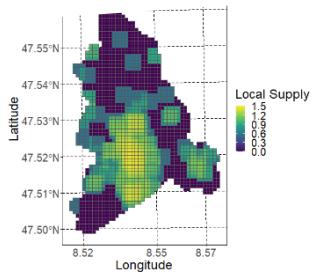
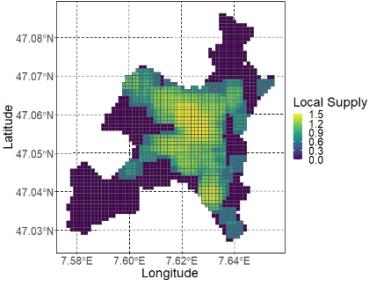


Einführung zu Deep Learning

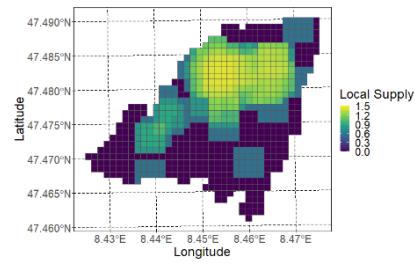
Dr. Yves Staudt



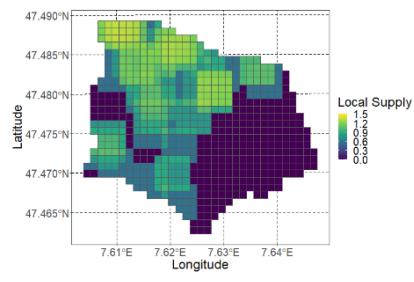
(a)
Grenchen



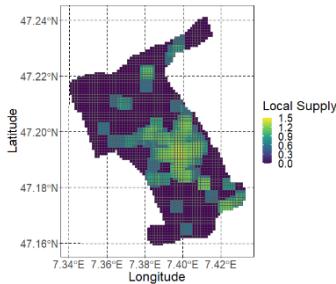
(b)
Oberglatt



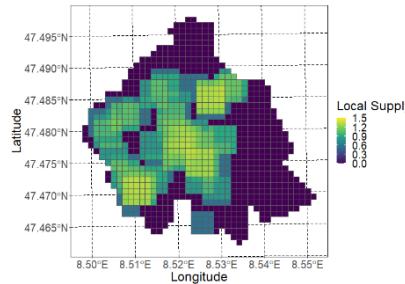
(c)
Reinach



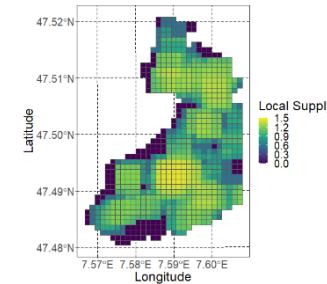
(d)
Rubigen



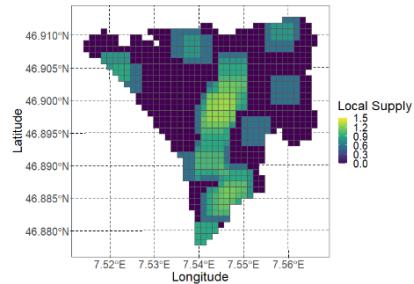
(e)
Buelach



(f)
Burgdorf



(g)
Dielsdorf



(h)
Dornach

Lernziel

Die Studierende sind in der Lage

- ein Deep Learning Modell zu beschreiben
- die Optimierung eines Deep Learning Modell zu beschreiben

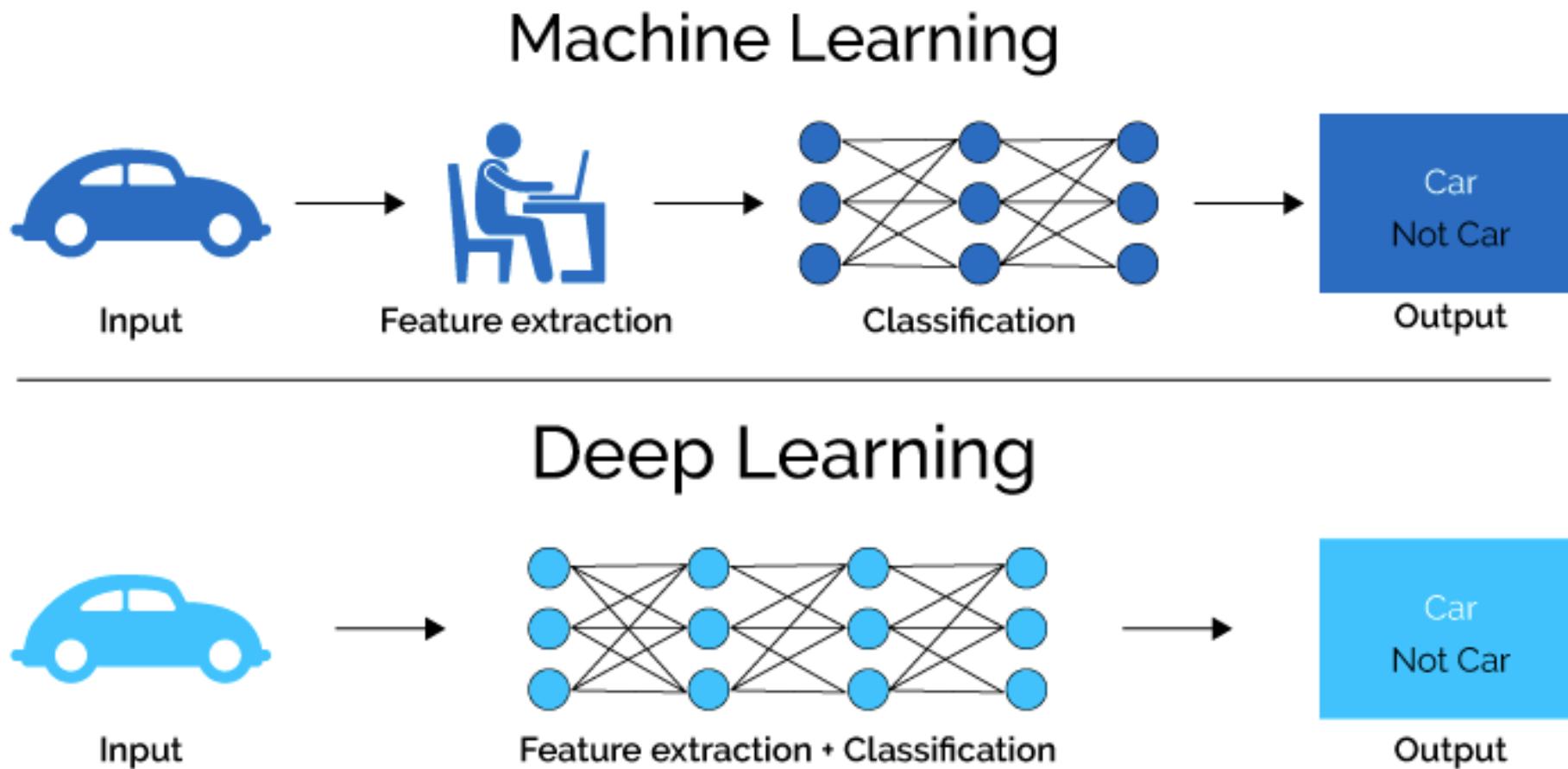
Wiederholung - Aufgabe

Begeben Sie sich in Gruppen von zwei und besprechen Sie sich über folgende Themen

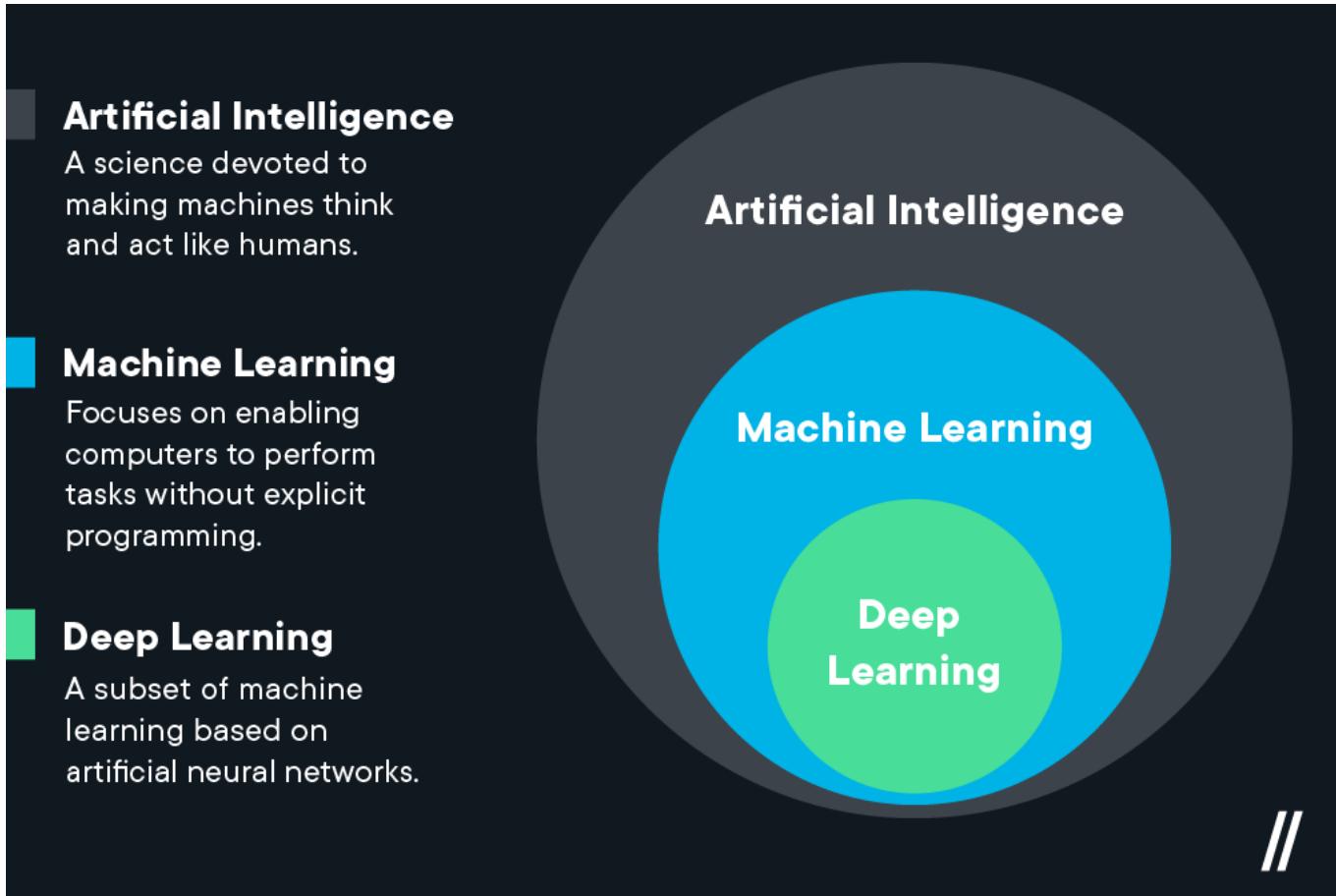
- Worin unterscheidet sich “supervised Learning” von “unsupervised Learning”?
- Worin unterscheidet sich eine Regression von einer Klassifikation?
- Worin unterscheidet sich “Overfitting” von ”Underfitting”?
- Wieso benötige ich ein Training und ein Test Sample?

Sie haben 10 Minuten Zeit sich zu besprechen und Ihre Antworten vorzubereiten.

Machine Learning vs Deep Learning



Artificial Intelligence – Deep Learning

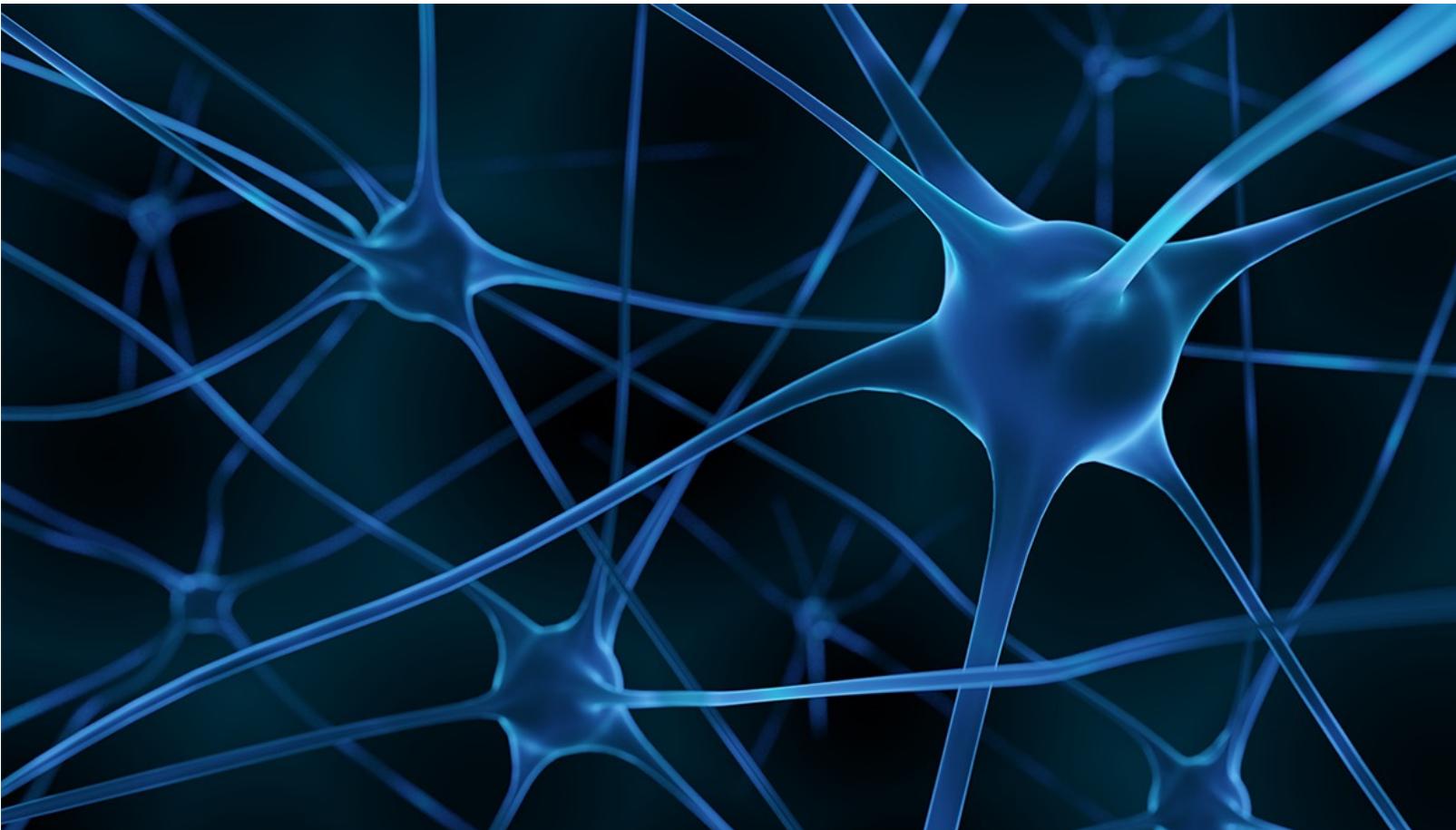


//

Einführung - Künstliche Intelligenz

- Anfangszeit: Lösen von schwierigen Problem der Menschen durch Computer
- Generell: Lösen von Problemen, welche einfach für den Mensch sind, jedoch schwierig, um formal zu beschreiben (Spracherkennung, Bilderkennung)
- Grundidee: Lernen von Erfahrungen basierend Konzepten der realen Welt

