

Wie funktioniert Künstliche Intelligenz und was sollte man darüber wissen?

Ein Vortrag für Einsteiger - von Alexandra Gessner und
Patrick Klügel



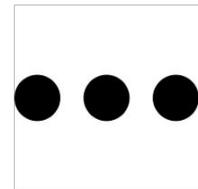
Agenda

1. Vorstellung, Einführung: Warum überhaupt ein Vortrag zu Künstlicher Intelligenz?
2. Was stellen wir uns unter Künstlicher Intelligenz vor?
3. Wie funktioniert maschinelles Lernen/Machine Learning?
4. Ethische Aspekte, soziale Überlegungen, aktuelle Fragen
5. Was passiert am Forschungsstandort Tübingen?
6. Fragen und Diskussion

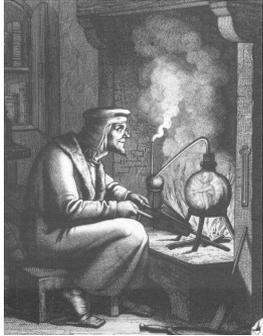
Ende: Machen Sie mit!

1. Warum überhaupt ein Vortrag zu Künstlicher Intelligenz (KI)?

- Einsatz von KI nimmt zu: Soziale Medien, Smartphone
- Bernhard Schölkopf: “Künstliche Intelligenz wird Leben und Arbeiten so stark ändern wie einst Elektrizität.”
- Wir brauchen mehr **KI-Kompetenz**: Wissen, kritische Reflektion, Souveränität
- Besondere Bedeutung von Tübingen: engagierte Stadtgesellschaft!



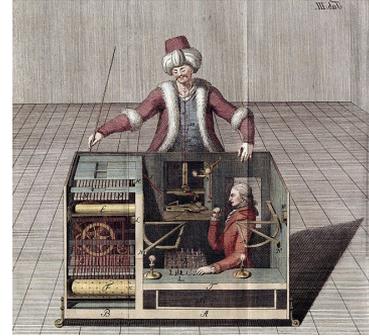
2. Was stellen wir uns unter “Künstlicher Intelligenz” vor?



Mittelalter



um 1770



1769



1844



1916



1984



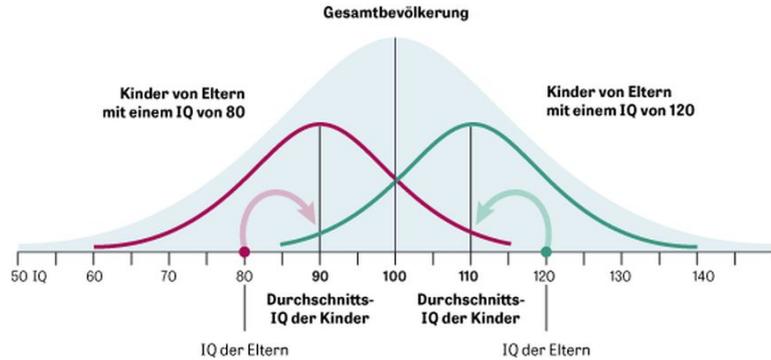
2014

...und heute?

Was ist eigentlich Intelligenz?

Alles strebt zur Mitte

Wie klug der Nachwuchs wird, hängt nicht nur von den Genen ab



3. Wie funktioniert ~~künstliche Intelligenz~~ maschinelles Lernen?

Künstliche Intelligenz existiert nicht im eigentlichen Sinne des Wortes, denn:

- Wir sind **weit entfernt** von einer allgemeinen KI (*engl. artificial general intelligence*)
 - "KI" löst **klar definierte, stark eingegrenzte** Probleme
- ➔ In der Forschung wird von **maschinellern Lernen** gesprochen

Was ist maschinelles Lernen (ML)?

Das maschinelle Lernen entwickelt Verfahren, mit denen Computer **anhand von Beispielen selbstständig bestimmte begrenzte Aufgaben** lösen können.

Anwendungsbeispiel:

FLORA incognita



The image shows a smartphone displaying the FLORA incognita app interface. The screen features a central flower icon with a blue plus sign, labeled "Pflanze erkennen". To the left is a menu icon labeled "Alle Arten", and to the right is an eye icon labeled "Meine Beobachtungen". The background of the app is a grid of various flower images. Below the smartphone are three download buttons: "Download on the App Store", "GET IT ON Google Play", and "EXPLORE IT ON AppGallery".

leben.natur.vielfalt
die Strategie

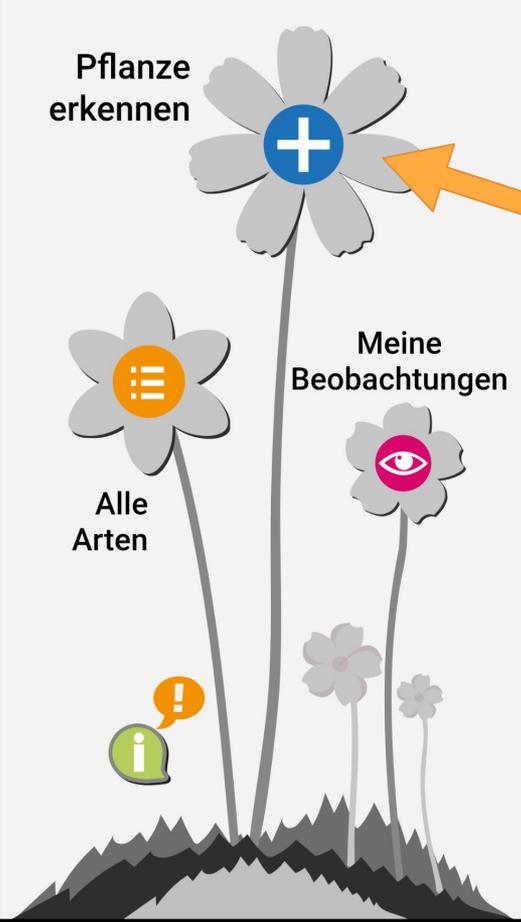


Freistaat
Thüringen



Ministerium
für Umwelt, Energie
und Naturschutz





× Beobachtung



Wuchsform



Wildblume oder Strauch



Baum



Gras



Farn

Der Standort der Pflanze erlaubt eine genauere Bestimmung.





FLORA
incognita

The logo for 'FLORA incognita' features the word 'FLORA' in large, blue, sans-serif capital letters. Below it, the word 'incognita' is written in a smaller, green, lowercase sans-serif font. To the right of the text is a stylized illustration of a smartphone with a white flower icon on its screen.

15:48



4G 93%



Arthur Chapman (CC BY-NC-SA 2.0)

Färber-Mädnichwe

Coreopsis tinctoria

Taxonomie

Gattung **Coreopsis**

Familie **Asteraceae (Korbblütengewächse)**

Beschreibung

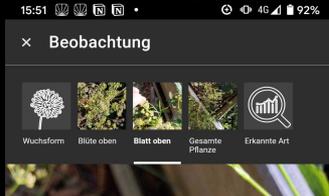
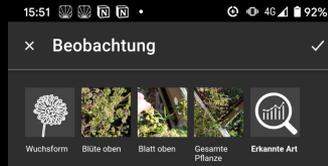
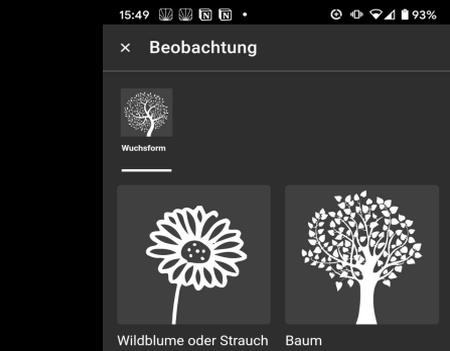
Schnell wachsende *Coreopsis tinctoria*-Pflanzen erreichen Höhen von 30 bis 100 cm (12 bis 40 Zoll). Die Blätter sind gefiedert, kahl und dünn und neigen dazu, an der Spitze

Verwendungszweck

Ökologie und Status

floristischer Status **Neophyt**
(DE)





Erkannte Art

Wir konnten die Art nicht mit ausreichender Sicherheit bestimmen. Unten sehen Sie die ähnlichsten Treffer.

