCHULE

# Künstliche Intelligenz vs. Machine Learning



# Machine Learning

## Definition

Ein Computerprogramm eignet sich Wissen aus vorhandenen Daten an (“Lernen”). Dabei wird versucht ein Muster in den Daten zu erkennen (eng. Pattern Recognition). Nach dem Lernen soll das Computerprogramm für neue Daten eine Entscheidung treffen.

# Machine Learning Paradigmen

- Supervised Learning
  - ▶ Regression
  - ▶ Classification
- Unsupervised Learning
  - ▶ Clustering
  - ▶ Generative Modelle

# Supervised Learning: Regression

- **Ziel:** Eine (numerische) Vorhersage für neue Daten.
- **gegebene Daten:** Für jeden Datensatz (der aus mehreren Merkmalen bestehen kann) ist das wahre Ergebnis bekannt.
- Beispiele:
  - ▶ Sage für eine Flugroute die Verspätung in Minuten vorher.
  - ▶ Sage für eine Immobilie den Verkaufspreis vorher.
  - ▶ Bestimme für einen Kunden und eine Werbeanzeige die Wahrscheinlichkeit, mit der ein Kunde die Werbeanzeige anklickt.

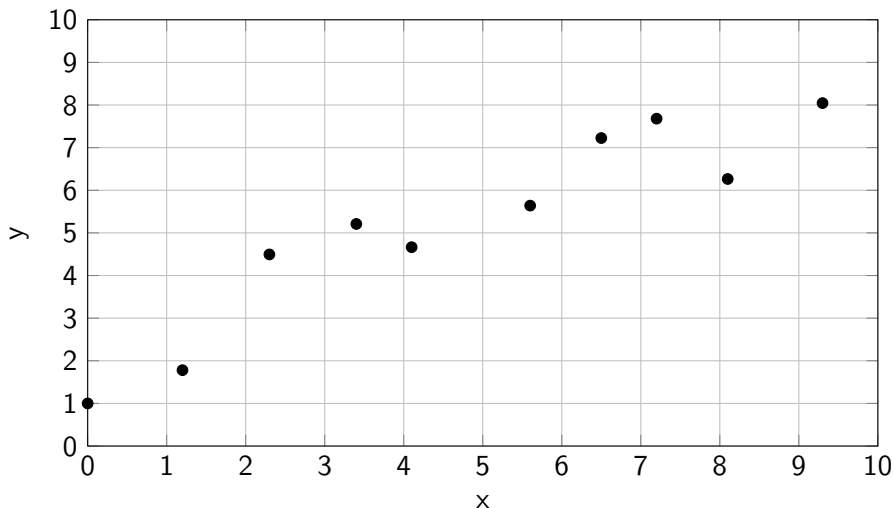
# Beispiel: Diabetes Daten-Set

Es gibt  $k = 442$  Datensätze. Jeder Datensatz besteht aus 11 Messwerten.

- Features
  - ▶ Alter
  - ▶ Geschlecht
  - ▶ BMI
  - ▶ Blutdruck (Durchschnitt)
  - ▶ Sechs Blutserummesswerte (z.B. Cholesterin, Blutzuckerspiegel, ...)
- Label (Target)
  - ▶ Quantitative Messung über das Krankheitsfortschreiten (“eine Zahl”) im Vergleich zu einer “Baseline” (Messung nach einem 1 Jahr).
- Ziel: für einen neuen Datensatz soll das Krankheitsfortschreiten ermittelt werden

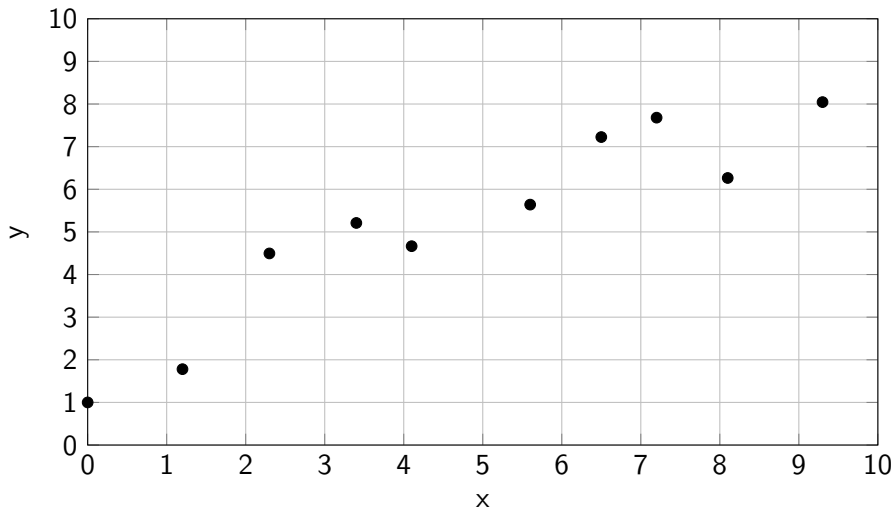
# Regression (abstrakt)

