

„Maschinelles Lernen“

(Und warum das vielleicht gruseliger ist, als es klingt)

Jonas Betzendahl

@jbetzend



Was mache ich?



Früher:

B. Sc. „Kognitive Informatik“

M. Sc. „Intelligente Systeme“

Technische Fakultät, Universität Bielefeld

Heute:

Doktorand der Informatik

Technische Fakultät, FAU Erlangen

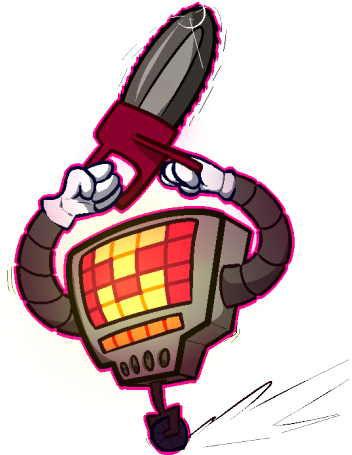
Small Talk in Intelligent Systems

Die häufigste Frage an meinen Studiengang:

Small Talk in Intelligent Systems

Die häufigste Frage an meinen Studiengang:

„Na, wie lange dauert
es noch bis zur
Roboterapokalypse?“



amazon

The Amazon logo, featuring the word "amazon" in a bold, dark gray sans-serif font. Below the text is a thick, orange curved arrow that starts under the 'a' and points towards the 'n', with the arrowhead pointing to the right.

amazon





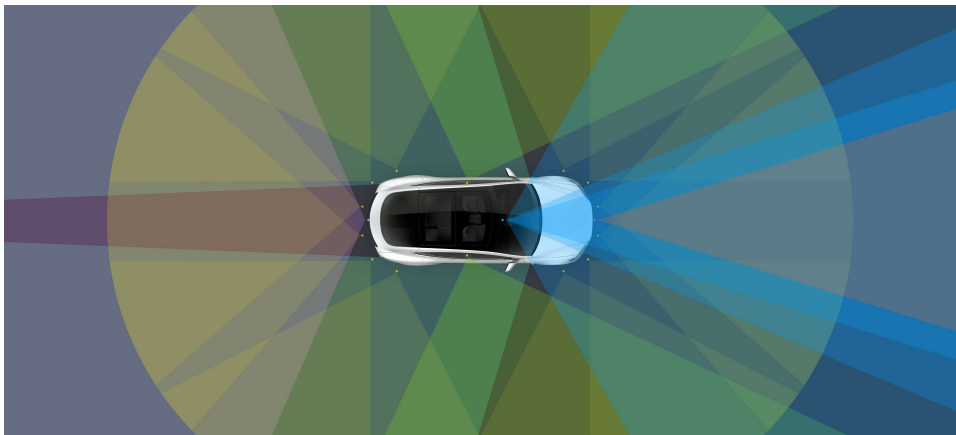
...zumindest habe ich bisher so immer meine Slams angefangen.

...zumindest habe ich bisher so immer meine Slams angefangen.

Wir müssen reden!

Die Errungenschaften von Maschinellem Lernen

Maschinelles Lernen ist prinzipiell sehr mächtig und nützlich...

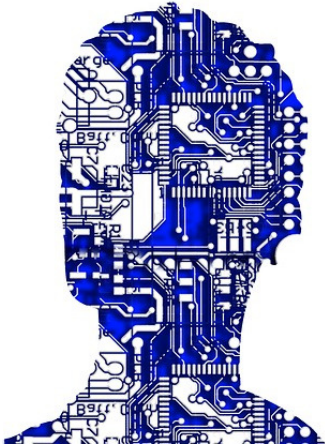


"Nachdem die Menschheit Jahrtausende damit verbracht hat, ihre Taktiken zu verbessern, erzählen uns die Computer, dass wir komplett daneben liegen."

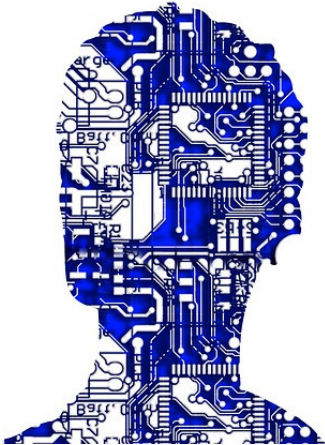
-- Ke Jie



Wie funktioniert
Maschinelles Lernen?