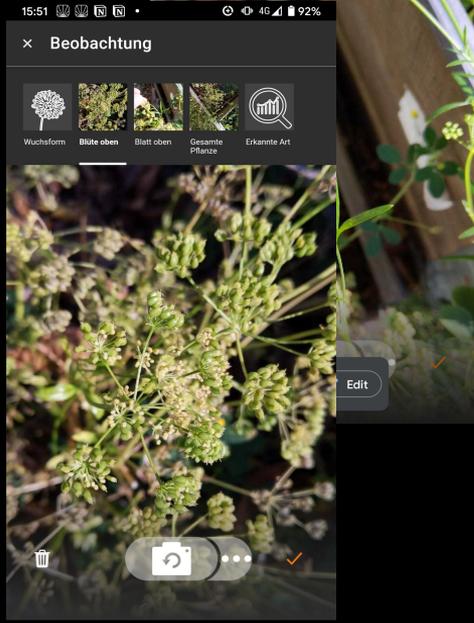
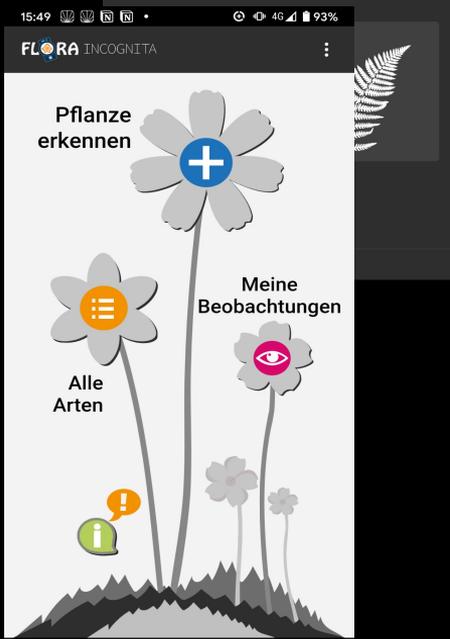
Ähnliche Treffer

- Wiesensilge**
Silaum silaus
 - Unbekannte Spezies**
Die Art vorschlagen...
- Vergleichen Sie die Steckbriefe und be...
die ähnlichste Art.
- Notiz



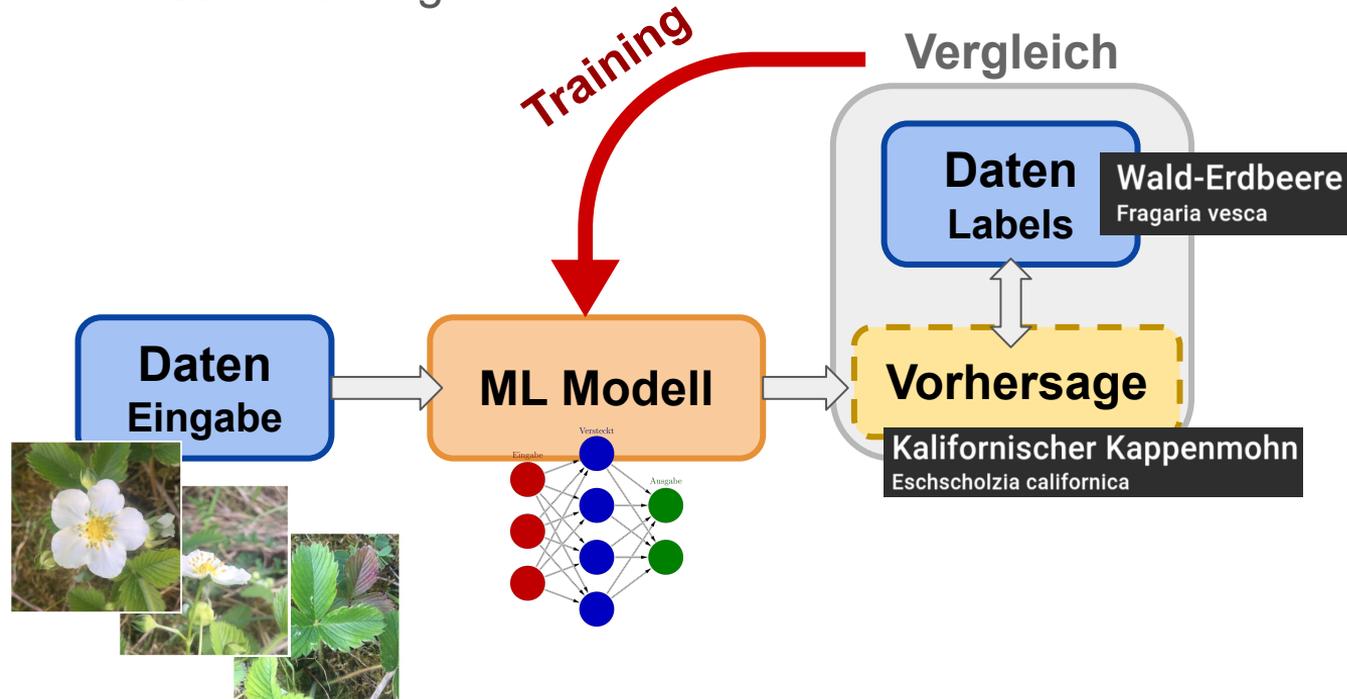
Wiesensilge
Silaum silaus

Taxonomie
Gattung Silaum (Silge) Familie Apiaceae (Doldenblütler)
Beschreibung
30-100 cm. Stängel kahl, kaum gerieft. Grundblätter 2-4fach gefiedert. Blattzipfel lineal lanzettlich, bespitzt. Dolden 5-10-strahlig. Strahlen verschieden lang. Hülle 0-3-
Verwendungszweck
Bestäubung
Ökologie und Status



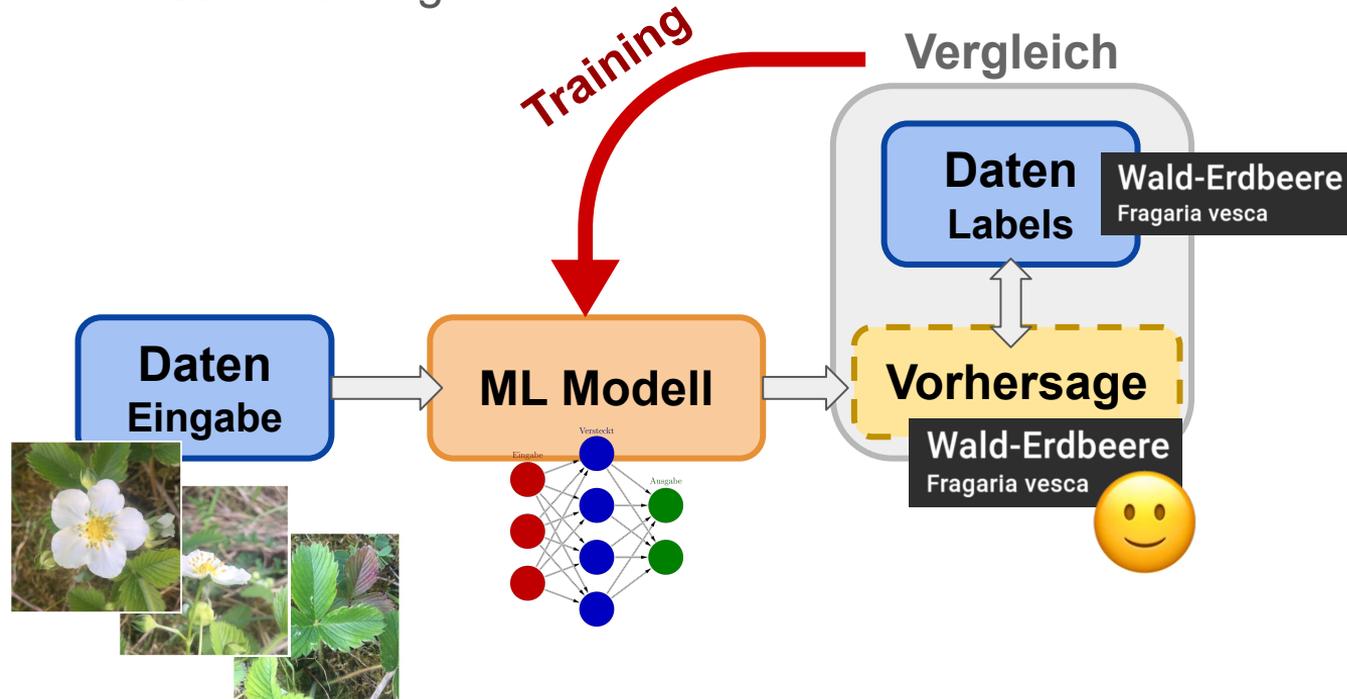
Woraus besteht eine ML Methode?