Neuronale Netze

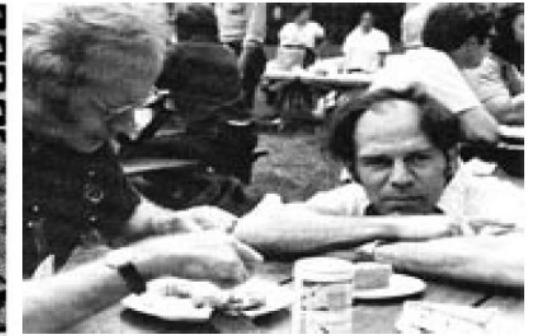
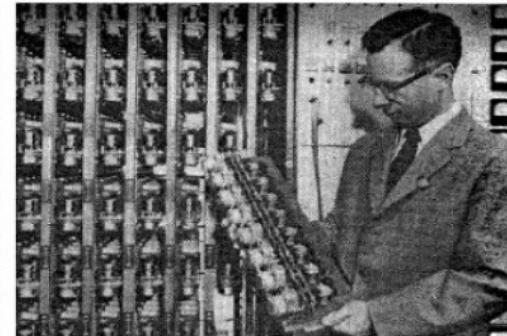
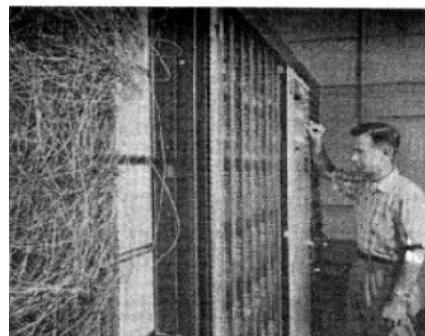
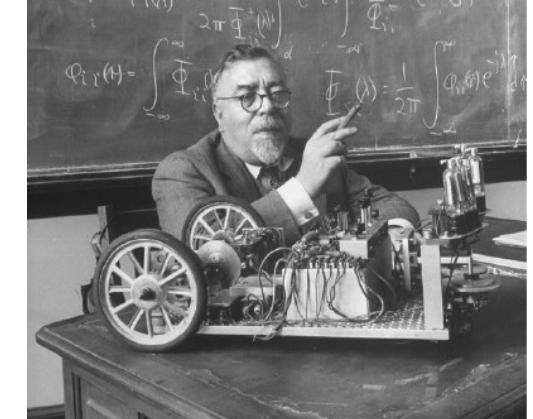
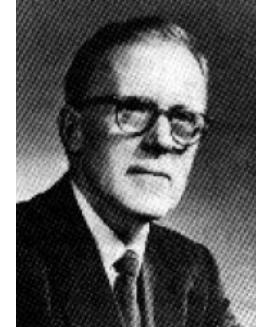


Neuronales Netz im Gehirn

Geschichte

Inspiration für das Forschungsfeld des Deep Learnings: Das Gehirn!

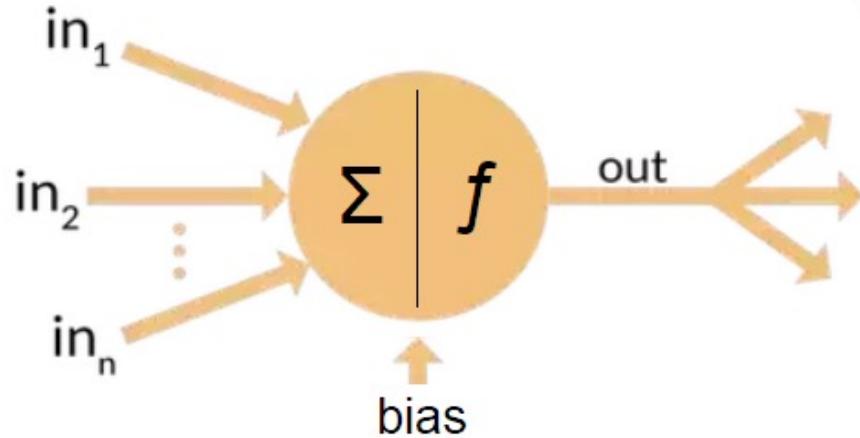
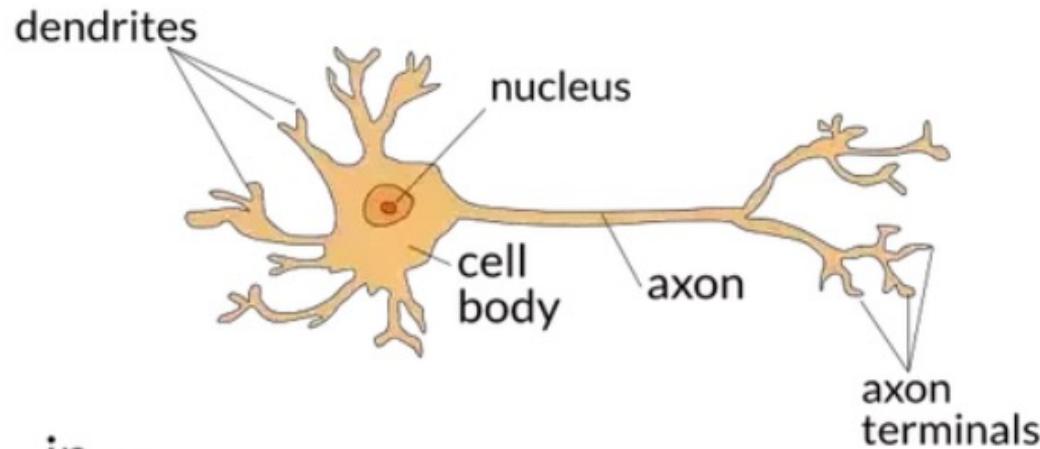
- McCulloch & Pitts (1943)
 - Network of binary neurons can do logic
- Donald Hebb (1947)
 - Hebbian synaptic plasticity
- Norbert Wiener (1948)
 - Cybernetics, optimal filter, feedback, autopoïesis, auto-organization
- Frank Rosenblatt (1957)
 - Perceptron
- Hubel & Wiesel (1960s)
 - Visual cortex architecture



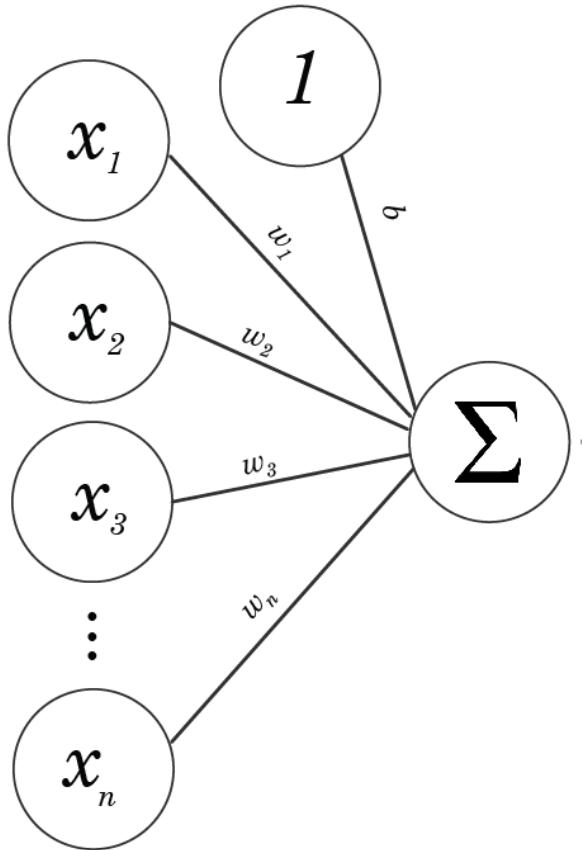
Geschichte

- Zwischen 1968 und 1984 starb das Forschungsfeld um die neuronalen Netze fast vollständig aus
- Isolierte Forschergruppe in Japan fuhr weiter (mit eigener Fundingstrategie)
- 1985 startete die Forschung wieder mit Back Propagation
 - Algorithmus zum Training von Multi-Layered Neural Nets
- Zwischen 1995 und 2010 starb die Forschung und Neural Nets wieder
 - Forscher im Machine Learning nutzen die Neural Nets nicht mehr
- 2010 fand man einen Ansatz, um Multi-Layer Neural Nets mit Bach Propagation für schnellere Speech Recognition zu nutzen

Verbindung zum Gehirn



Perceptron



Darstellung eines Perceptron

Visualisation of the idea of the linear regression (Gonick and Smith, 2005)

CAN WE PREDICT A STUDENT'S WEIGHT y FROM HIS OR HER HEIGHT x ?

Regression analysis

FITS A STRAIGHT LINE TO THIS MESSY SCATTERPLOT.
 x IS CALLED THE INDEPENDENT OR PREDICTOR VARIABLE, AND y IS THE DEPENDENT OR RESPONSE VARIABLE. THE REGRESSION OR PREDICTION LINE HAS THE FORM

$$y = a + bx$$



Linear regression

Linear regression: assumes that there is a linear relationship between X and Y

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

where

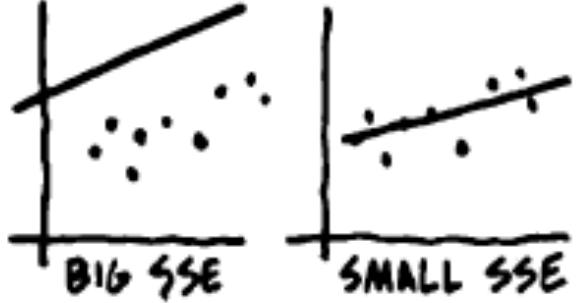
- β_0 is the intercept
- β_1 is the slope

β_0 and β_1 are the unknown model coefficients

β_0 and β_1 are estimated by using data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

Goal: To find an intercept $\hat{\beta}_0$ and a slope $\hat{\beta}_1$ such that the resulting line is as close as possible to the data point
Unbiased Estimator

(James et al., 2013; Kuhn and Johnson, 2013; Hastie et al. 2017)



The regression or least squares line

IS THE LINE WITH THE SMALLEST SSE.



Visualization of the multiple linear regression (Gonick and Smith, 2005)

Multiple linear regression

WE CAN USE THE SAME BASIC IDEAS TO ANALYZE RELATIONSHIPS BETWEEN A DEPENDENT VARIABLE AND SEVERAL INDEPENDENT VARIABLES:

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

FOR EXAMPLE, WEIGHT IS DETERMINED BY A NUMBER OF FACTORS OTHER THAN HEIGHT: AGE, SEX, DIET, BODY TYPE, ETC.

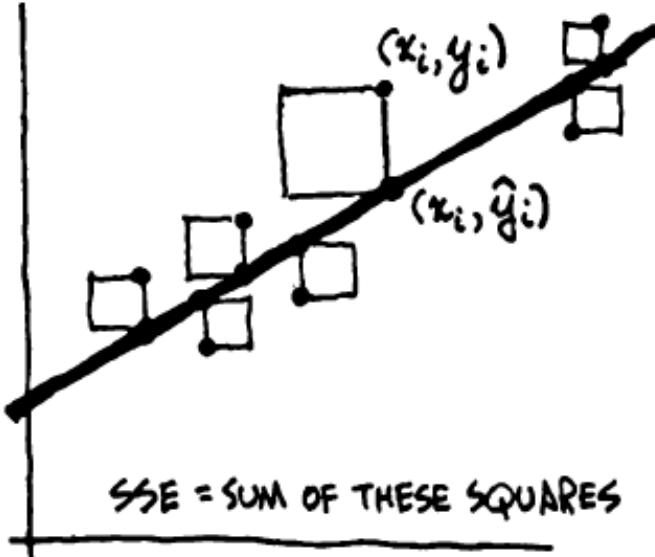
MATRIX ALGEBRA AND A COMPUTER COMBINE TO MAKE SUCH PROBLEMS EASY TO ANALYZE.