Finde eine Funktion, welche für neue Daten eine möglichst exakte Vorhersage macht unter der Berücksichtigung der gegebenen Daten.



# Walkthrough: **Lineare** Regression

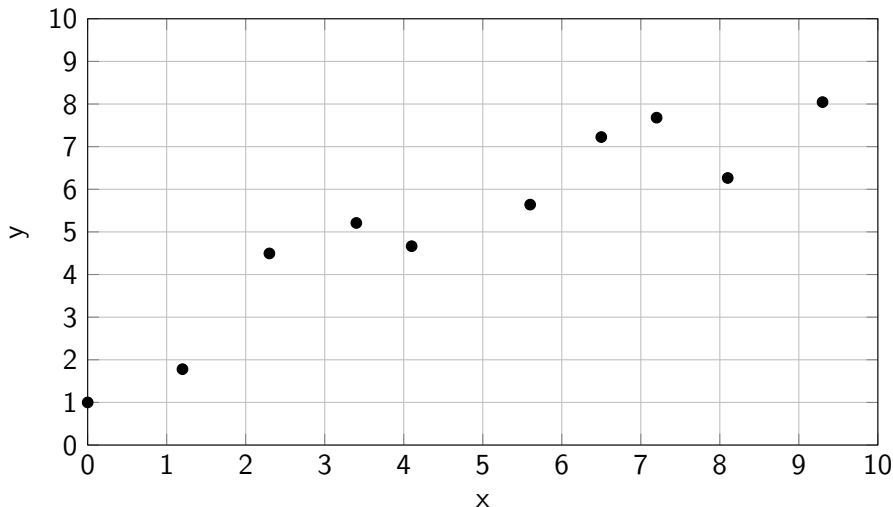
Finde eine **lineare** Funktion, welche für neue Daten eine möglichst exakte Vorhersage macht unter der Berücksichtigung der gegebenen Daten.





# Walkthrough: **Lineare** Regression

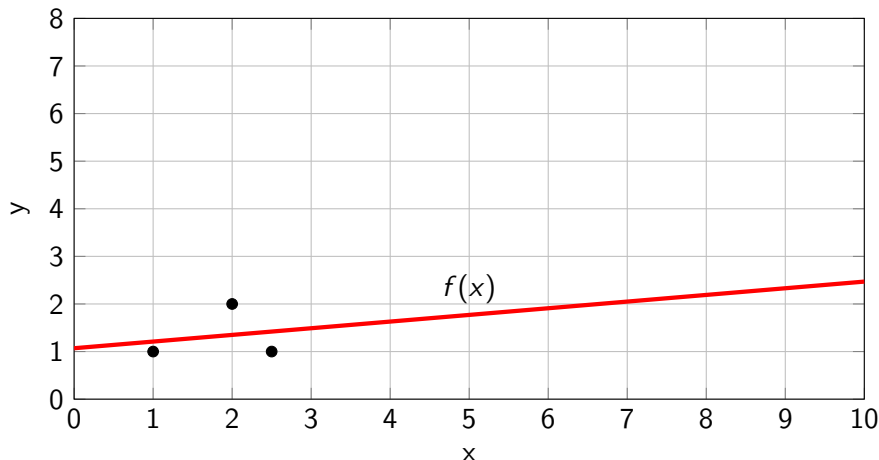
Finde eine **lineare** Funktion, welche für neue Daten eine möglichst exakte Vorhersage macht unter der Berücksichtigung der gegebenen Daten.