Phase I: Training



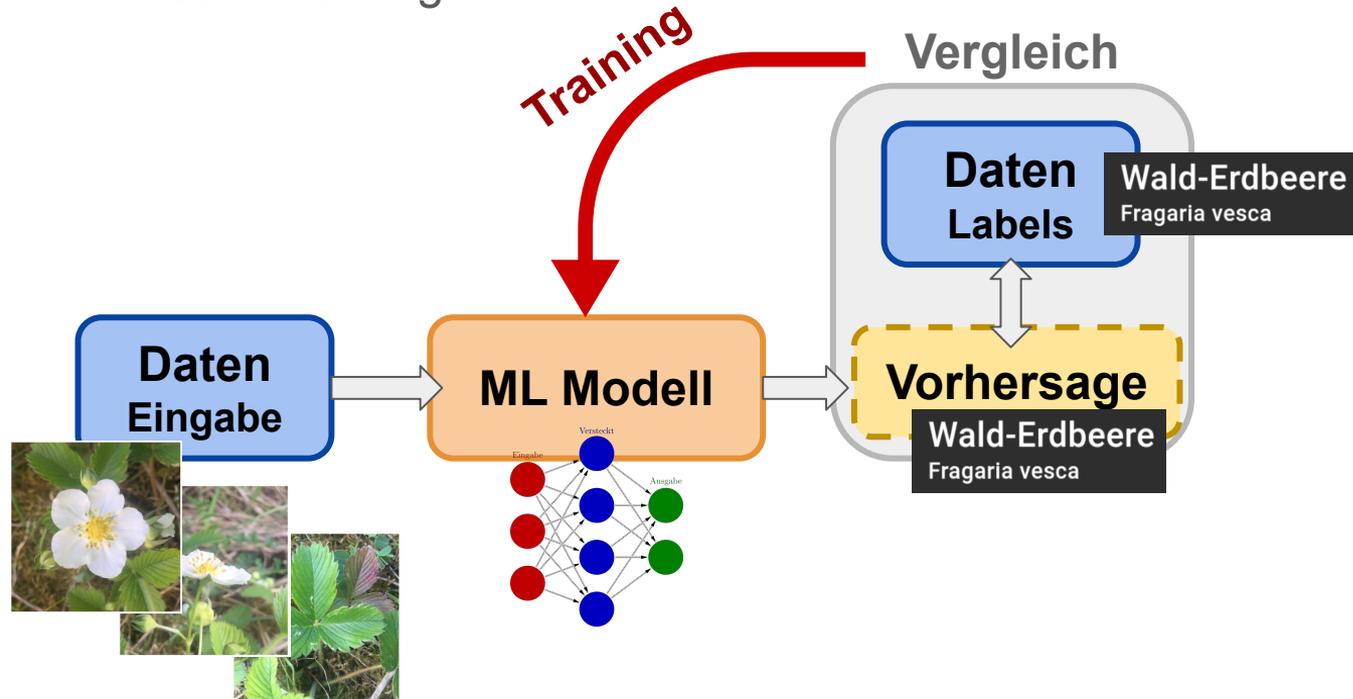
Woraus besteht eine ML Methode?

Phase I: Training

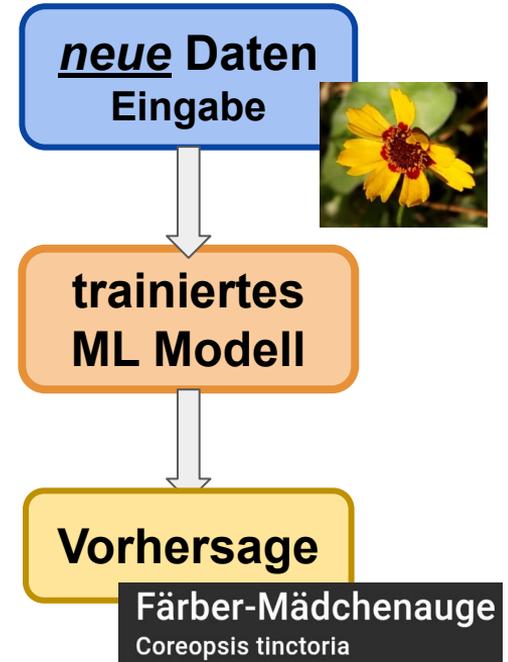


Woraus besteht eine ML Methode?

Phase I: Training

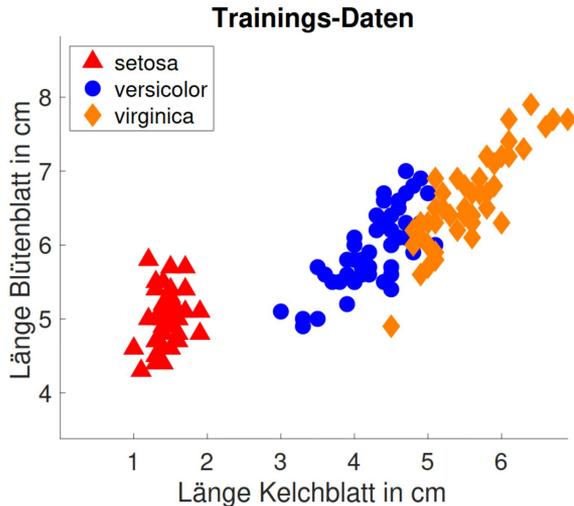


Phase II: Auswertung



Was bedeutet Training?

Aufgabe:
Klassifizierung von
Schwertlilien



Iris Versicolor



Iris Setosa



Iris Virginica

Daten:
Maße von
Blütenblättern

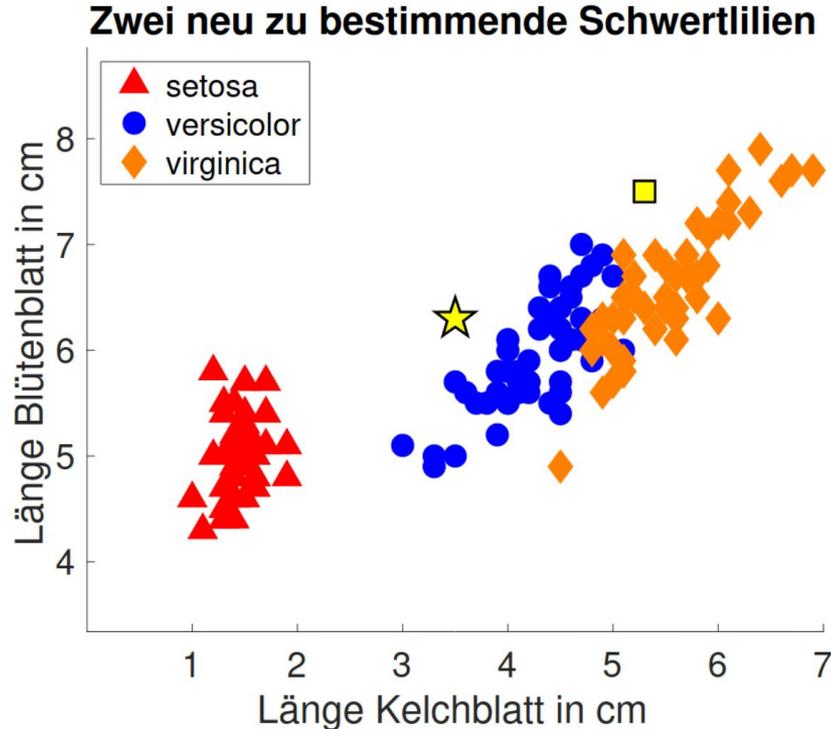
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Prof. v. Luxburg



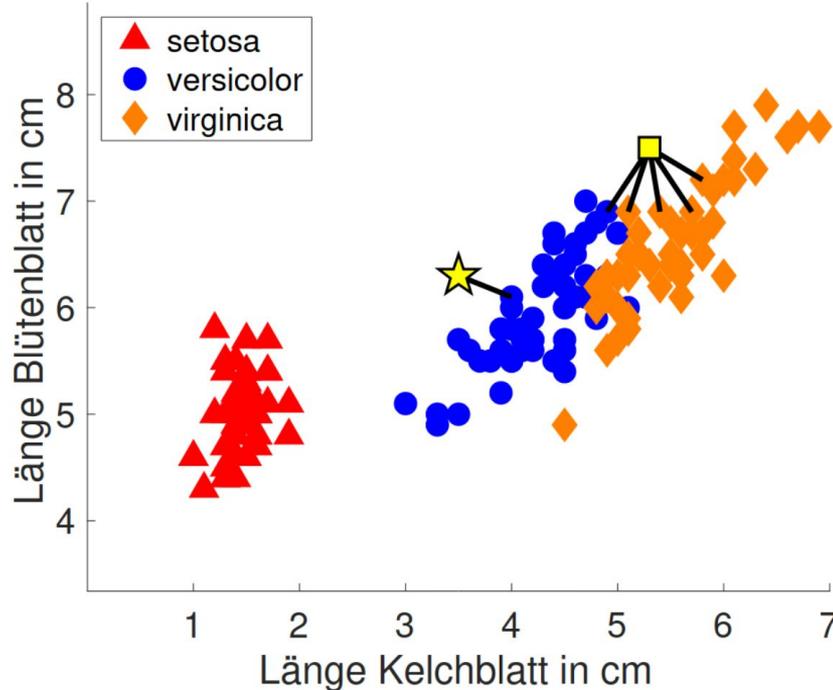
Klassifikation **neuer** Schwertlilien



Wie würden Sie
die neuen
Schwertlilien
einordnen?