Visualisation of the least squares error (Gonick and Smith, 2005)

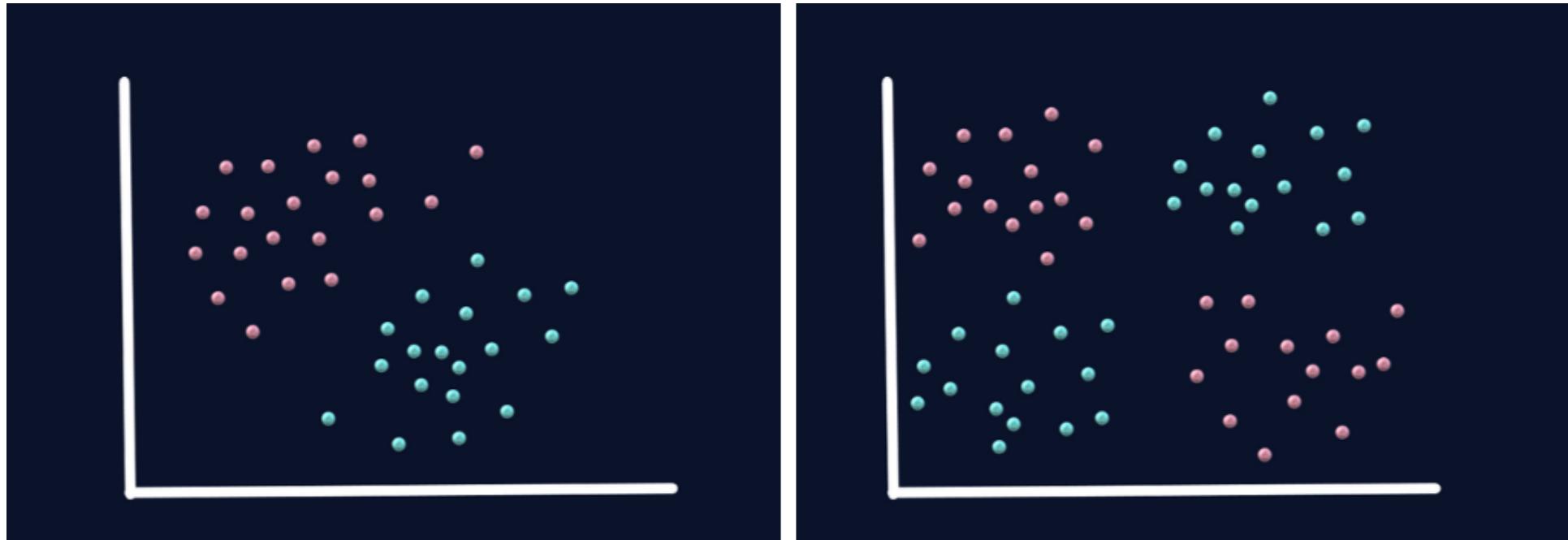
THE IDEA IS TO MINIMIZE THE TOTAL SPREAD OF THE y VALUES FROM THE LINE. JUST AS WHEN WE DEFINED THE VARIANCE, WE LOOK AT ALL THE SQUARED y DISTANCES FROM THE LINE, AND ADD THEM UP TO GET THE SUM OF SQUARED ERRORS:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



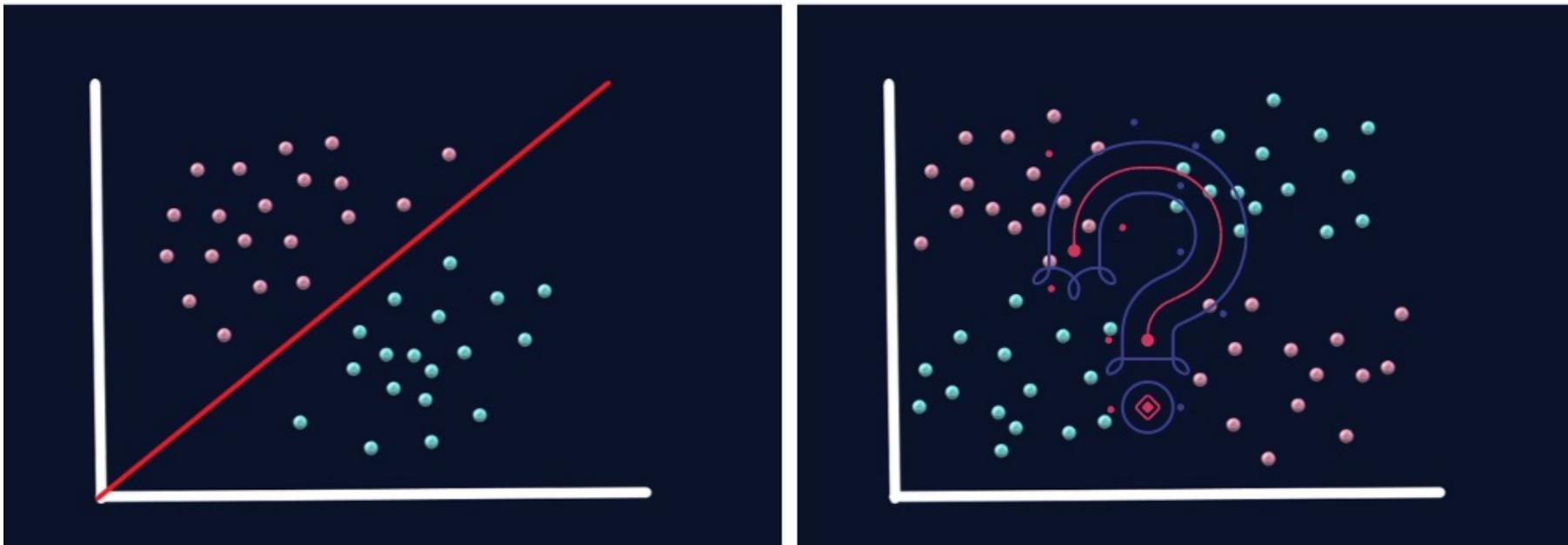
IT'S AN AGGREGATE MEASURE OF HOW MUCH THE LINE'S "PREDICTED y_i ," OR \hat{y}_i , DIFFER FROM THE ACTUAL DATA VALUES y_i .

XOR Challenge



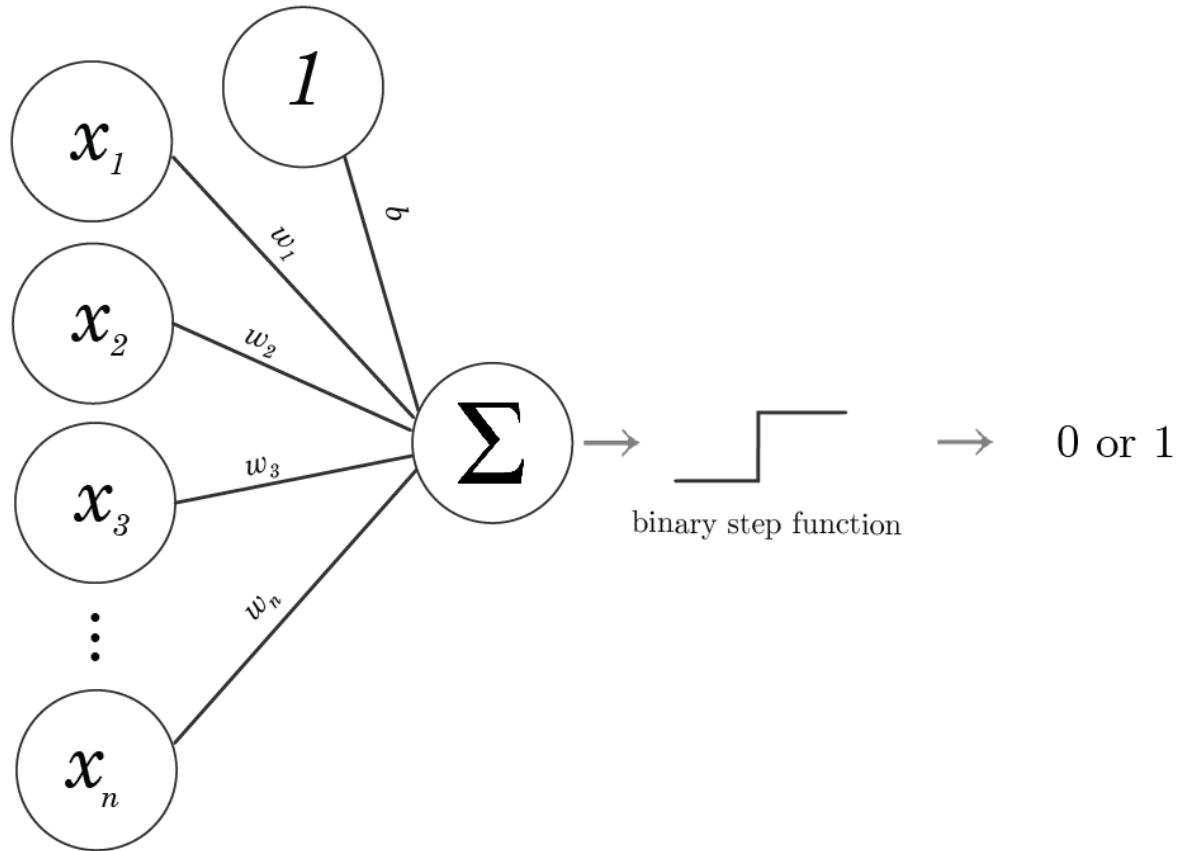
XOR Challenge

XOR Challenge



XOR Challenge

Perceptron Erweiterung



Perceptron mit Aktivierungsfunktion

Why not Linear Regression

- If we use linear regression, some of our estimates might be outside the $[0,1]$ interval, making them hard to interpret as probability.
- Modeling the probability in a linear regression will lead to probabilities out of the range

$$p(X) = \beta_0 + \beta_1 X$$

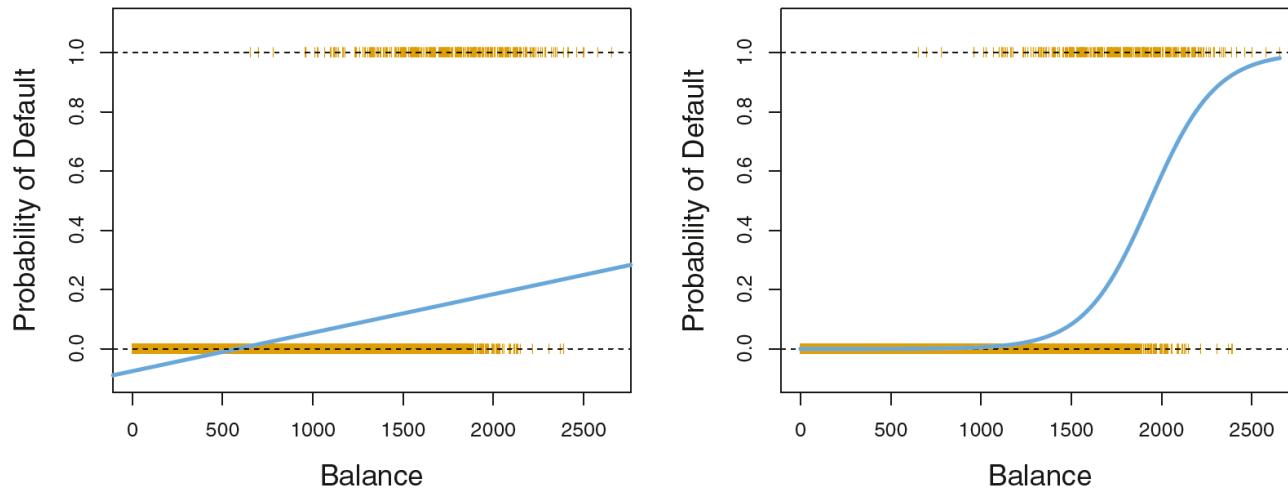
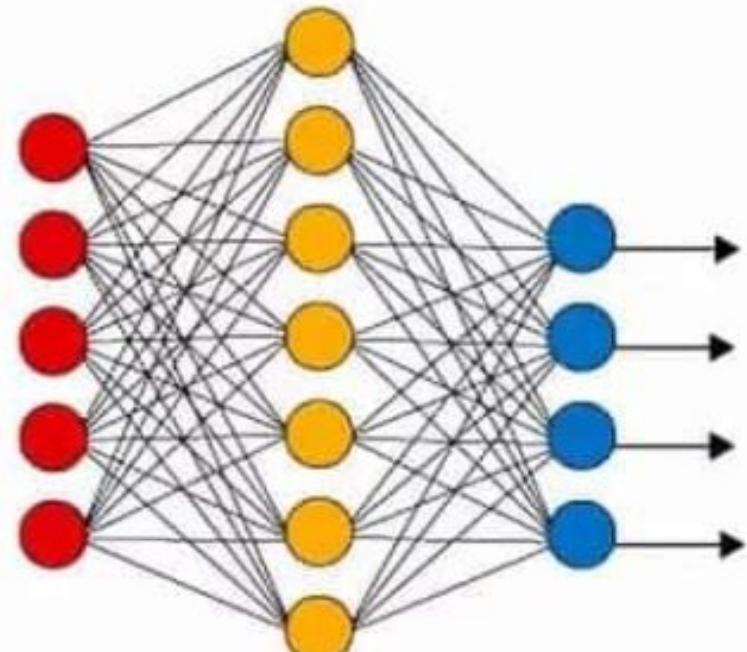


FIGURE 4.2. Classification using the `Default` data. Left: Estimated probability of `default` using linear regression. Some estimated probabilities are negative! The orange ticks indicate the 0/1 values coded for `default`(No or Yes). Right: Predicted probabilities of `default` using logistic regression. All probabilities lie between 0 and 1.