# Walkthrough: **Lineare** Regression (1-D)

Gegeben:  $D = \{(1/1), (2/2), (2,5/1)\}$

Methode der kleinsten Quadrate ergibt:  $f(x) = 0,14 \cdot x_1 + 1,07$



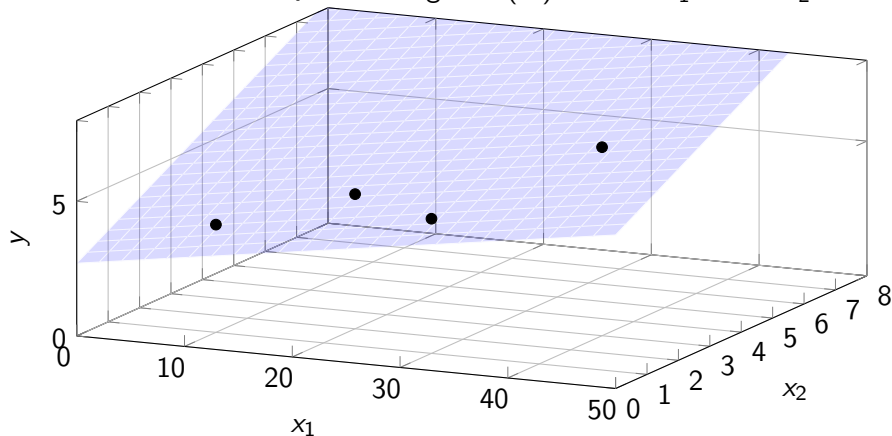
Vorhersage:  $x = 8 \Rightarrow f(8) = 2,19$

## Walkthrough: **Lineare** Regression (2-D)

$$D = \{(\vec{x}_1 = \begin{bmatrix} 10 \\ 1 \end{bmatrix}, 4), (\vec{x}_2 = \begin{bmatrix} 20 \\ 2 \end{bmatrix}, 5), (\vec{x}_3 = \begin{bmatrix} 30 \\ 1 \end{bmatrix}, 5), (\vec{x}_4 = \begin{bmatrix} 40 \\ 3 \end{bmatrix}, 7)\}$$

Gesucht:  $f(\vec{x}) = w_1 \cdot x_1 + w_2 \cdot x_2 + w_0$

Methode der kleinsten Quadrate ergibt:  $f(\vec{x}) = 0.06 \cdot x_1 + 0.7 \cdot x_2 + 2.7$



# Walkthrough: **Lineare** Regression (10-D)

Diabetes Daten-Set:  $D = \{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_{442}, y_{442})\}$

$$\vec{x}_1 = \begin{bmatrix} 59 \\ 2 \\ 32.1 \\ 101 \\ 157 \\ 93.2 \\ 38 \\ 4 \\ 4.8598 \\ 87 \end{bmatrix} \quad \vec{x}_2 = \begin{bmatrix} 48 \\ 1 \\ 21.6 \\ 87 \\ 183 \\ 103.2 \\ 70 \\ 3 \\ 3.8918 \\ 69 \end{bmatrix} \quad \dots \quad \vec{x}_{442} = \begin{bmatrix} 36 \\ 1 \\ 19.6 \\ 71 \\ 250 \\ 133.2 \\ 97 \\ 3 \\ 4.5951 \\ 92 \end{bmatrix}$$

$$\vec{y} = X \cdot \vec{w} \Rightarrow \vec{w} = (X^T \cdot X)^{-1} \cdot X^T \cdot \vec{y}$$

$$\vec{y} \in \mathbb{R}^{442}, X \in \mathbb{R}^{442 \times 11} \text{ und } \vec{w} \in \mathbb{R}^{11}$$

# Walkthrough: **Lineare** Regression (10-D) Lösung

$$f(\vec{x}) = w_1 \cdot x_1 + w_2 \cdot x_2 + \cdots + w_{10} \cdot x_{10} + w_0$$

$$\vec{w} = \begin{bmatrix} -0.0364 \\ -22.8596 \\ 5.603 \\ 1.1168 \\ -1.09 \\ 0.7465 \\ 0.372 \\ 6.5338 \\ 68.4831 \\ 0.2801 \\ -334.5671 \end{bmatrix}$$

Vorhersage für  $\vec{x}_{neu}$ :  $y = \vec{x}_{neu} \cdot \vec{w}$  (Skalarprodukt)