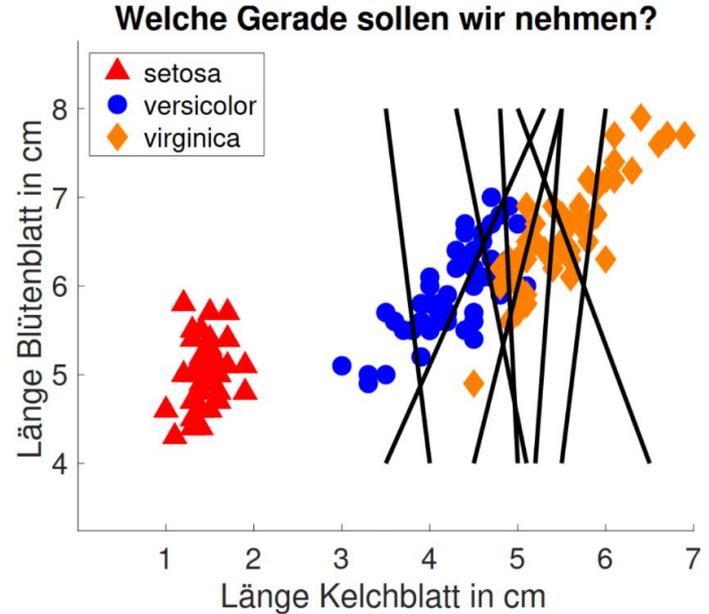
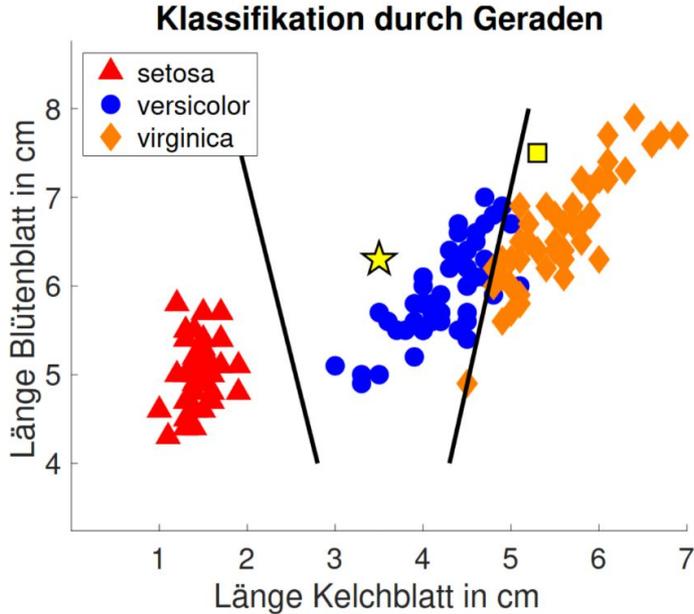
Klassifikation **neuer** Schwertlilien

Klassifikation durch nächste Nachbarn



Lernregel:
k=1,2,3,... nächste
Nachbarn

Klassifikation neuer Schwertlilien

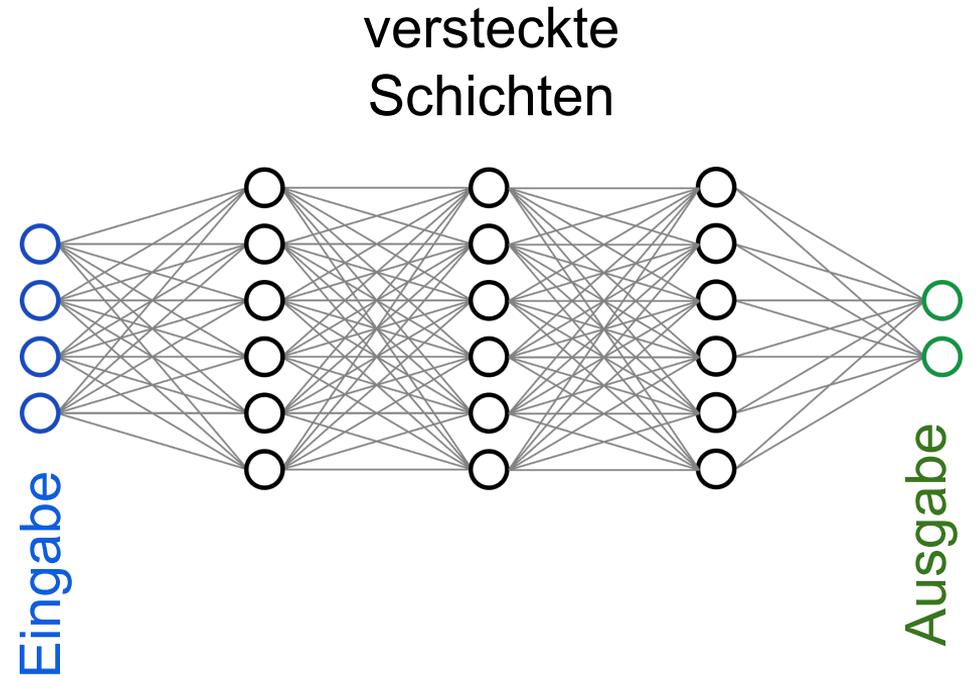
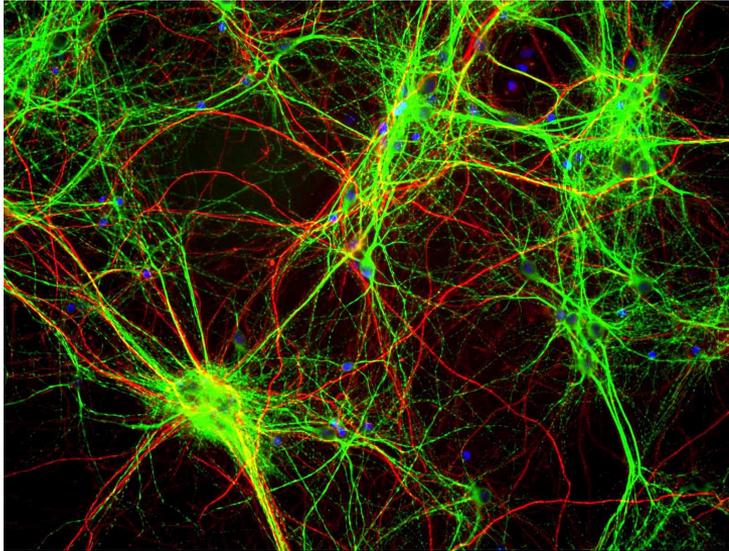