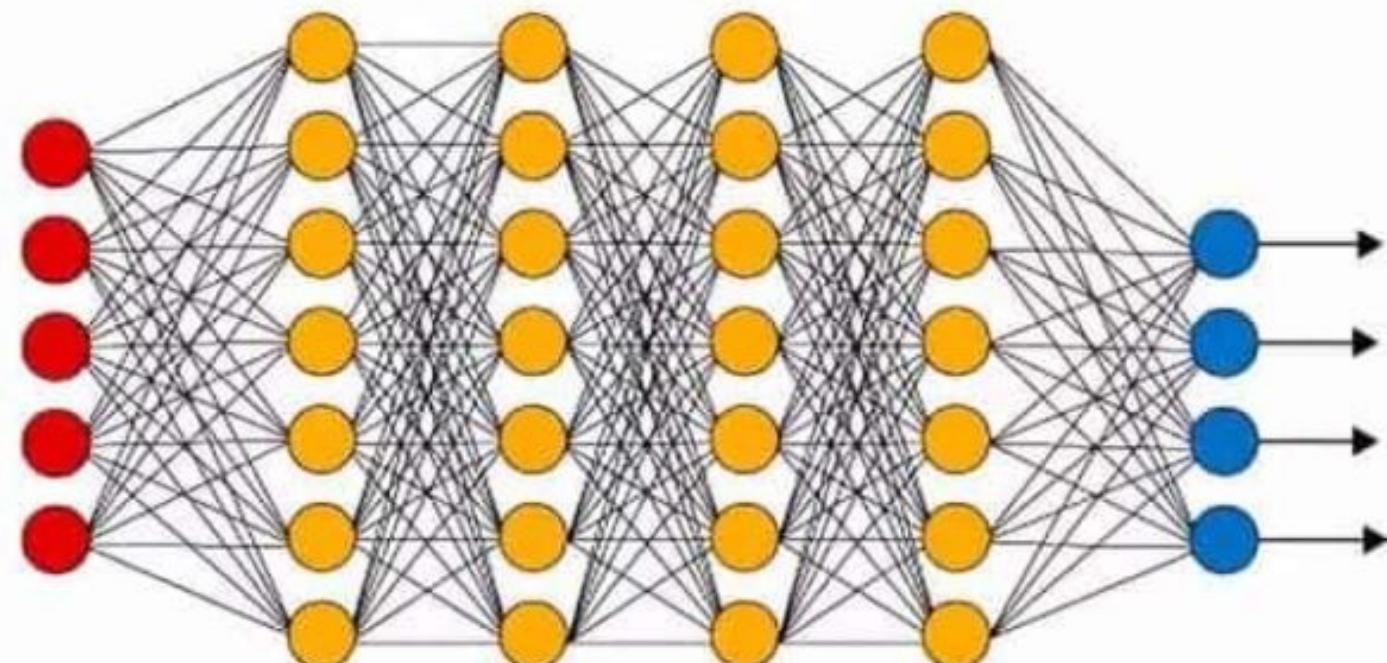
Deep Learning

Simple Neural Network



● Input Layer

Deep Learning Neural Network



● Hidden Layer

● Output Layer

Deep Learning

MNIST Datenbank

<https://de.wikipedia.org/wiki/MNIST-Datenbank>

MNIST-Datenbank

Die **MNIST-Datenbank** (*Modified National Institute of Standards and Technology database*^[1]) ist eine öffentlich verfügbare Datenbank von handgeschriebenen Ziffern. Wobei jede Ziffer als 28×28 Pixel großes Graustufen-Bild gespeichert ist^[1]. Die MNIST-Datenbank besteht aus 60.000 Beispielen im Trainingsdatensatz und 10.000 Beispielen im Testdatensatz. Der Datensatz dient dem Trainieren von Klassifikatoren, unter anderem **Convolutional Neural Networks**, im weiteren Sinne dem **Maschinellen Lernen** bzw. der **Künstlichen Intelligenz**. Eine der MNIST-Datenbank nachempfundene Bilddatenbank ist die *Fashion-MNIST*^[2] von **Zalando**, ebenfalls bestehend aus 60.000 Beispielen im Trainingsdatensatz und 10.000 Beispielen im Testdatensatz. François Chollet, einer der **Keras**-Hauptentwickler sieht MNIST nicht mehr als relevanten Benchmark^[3] an, allerdings wird die MNIST-Datenbank gerne als Einsteigerbeispiel für maschinelles Lernen^{[4][5]} verwendet.

Einelnachweise [Bearbeiten | Quelltext bearbeiten]

1. ↑ ^{a b} *THE MNIST DATABASE of handwritten digits.* ↗ Yann LeCun, Courant Institute, NYU Corinna Cortes, Google Labs, New York Christopher J.C. Burges, Microsoft Research, Redmond, abgerufen am 22. Juli 2020 (englisch).
2. ↑ *Fashion-MNIST.* ↗ zalando, abgerufen am 23. Juli 2020 (englisch).
3. ↑ *Lots of issues with MNIST, but most of all, it is really not representative of CV tasks. Please at least try CIFAR10, of comparable size.* ↗ François Chollet (Twitter tweet), abgerufen am 23. Juli 2020 (englisch).
4. ↑ *Intro to TensorFlow for Deep Learning.* ↗ Abgerufen am 23. Juli 2020 (englisch).
5. ↑ *Tutorial.* ↗ Abgerufen am 23. Juli 2020 (englisch).



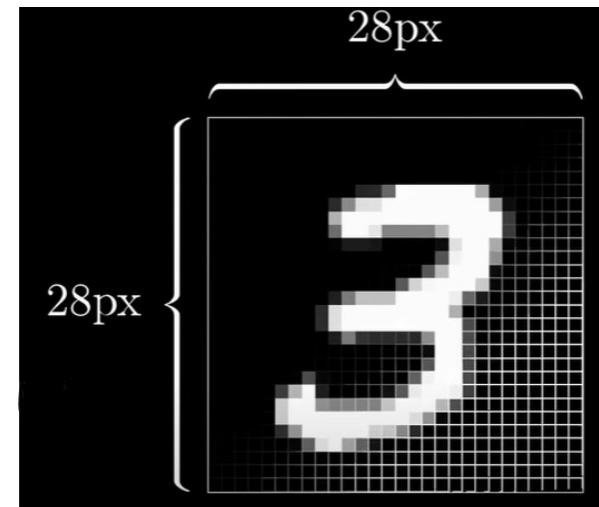
Beispiele aus dem MNIST-Testdatensatz

Kategorien: Maschinelles Lernen | Künstliche Intelligenz | Neuroinformatik | Künstliche Neuronale Netze

Zahlen erkennen

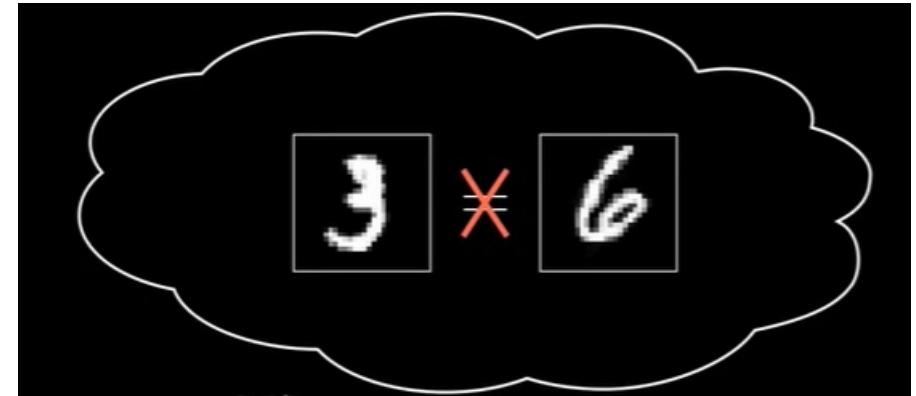
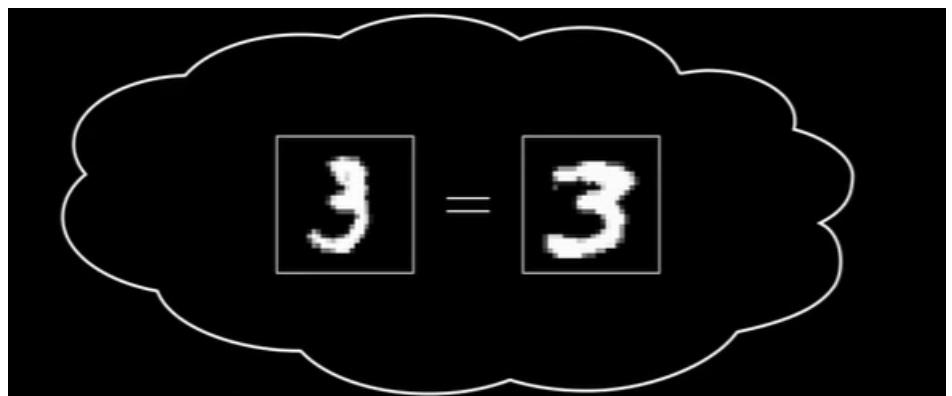
Beispiel: Handgeschriebene Zahl/Ziffer

- 28 x 28 Pixels Auflösung
- Jedes neue Aufschreiben erzeugt mehr oder weniger anderes Bild
- Für den menschlichen Betrachter leicht als Ziffer «3» zu interpretieren, für den Computer sehr schwierig
- Warum kann das Gehirn die Ziffer «3» so leicht erkennen?