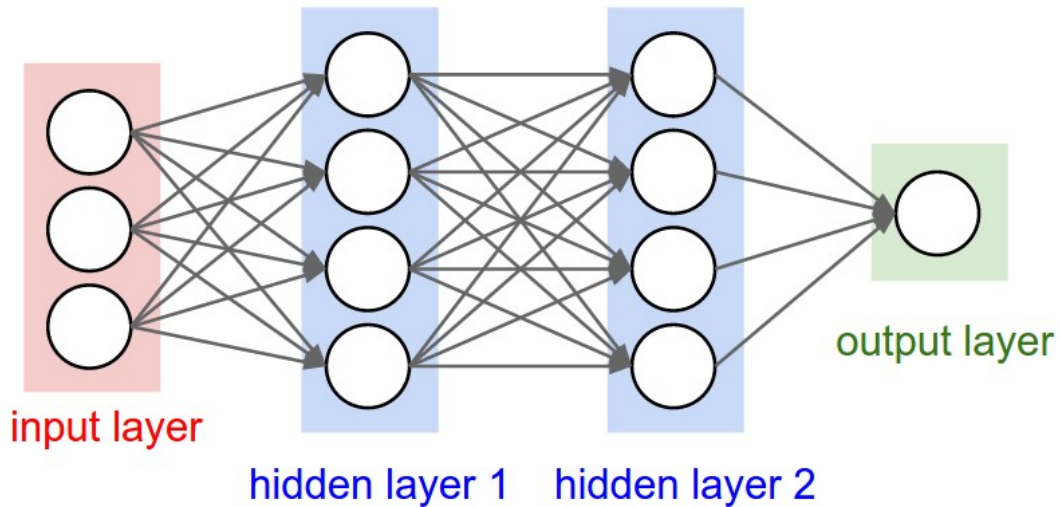


Maschinelles Lernen simuliert einen Vorgang nicht unähnlich dem im menschlichen Gehirn selbst.



Maschinelles Lernen simuliert einen Vorgang nicht unähnlich dem im menschlichen Gehirn selbst.

Ein (künstliches) *neuronales Netz* wird simuliert und trainiert mit *Testdaten*, bis es akzeptable Leistungen bringt.





FIRE TOGETHER

=

WIRE TOGETHER

Die Fehler von Maschinellen Lernen

"What is the cost of a
train ticket from Sydney
to Brisbane"
tap to edit

The answer is about
28.8 quadrillion
kilometer US dollars
squared.

Input interpretation

The Train (movie)	production budget
The Train (movie)	total US box office receipts
distance	from Sydney, New South Wales
	to Brisbane, Queensland

Result

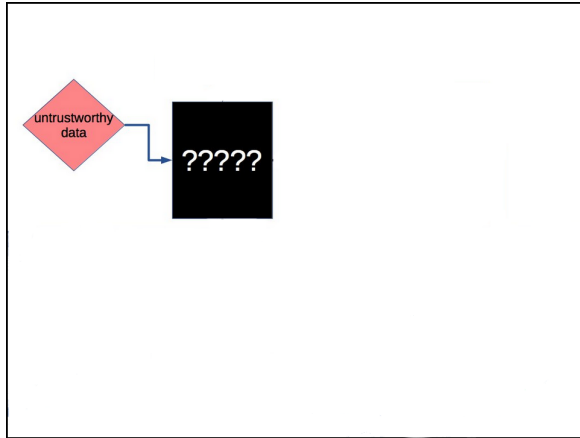
28.8488 quadrillion km \$² (kilometer US dollars squared)

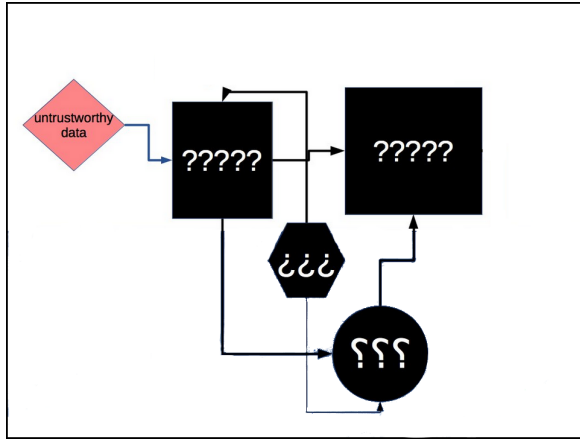
WolframAlpha

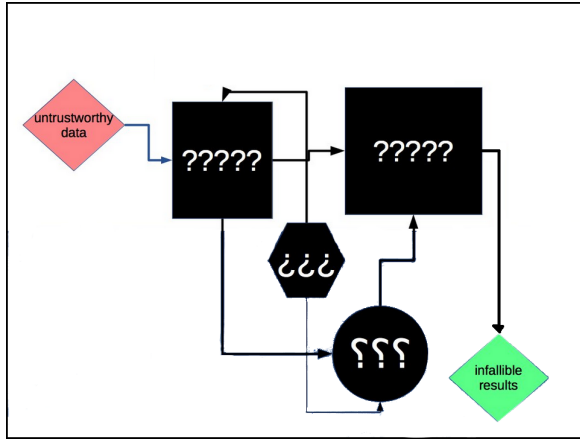
Maschinelles Lernen ist
(genau wie menschliches) alles andere als
fehlerfrei!