Modell
Trenngeraden

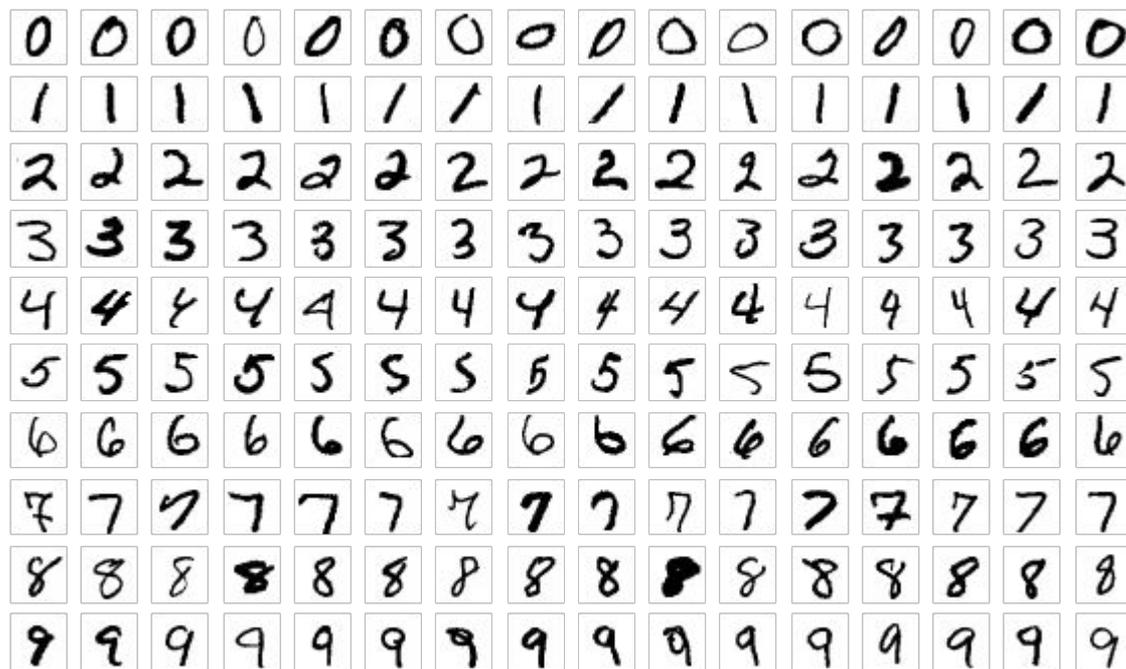
Bewertungsfunktion
Trainingsfehler

Lernregel
Trainingsfehler minimieren

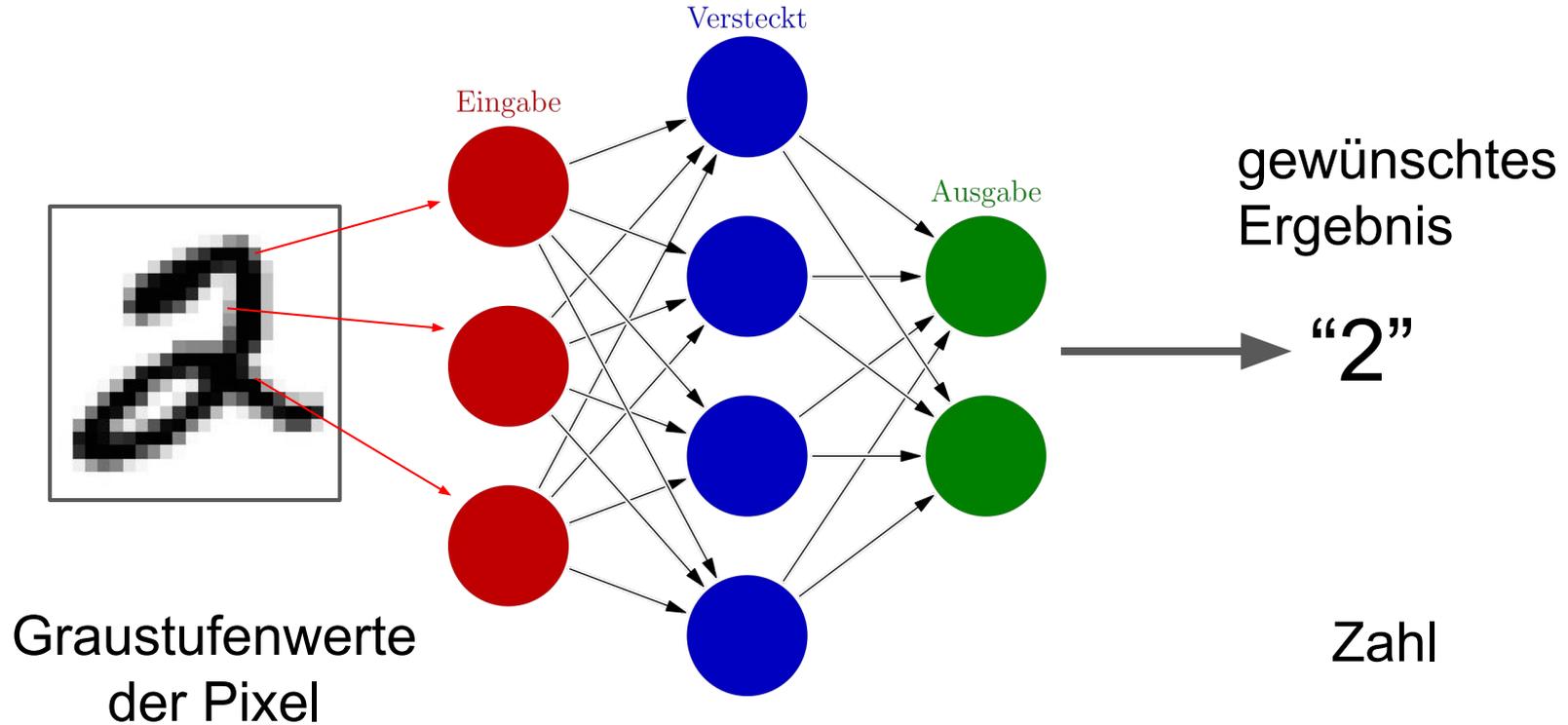
Künstliche neuronale Netze



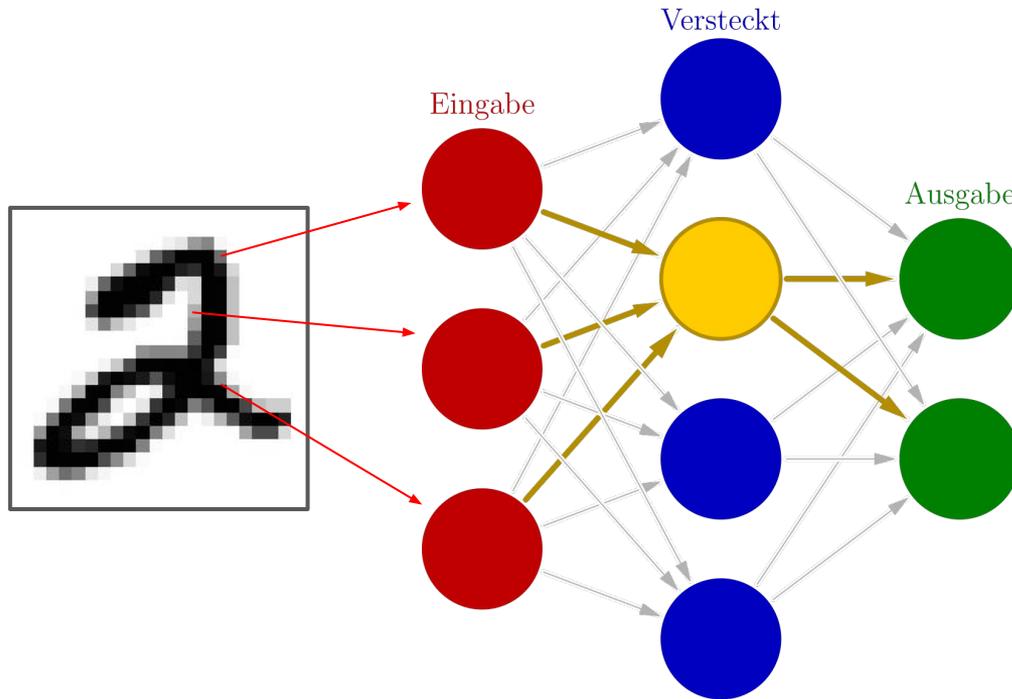
Beispiel: Erkennung handgeschriebener Zahlen