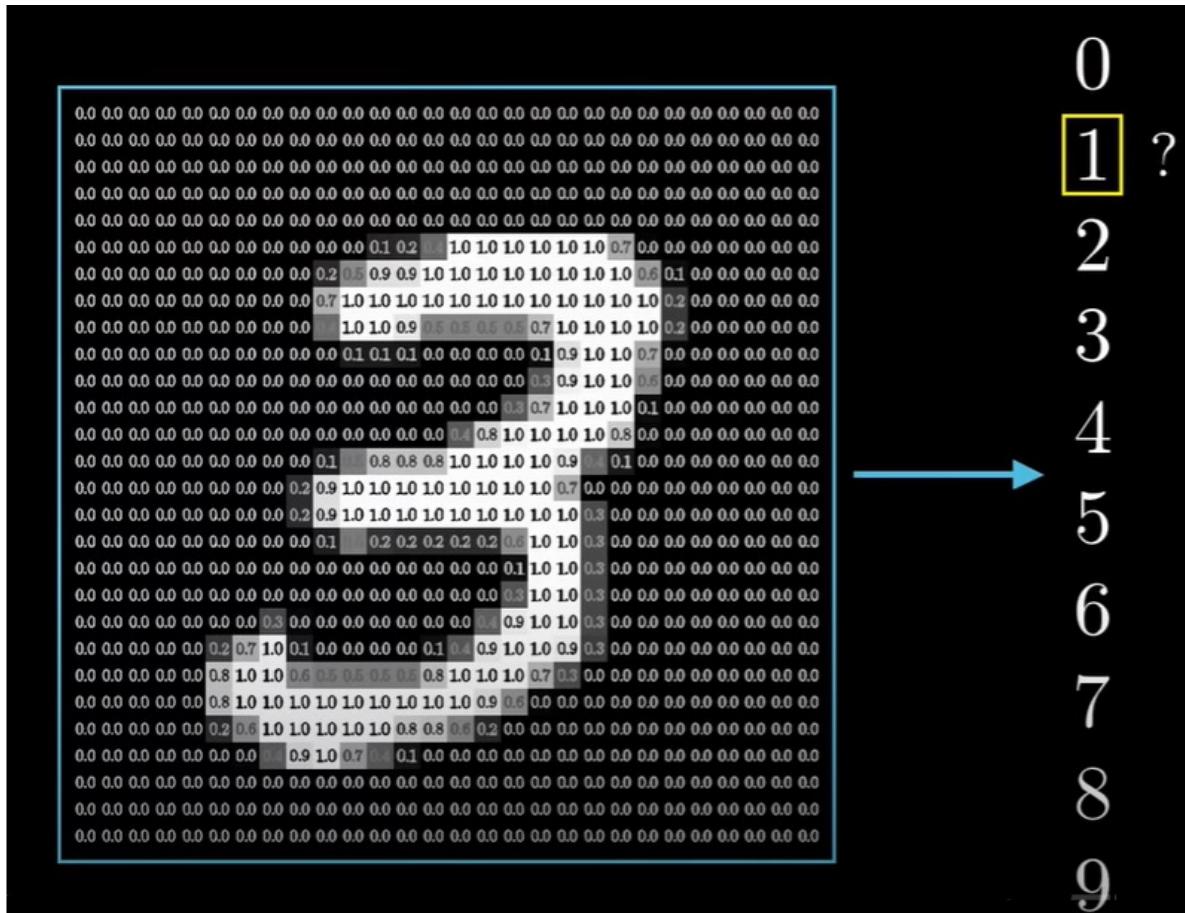


Zahl erkennen

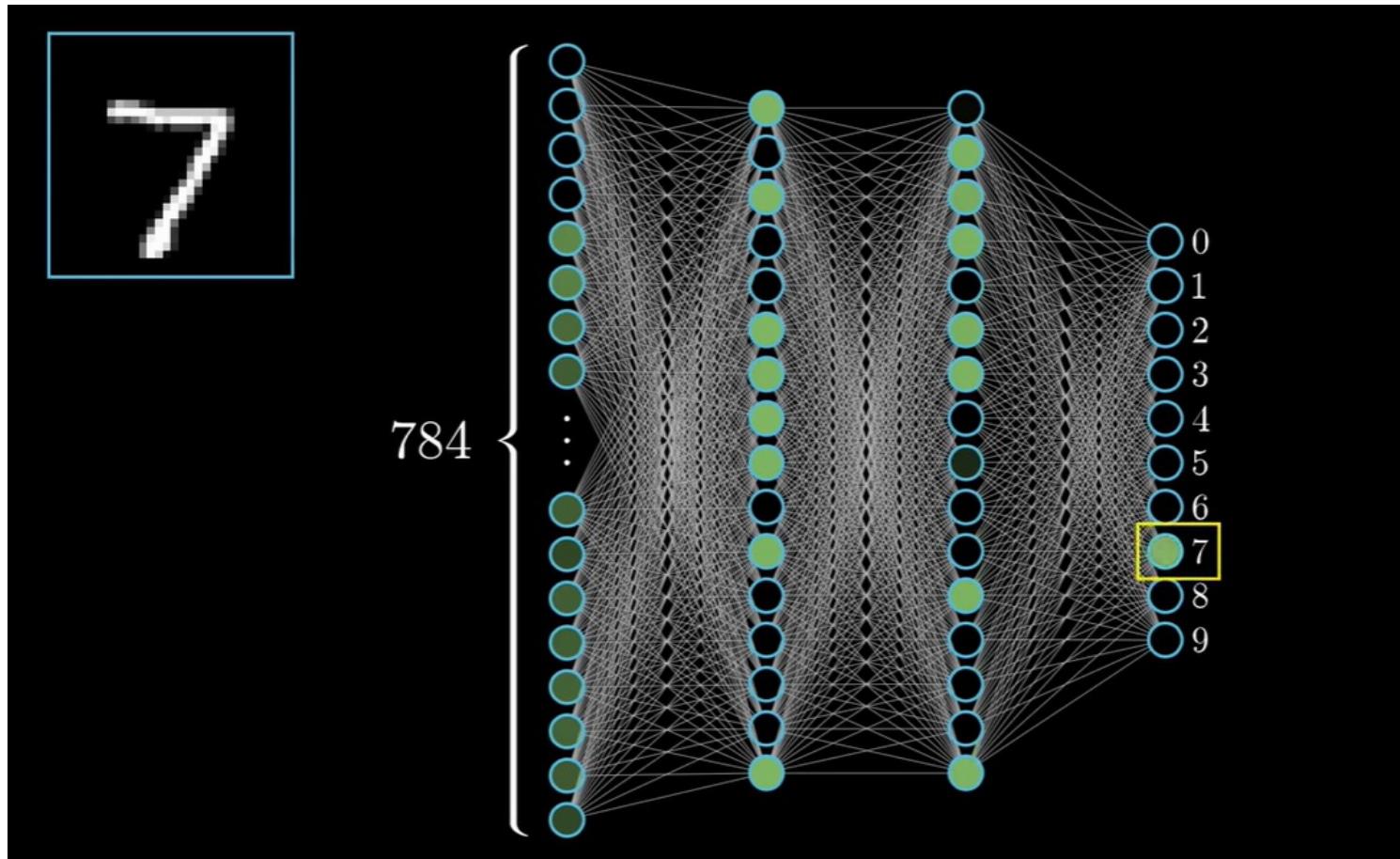
Beispiel: Handgeschriebene Zahl/Ziffer



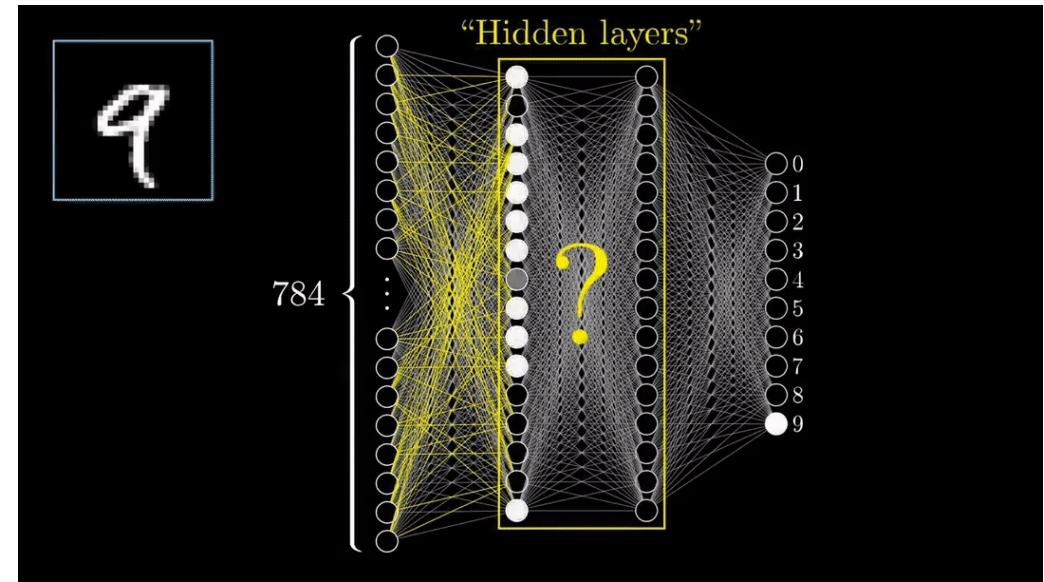
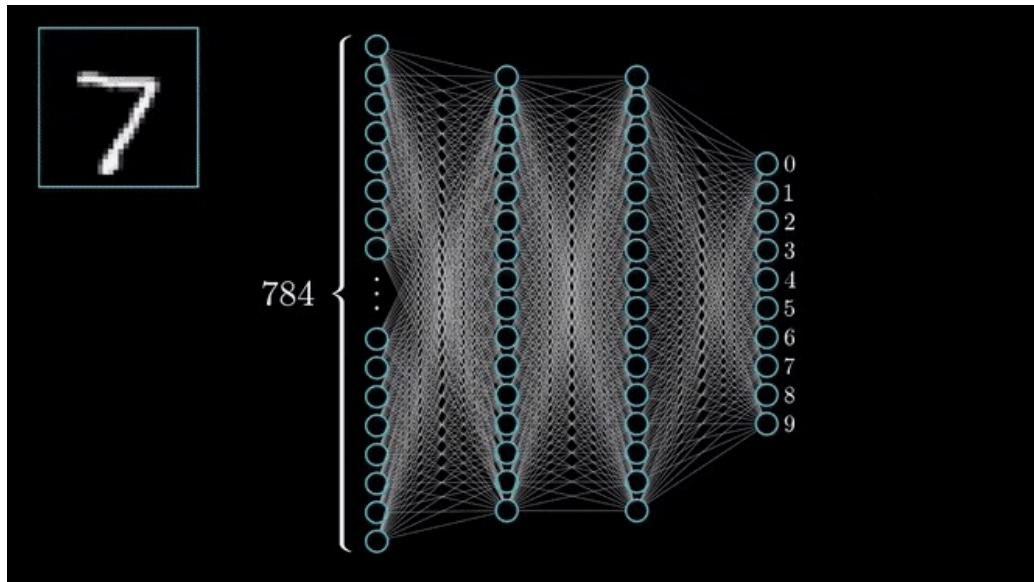
Zahl erkennen