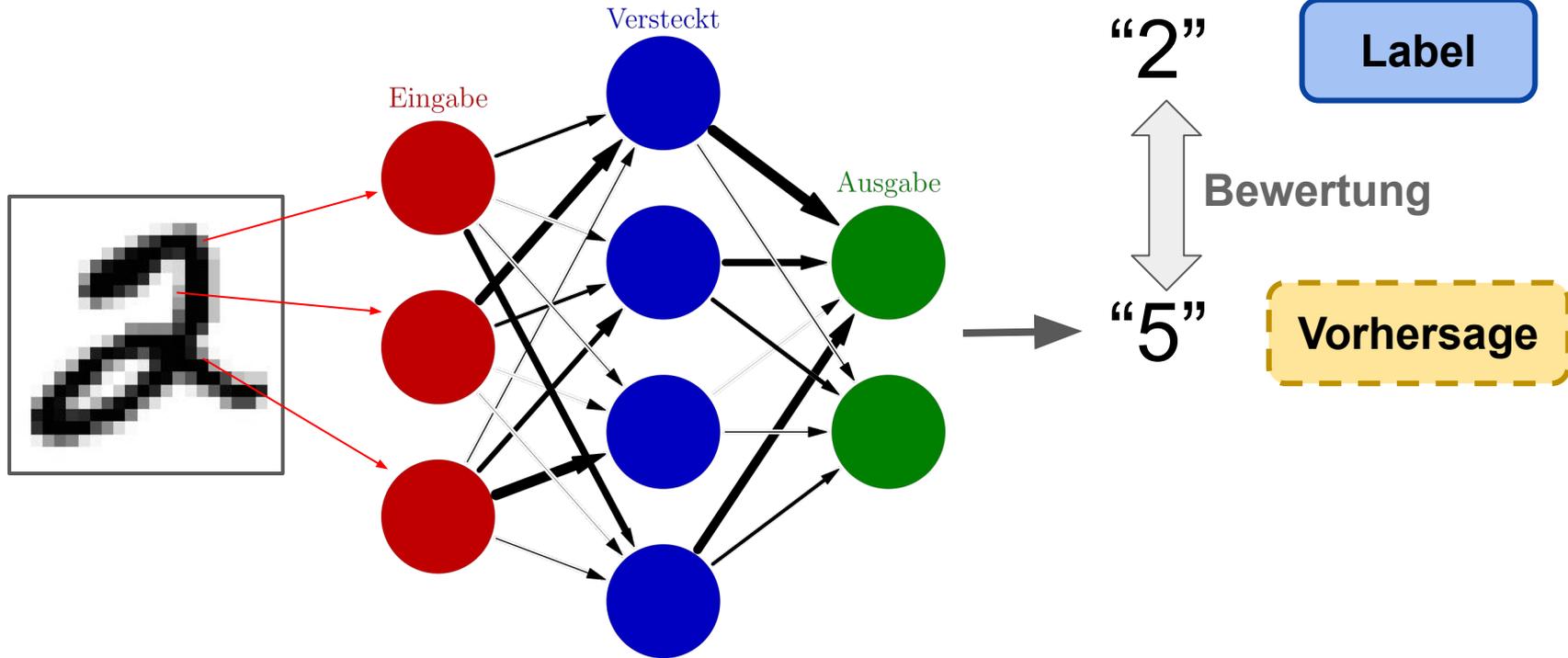
Künstliche neuronale Netze



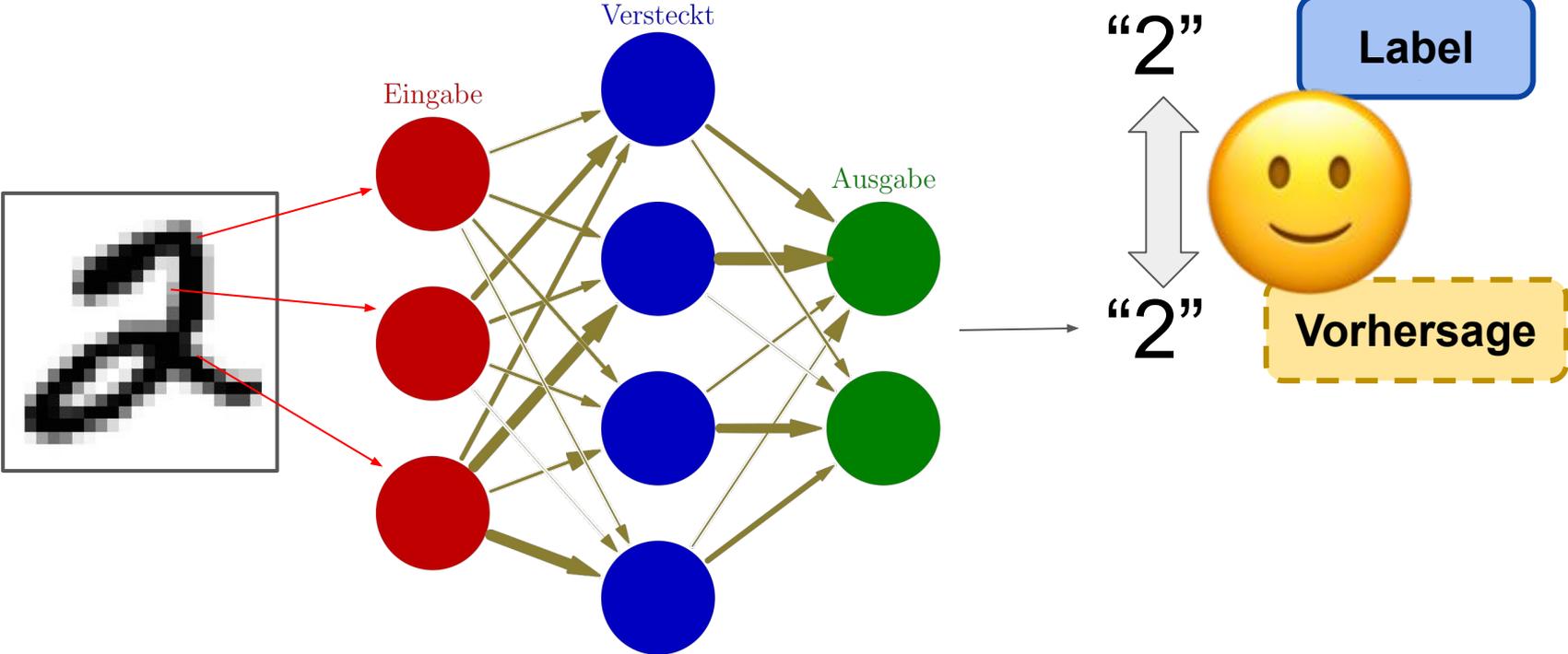
Was passiert in einem Neuron?



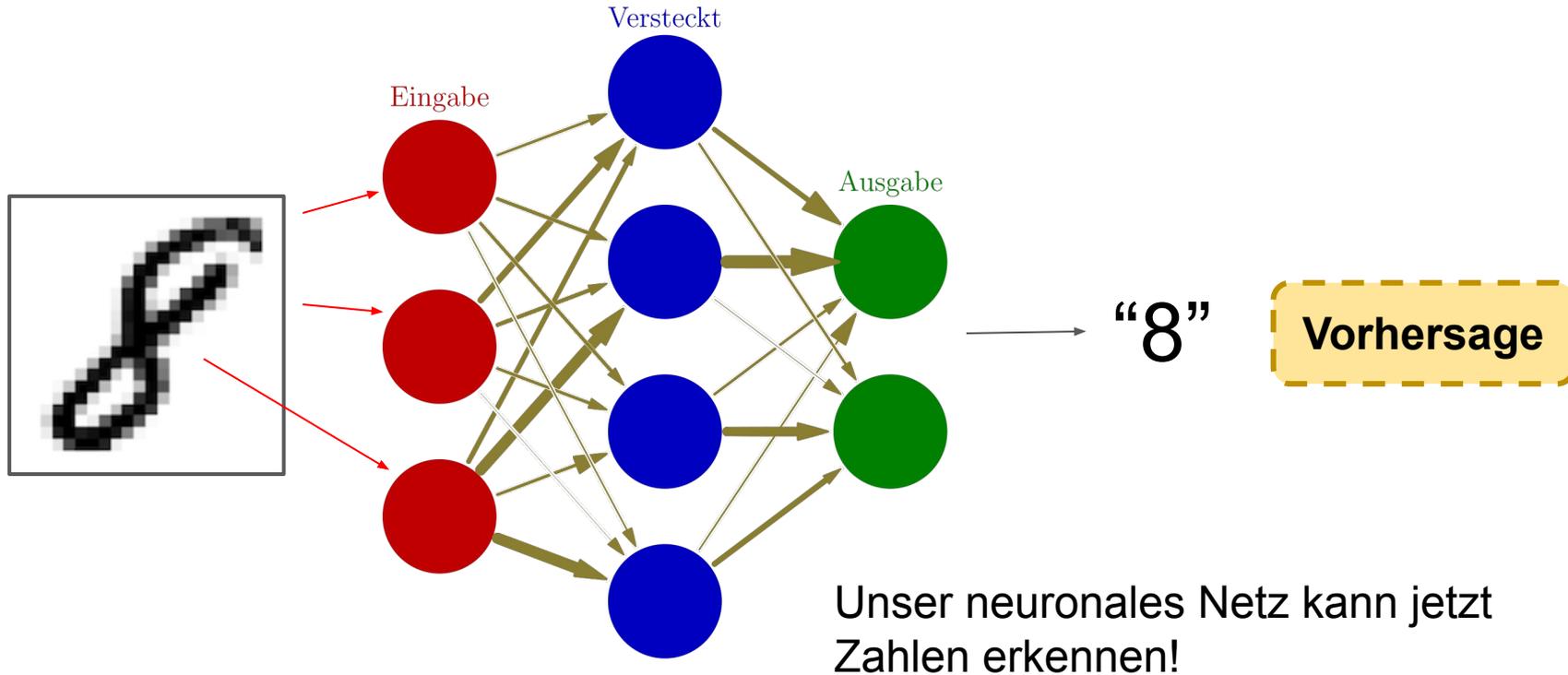
Training eines neuronalen Netzes



Ergebnis von Phase I: Training



Phase II: Auswertung



Zusammenfassung: Bausteine von ML Methoden

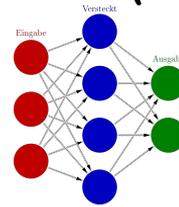
1. Trainingsdaten

Beispiele von Eingabe und Ausgabe



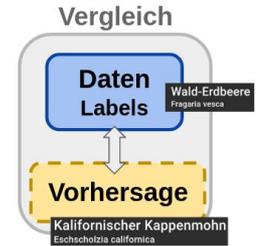
2. Modell

Funktion, die Eingaben auf mögliche Ausgaben abbildet



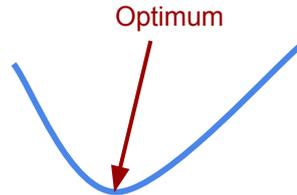
3. Bewertungsfunktion

Vergleich der Vorhersage und der bekannten Ausgabe



4. Lernmethodik (Lernalgorithmus)

i.d.R. ein Optimierungsverfahren

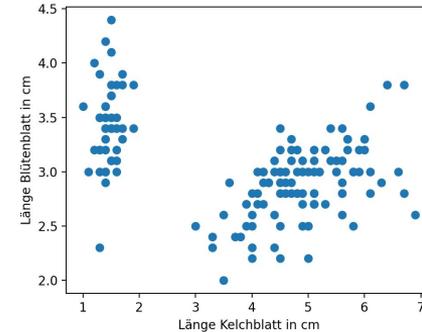


Überblick: Kategorien des maschinellen Lernens

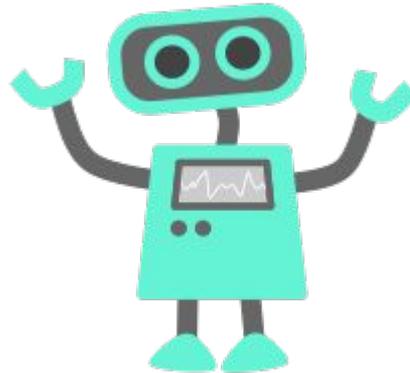
- **beaufsichtigtes Lernen:** Daten haben Labels



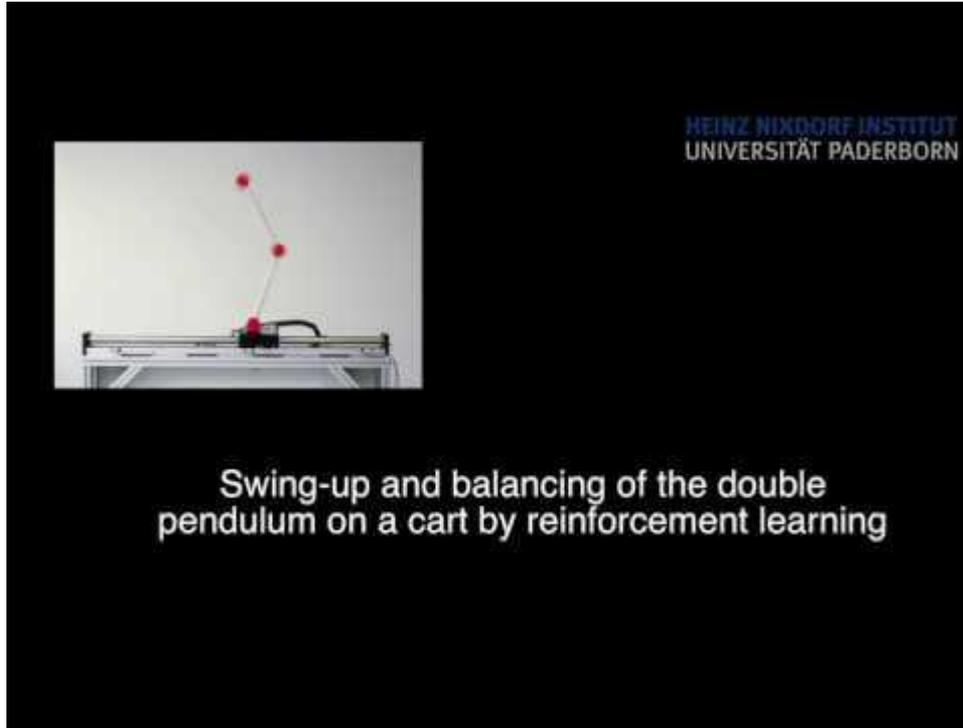
- **unbeaufsichtigtes Lernen:**
Struktur finden. Mustererkennung



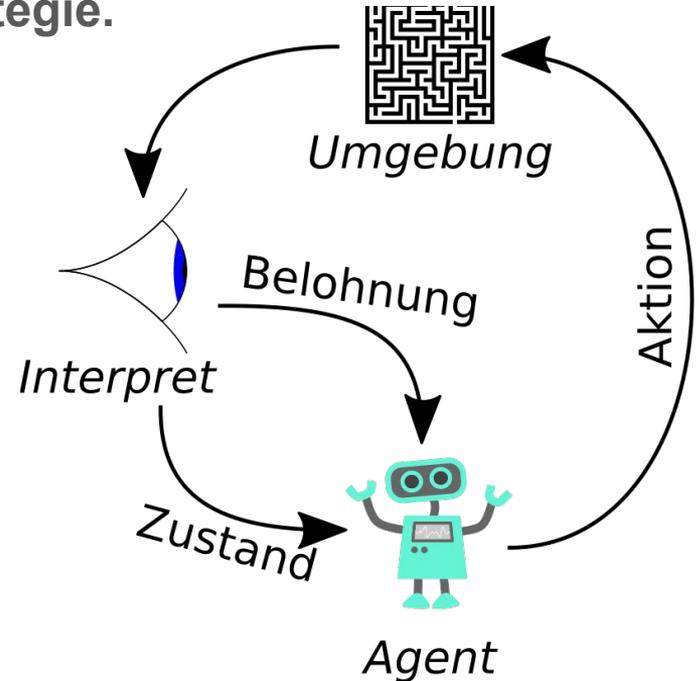
- **bestärkendes Lernen**



Bestärkendes Lernen (engl. Reinforcement Learning)



Ein Agent erlernt selbstständig **eine Strategie**.





Rückfragen?

4. Ethische und soziale Überlegungen; Herausforderungen

- **Daten bestimmen unser Zeitalter**
- **Was passiert mit meinen Daten?**

Sozialkreditsystem in China



Whatsapp, Facebook



Aktuelle Herausforderungen I: Wie fair sind Algorithmen?



Max Planck Institute for
Intelligent Systems



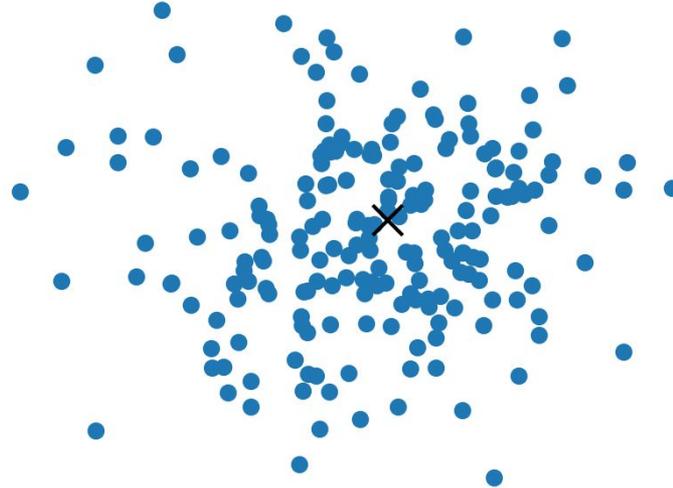
Prof. Hardt

Woraus ergibt sich eigentlich unsere Kreditwürdigkeit?

Was, wenn

- Alter
- Geschlecht
- Herkunft

eine Rolle spielen?



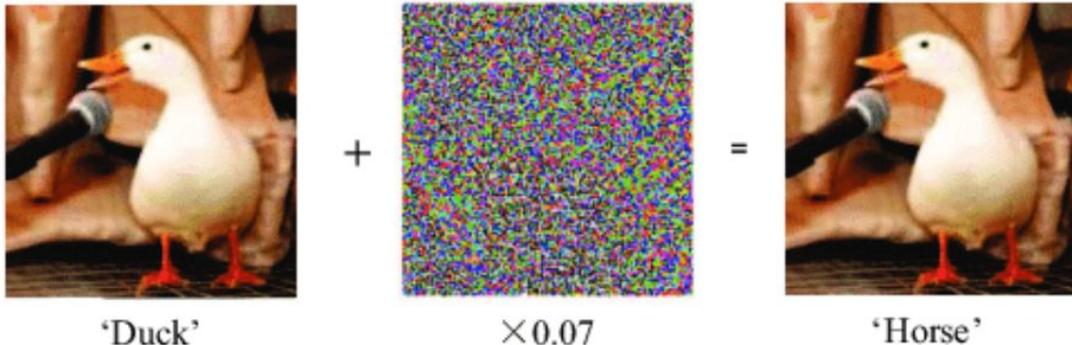
Achtung!

Selbst wenn geschützte Attribute nicht im Training berücksichtigt werden, können diese auf andere Aspekten der Daten Einfluss haben, und damit zur Entscheidung beitragen

Aktuelle Herausforderungen II: Robustheit gegen bösartige Attacken



Dr. Brendel



Merke:

Der Erkennungsprozess in einem neuronalen Netz unterscheidet sich grundlegend von dem eines Menschen!



Aktuelle Herausforderungen III: Erklärbarkeit von ML Methoden

Wie ist die Entscheidung entstanden?