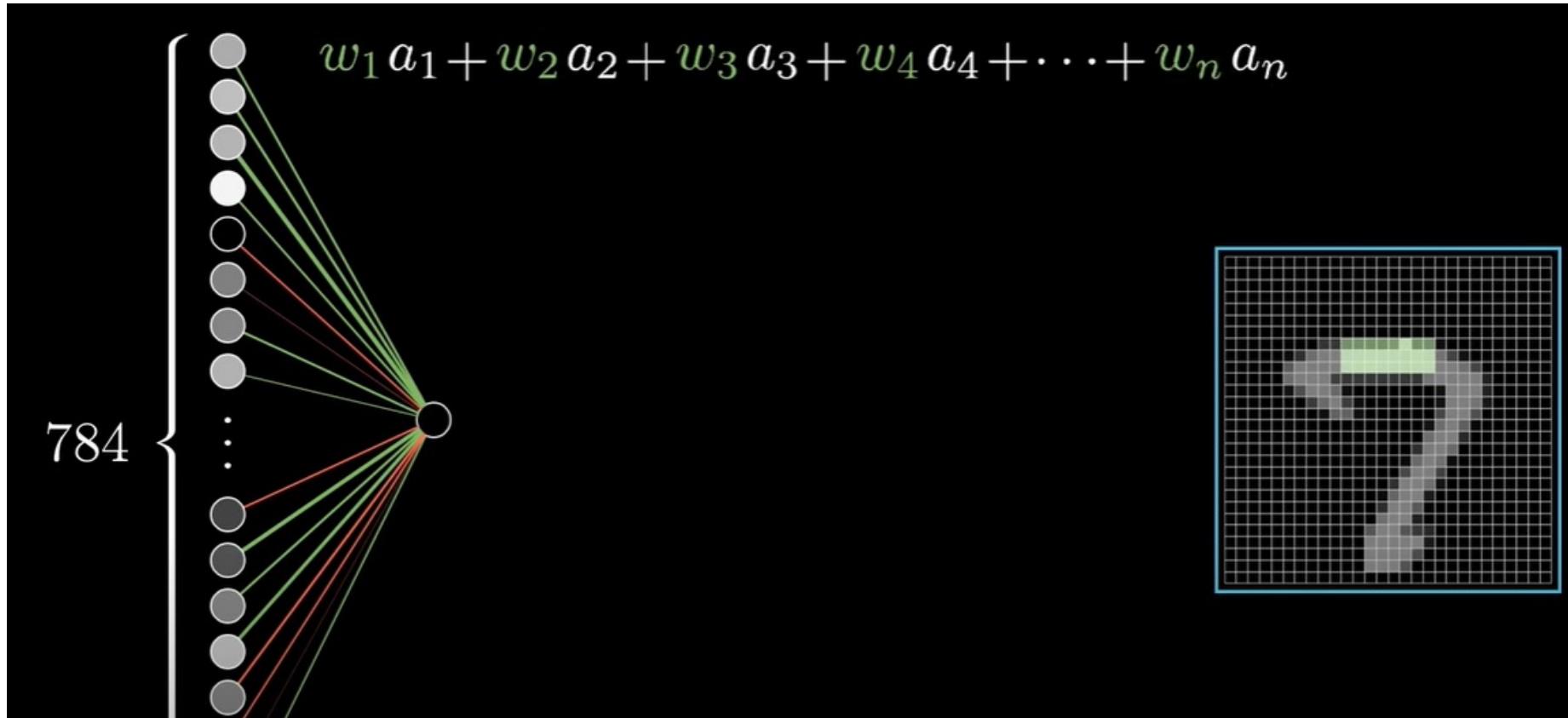
Zahl erkennen – Neuronales Netz



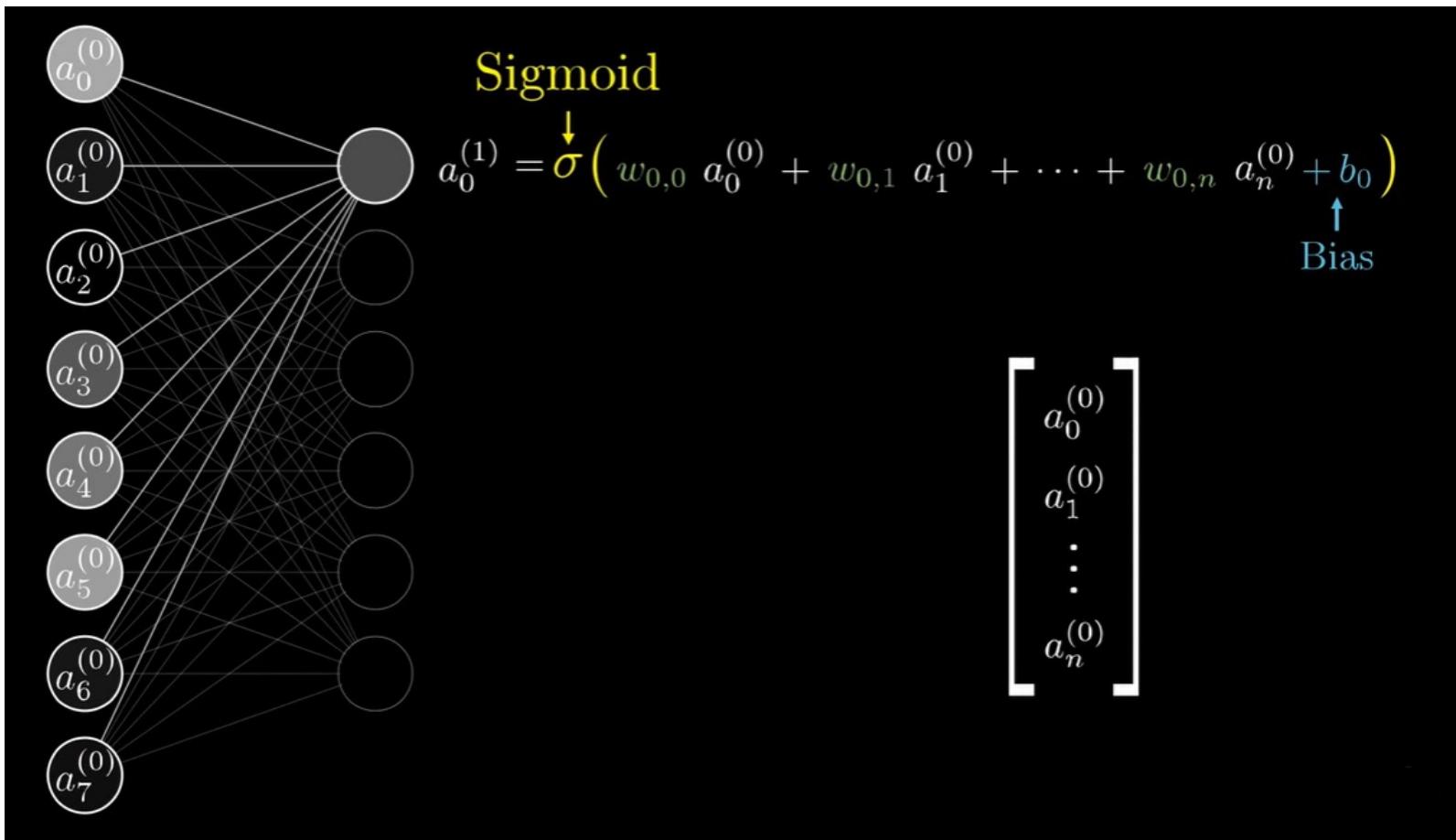
Zahl erkennen – Neuronales Netz



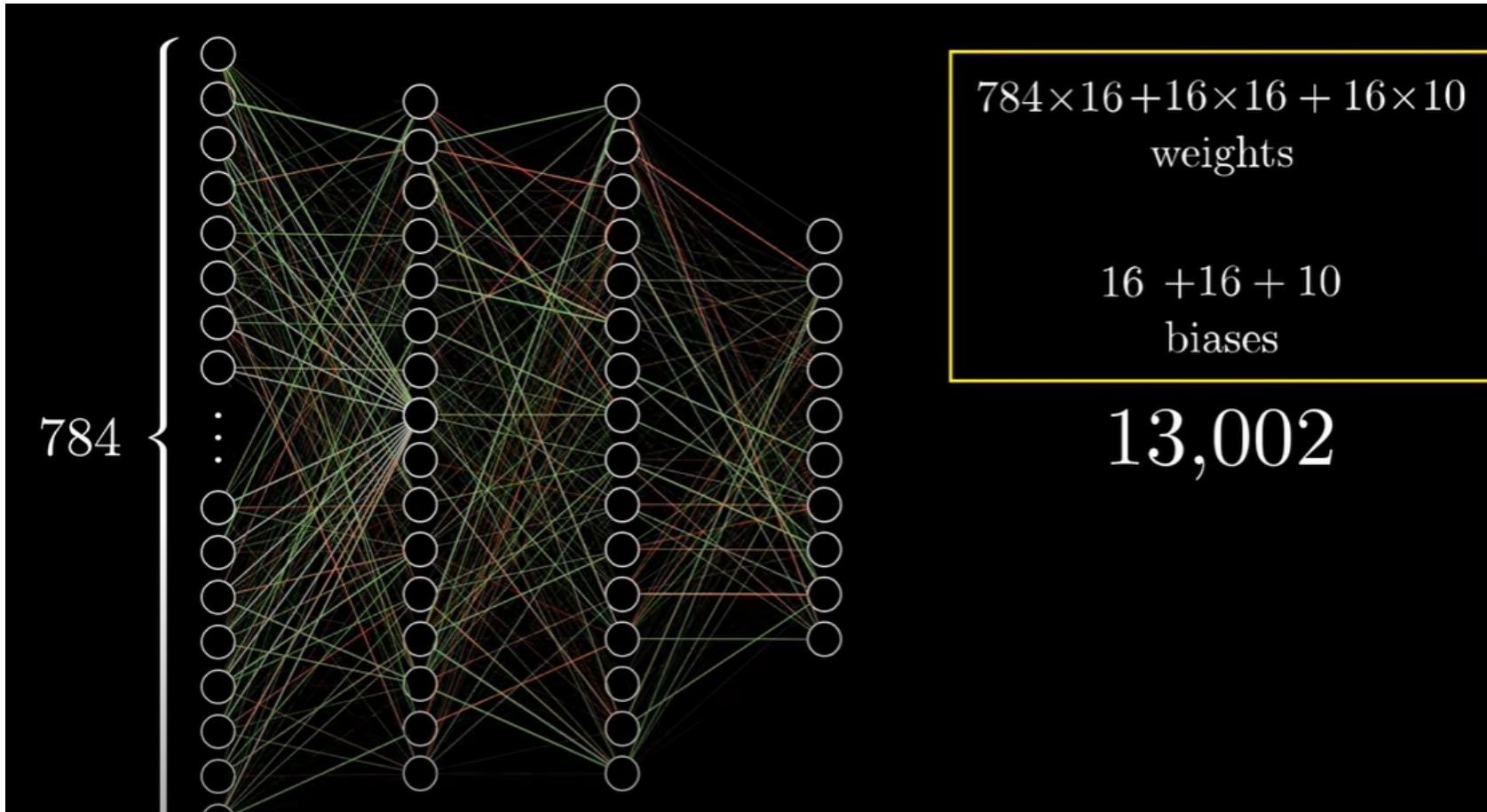
Berechnung Neuronales Netz



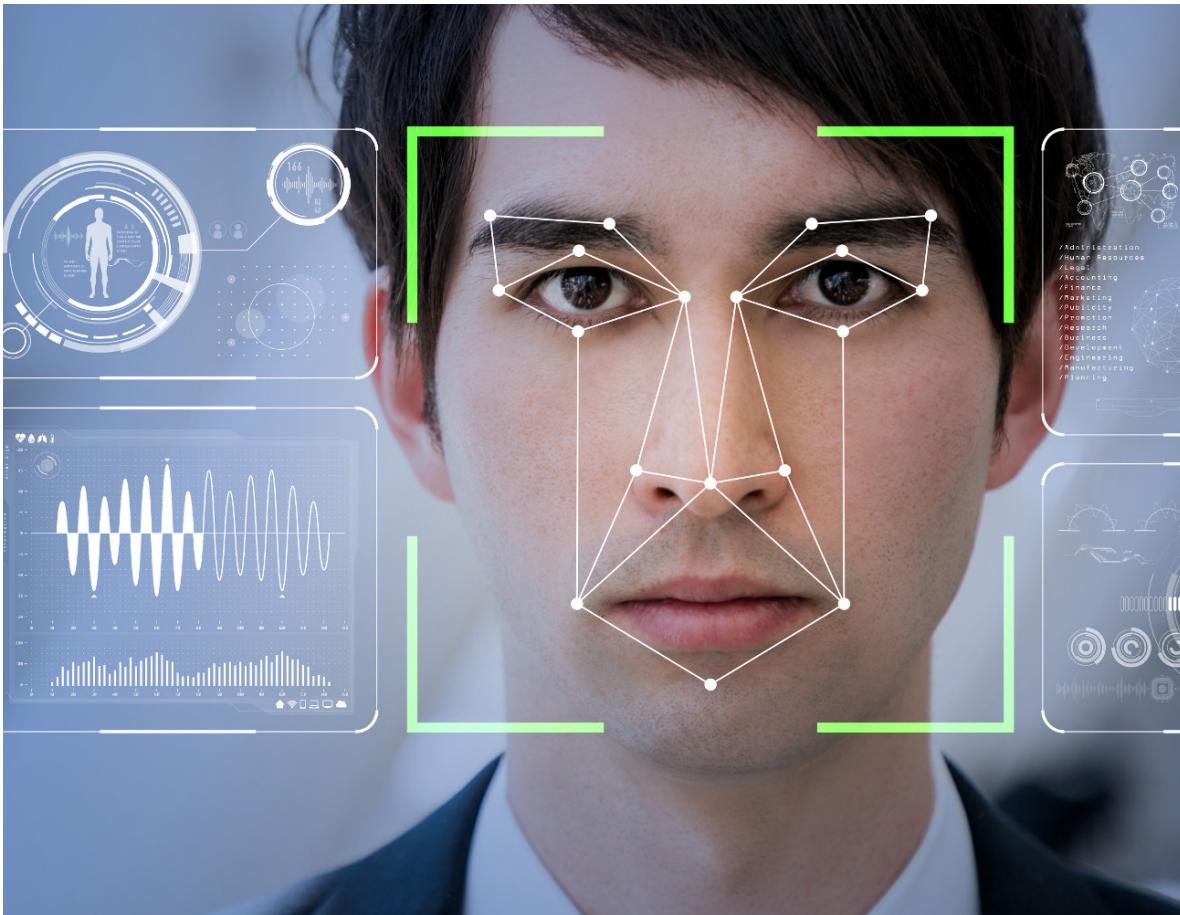
Berechnung Neuronales Netz



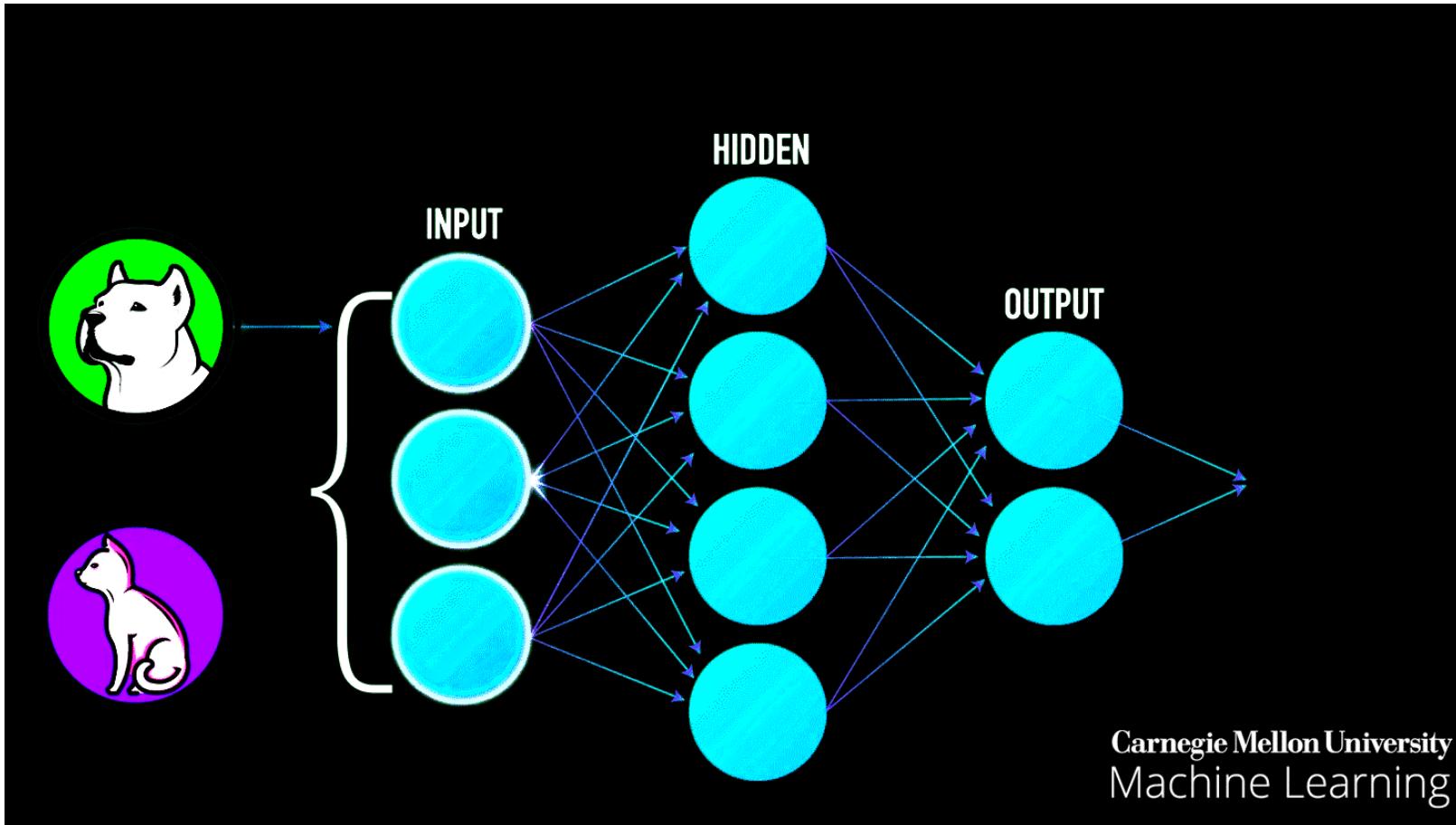
Anzahl Parameter



Fotografie



Neuronales Netz - Klassifikation



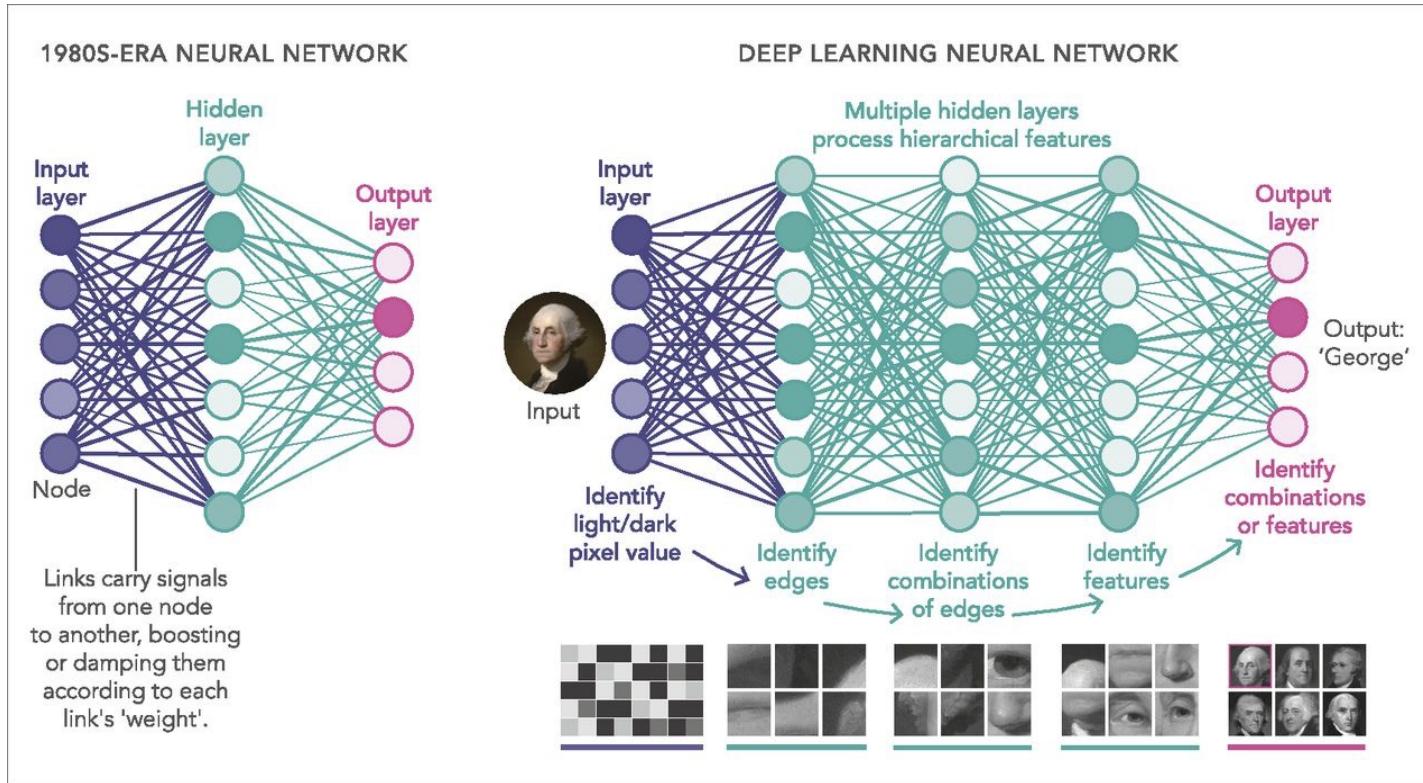
Auftrag

Welche Anwendungsgebiete kennen Sie?