Trainings-
daten

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



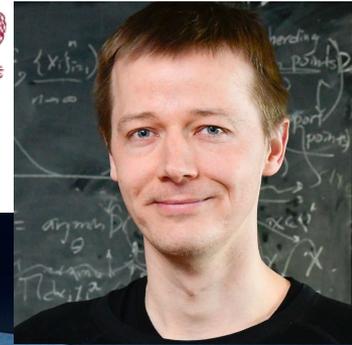
Prof. Akata

Testdaten



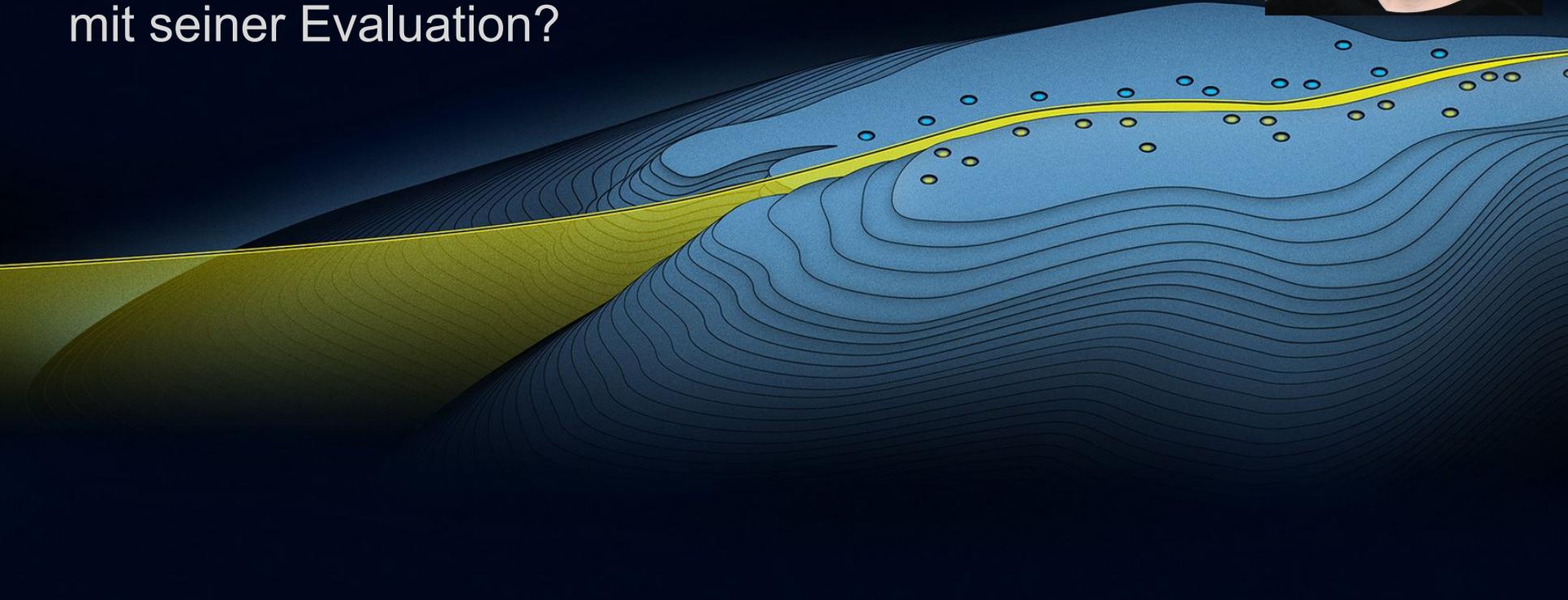
Aktuelle Herausforderungen IV: Quantifizierung von Unsicherheit

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Wie sicher ist sich der Algorithmus
mit seiner Evaluation?

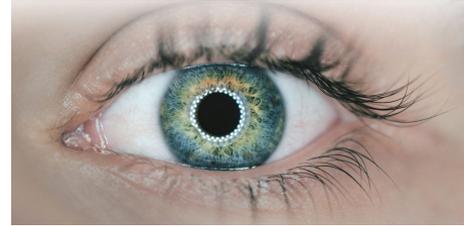
Prof. Hennig



Beiträge und Chancen des maschinellen Lernens

- **Medizin:** Krankheitserkennung, personalisierte Medizin

- Erkennung von Retinaveränderungen



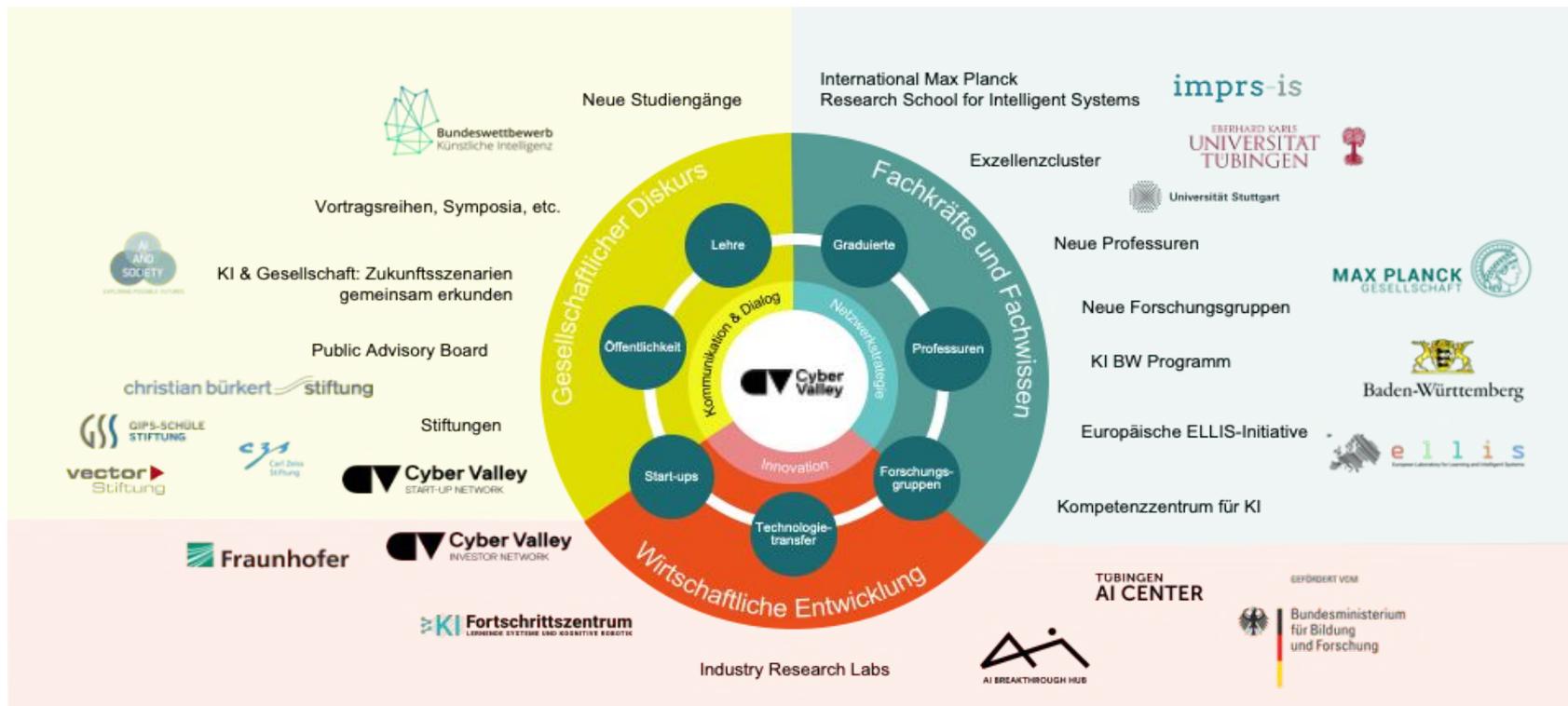
- **Automatische Übersetzung**  **DeepL**

- **Industrie:** Vorhersage von Fehlern in Bauteilen

- **Attributionsforschung:**
Wetterextreme dem Klimawandel
zuordnen



5. Was passiert am Forschungsstandort Tübingen?



DAIMLER



6. Fragen und Diskussion

- Wo soll KI auf keinen Fall eingesetzt werden?
- Wo sollte der “KI-Handwerkskoffer” verstärkt eingesetzt werden?
- Was wünsche ich mir von den KI-Forschenden in Tübingen?
- Welche Fragen sollten wir in weiteren Veranstaltungen beantworten?

Cyber Valley Public Engagement: Machen Sie mit!

Unser Ziel ist eine ethisch und sozial reflektierte Forschung auf dem Gebiet der künstlichen Intelligenz. Wir laden Sie deshalb ein zu unseren regelmäßig stattfindenden Dialogformaten für mehr Zugänglichkeit und Austausch zwischen Gesellschaft und KI-Forschenden.



www.cyber-valley.de/public-engagement

public.engagement@cyber-valley.de

Botschaften für den Nachhauseweg

zu **KI**: der Begriff ist irreführend → **“maschinelles Lernen”**

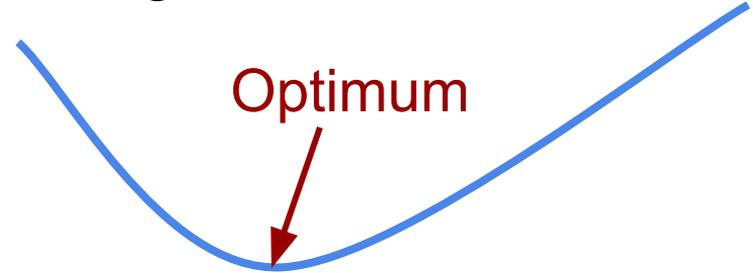
Funktionsweise:

- Beim maschinellen Lernen lernen Computer, **bestimmte** Probleme **anhand von Beispielen** zu lösen.

ML Methoden können manche Dinge gut, die Menschen nicht können und umgekehrt → **ML als Unterstützung des Menschen**

Mathematische Optimierung als Lernalgorithmus

Bewertungsfunktion zählt, wie oft wir falsch liegen (= **Trainingsfehler**)



Wir wollen, dass der Trainingsfehler **klein** wird.

⇒ “**Gradientenabstieg**”