Neuronales Netz

Lernen bedeutet in diesem Zusammenhang

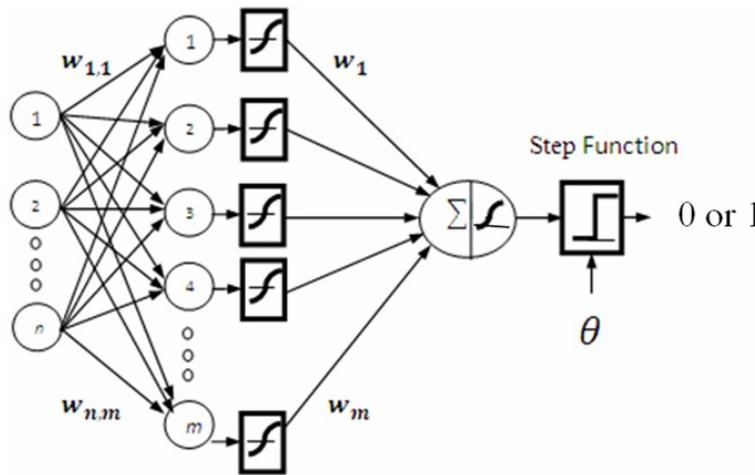
- Valide Werte für alle diese Gewichte und Biases zu finden
- Um das Problem schliesslich lösen zu können
 - Nämlich Ziffern zu klassifizieren



Klassifikation

Binäre Klassifikation

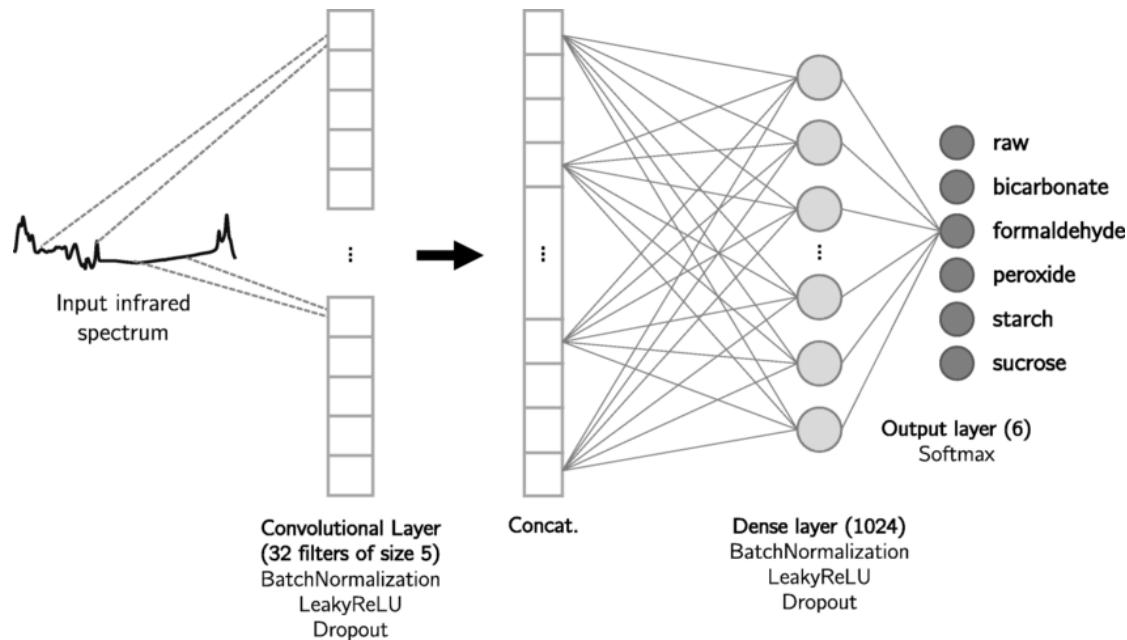
- Typischerweise positive und negative Klasse
 - Positiv, wenn berechnete Wahrscheinlichkeit höher als gegebene Schwelle
 - Negativ, wenn berechnete Wahrscheinlichkeit niedriger als gegebene Schwelle
 - Ja/Nein Antworten, Hund/Katze,...
 - Jedoch: Meistens mehr als 2 Entscheidungen



Multi-Klassifikation

Multiklassen Klassifikation

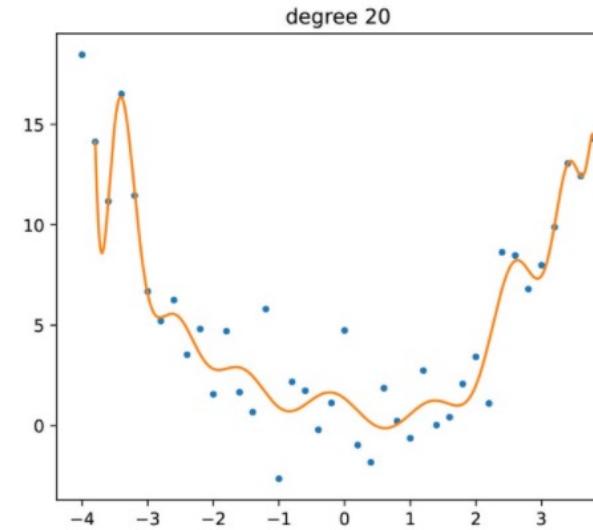
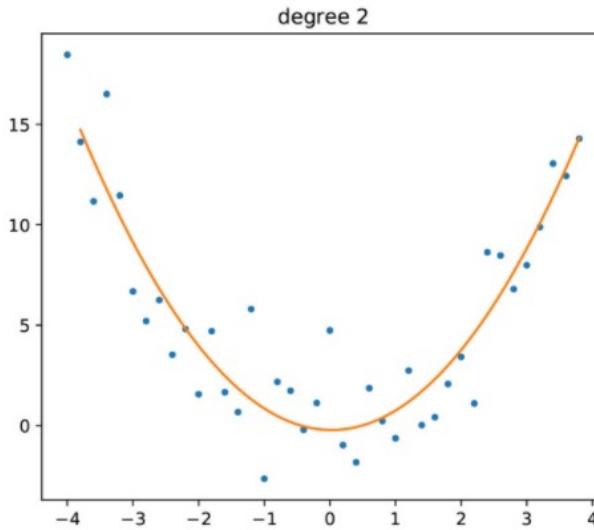
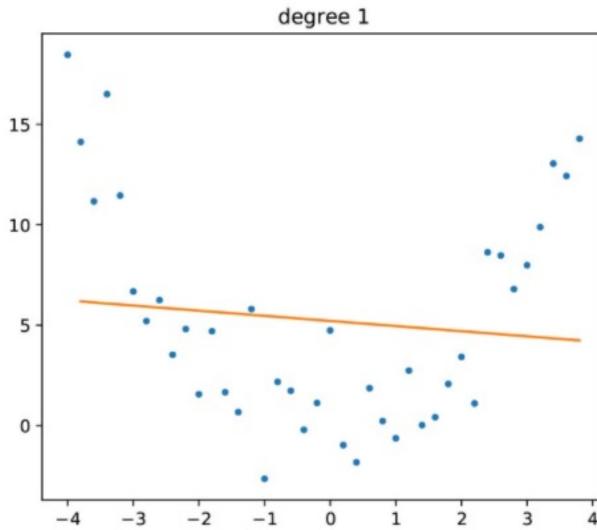
- Klassifizierung von Daten in mehrere Klassen
- 3 oder mehr Klassen, sonst binäre Klassifikation
- Beispiel: Erkennung handgeschriebener Ziffern



Regression

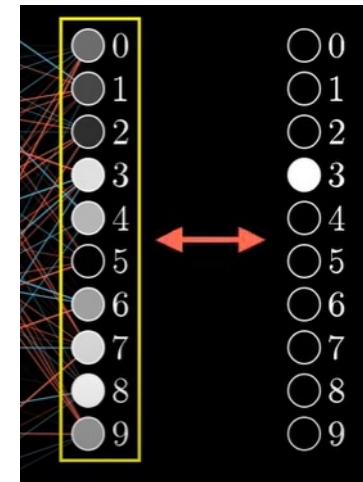
Regression

- Vorhersage kontinuierlicher/numerischer Werte
- Ziel eines Regressionsalgorithms'
 - Finde und plotte eine Best-Fit Linie oder Kurve



Verlustfunktion (Loss function)

- Eine Verlustfunktion misst in der Statistik und im Data Science wie gut ein Modell/Algorithmus den Datensatz beschreibt.
- Eine Verlustfunktion misst wie sich die Vorhersage vom Modell vom reellen Wert unterscheidet
- Die Verlustfunktion unterscheiden sich im Falle von Regression und Klassifikation.



Was ist der Unterschied zwischen dem berechneten Ergebnis und dem tatsächlichen Ergebnis (Label)?

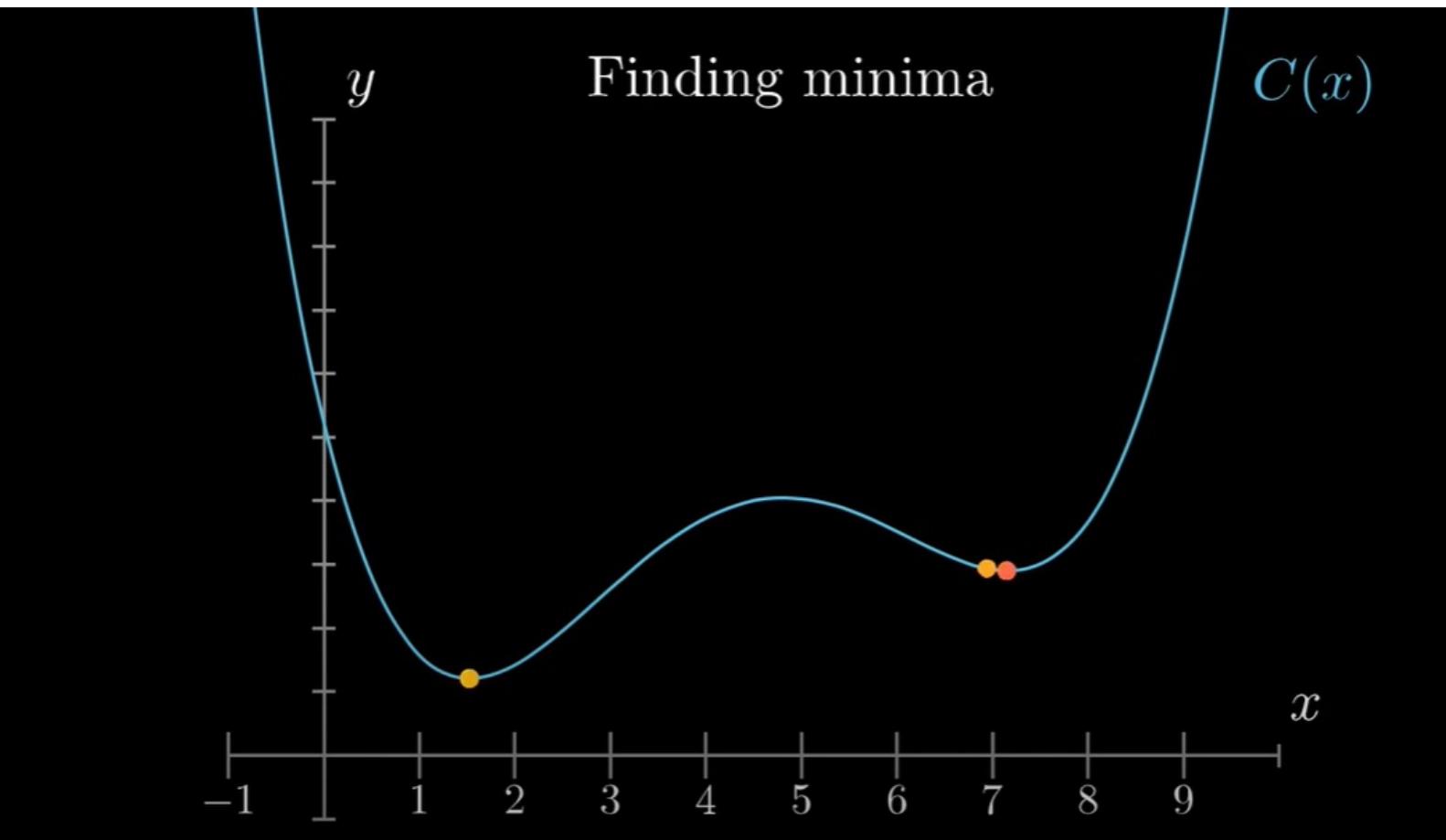
Auftrag

Setzen Sie sich mit einem Kollegen zusammen und halten Sie alle Verlustfunktionen die sie kennen fest.

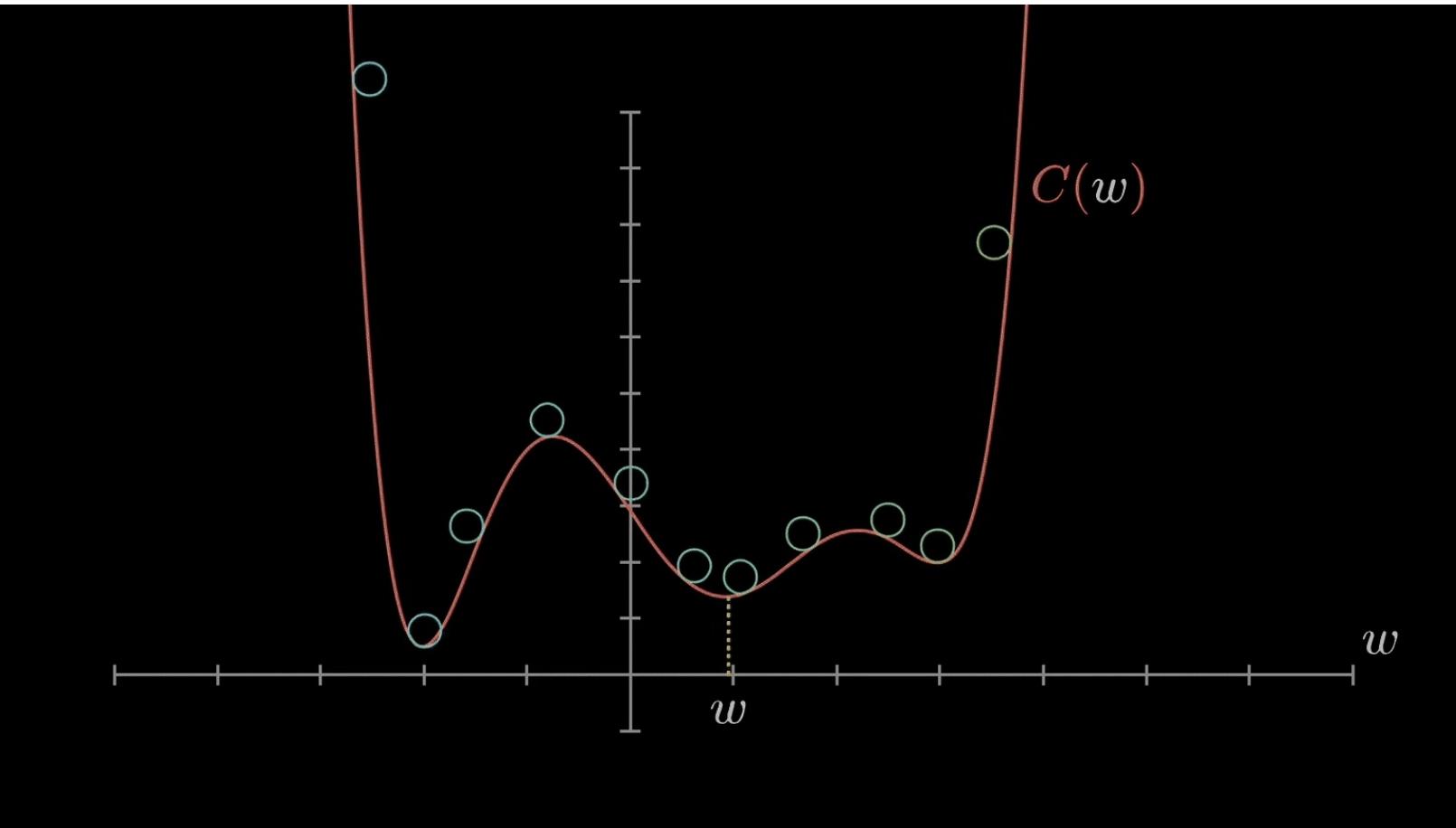
Gradient Descent

- Deep Learning Modelle werden mithilfe des Gradienteverfahren optimiert
- Das Gradientenverfahren minimiert die Verlustfunktion/Kostenfunktion
- Die Verlustfunktion ist abhängig vom Problem
- Die Verlustfunktion muss ableitbar sein
- Mögliche Funktionen sind
 - Mean Squared Error (MSE): $\frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2$
 - Binary Cross-Entropy Loss: $-(y \cdot \log \hat{y} + (1 - y) \cdot \log(1 - \hat{y}))$
 - Categorical Corss-Entropy: $\sum_{i=1}^n y \cdot \log \hat{y}$

Gradient Descent

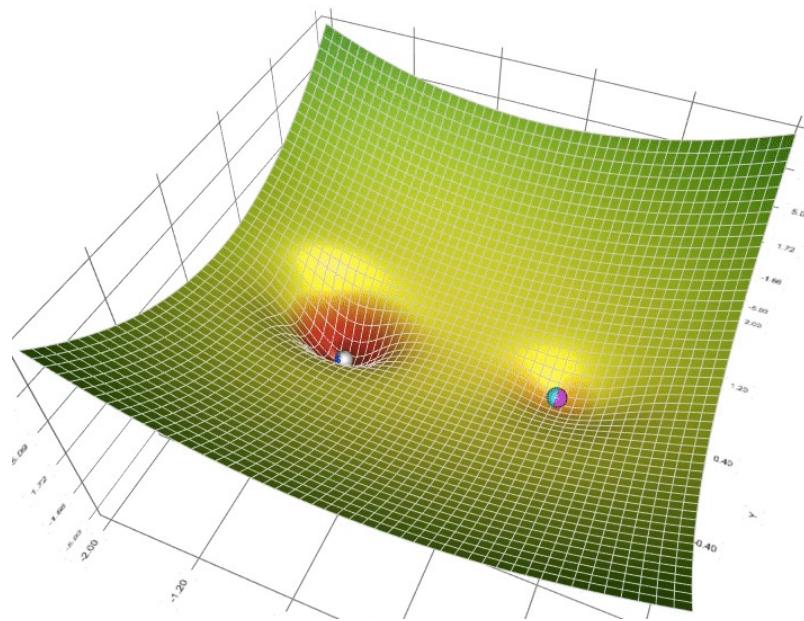


Gradient Descent



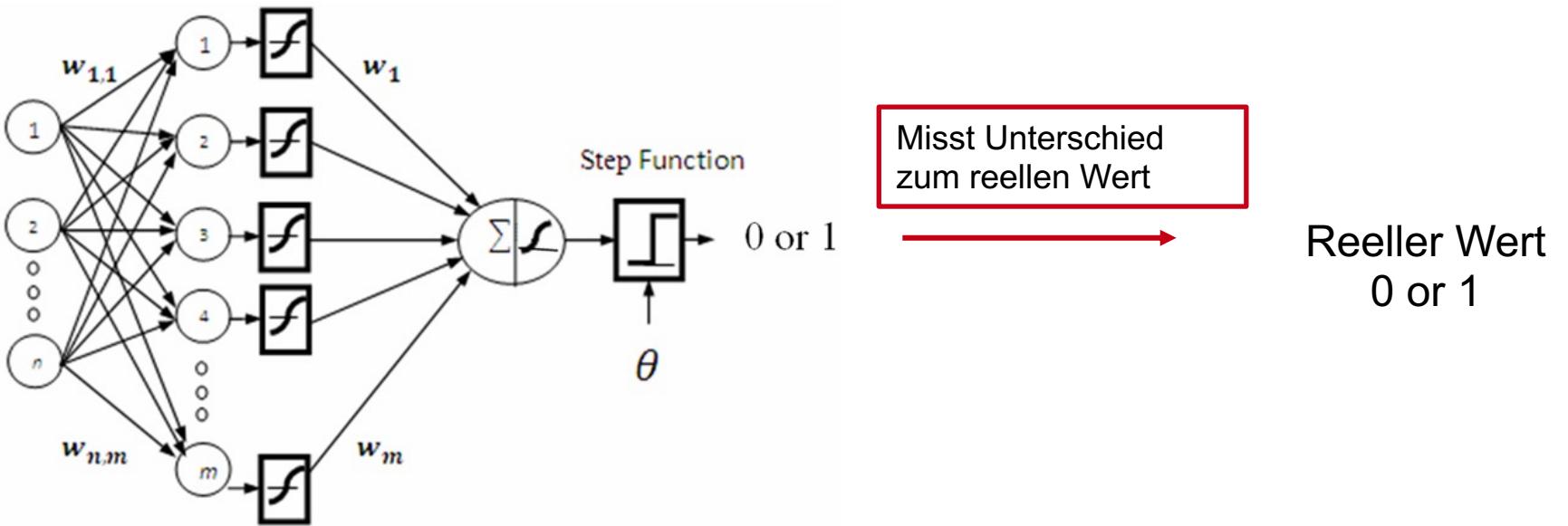
Schwierigkeit Gradient Descent

- Das Konvergieren in ein lokales Minimum einer Kostenfunktion
 - Durch wiederholtes Anwenden des gradiellen Abstiegs

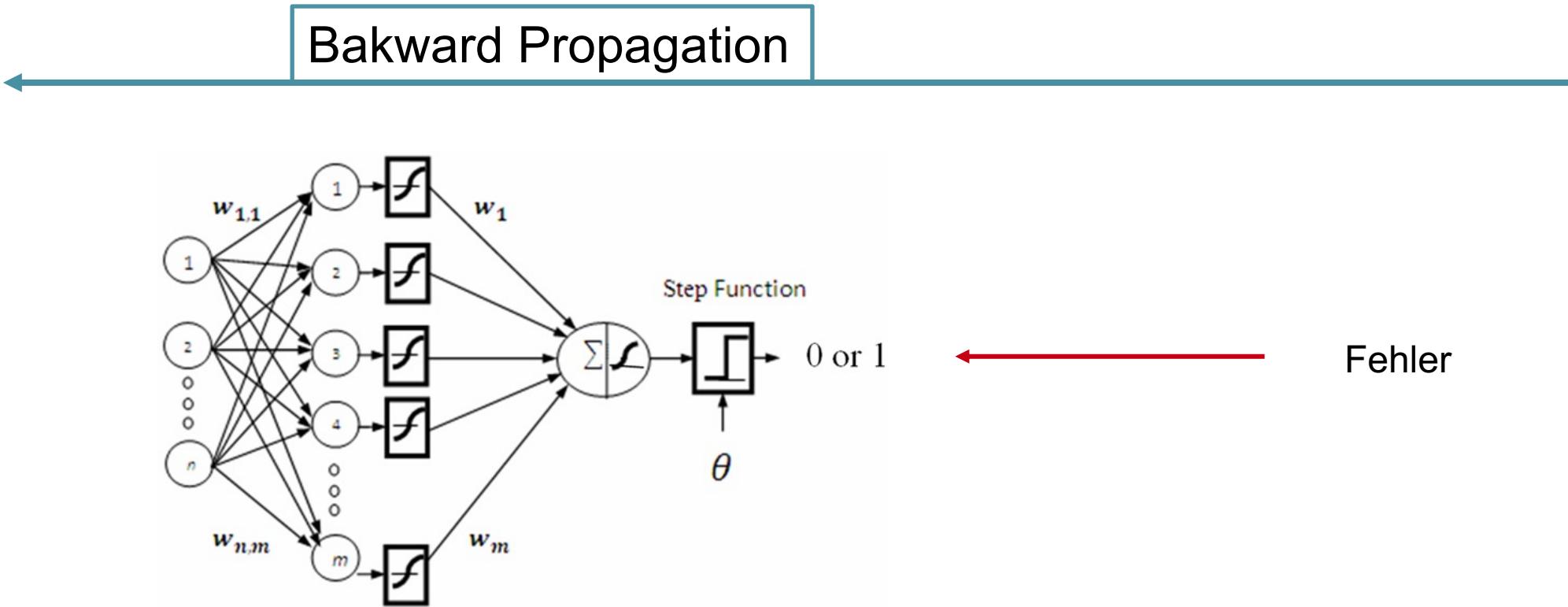


Forward Propagation

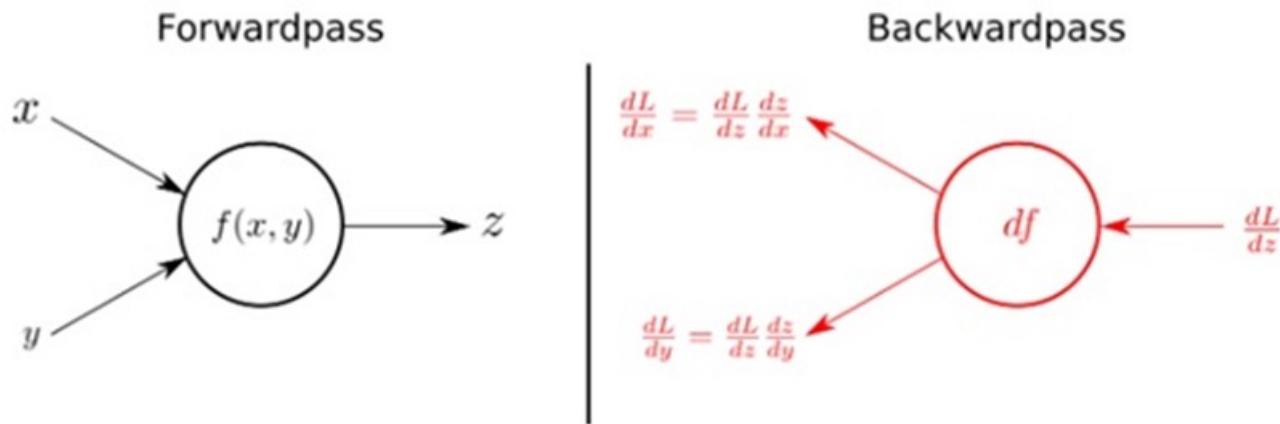
Forward Propagation



Backward Propagation



Forward and Bachward Propagation



Fragen



Darstellung eines Fragesymbol aufgerufen von der Webseite
<https://www.qnigge.de/news/detail/modul-v/#images> am
12.07.2021.

Fachhochschule Graubünden
Pulvermühlestrasse 57
7000 Chur
T +41 81 286 24 24
info@fhgr.ch

**Vielen Dank für Ihre Aufmerksamkeit.
Grazia fitg per l'attenziun.
Grazie per l'attenzione.**

Fachhochschule Graubünden
Scola auta spezialisada dal Grischun
Scuola universitaria professionale dei Grigioni
University of Applied Sciences of the Grisons

swissuniversities



Institutionell akkreditiert nach
HFKG 2018–2025