Computer haben noch einen langen Weg,
bis sie z.B. Sprache so gut verstehen wie
Menschen.

untrustworthy
data







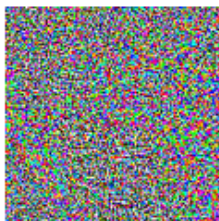


x

“panda”

57.7% confidence

$+ .007 \times$



$\text{sign}(\nabla_x J(\theta, x, y))$

“nematode”

8.2% confidence

$=$



$x +$

$\epsilon \text{sign}(\nabla_x J(\theta, x, y))$

“gibbon”

99.3 % confidence

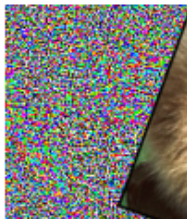


x

“panda”

57.7% confidence

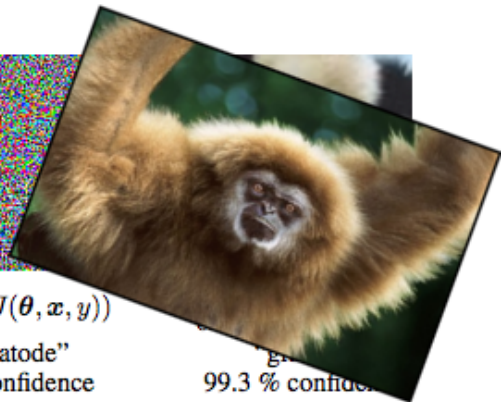
$+ .007 \times$



$\text{sign}(\nabla_x J(\theta, x, y))$

“nematode”

8.2% confidence



“gorilla”

99.3 % confidence

Wat lernt misch datt?

Maschinelles Lernen liefert oft
nur *Ergebnisse*, keine *Begründungen*.

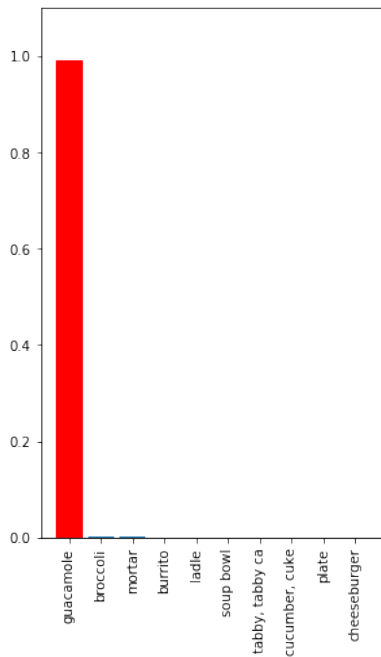
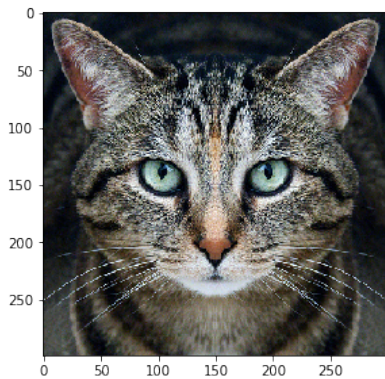
Wat lernt misch datt?

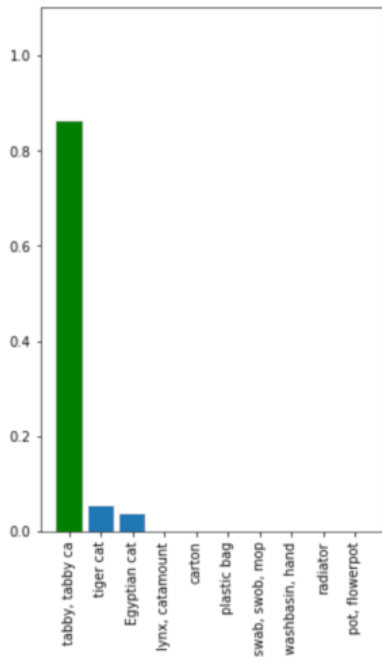
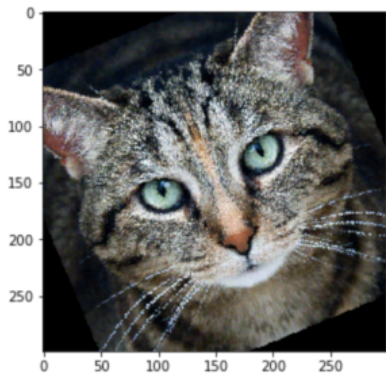
Maschinelles Lernen liefert oft
nur *Ergebnisse*, keine *Begründungen*.

Außerdem ist das Ergebnis höchstens
so allgemein wie die Trainingsdaten.

„Adversarial Objects“ (Feindliche Objekte)

(*Subs., plural*) Objekte, die für das menschliche Auge herkömmlich erscheinen, aber für den Computer radikal anders aussehen.





Feindliche 3D-gedruckte Schildkröte:



Ein paar offensichtliche Probleme:

Ein paar offensichtliche Probleme:

- Gratis T-Shirts die am Flughafen als Waffen erkannt werden

Ein paar offensichtliche Probleme:

- Gratis T-Shirts die am Flughafen als Waffen erkannt werden
- Plakate neben der Autobahn die als Stoppschilder erkannt werden

Ein paar offensichtliche Probleme:

- Gratis T-Shirts die am Flughafen als Waffen erkannt werden
- Plakate neben der Autobahn die als Stoppschilder erkannt werden
- Lars

Ein paar offensichtliche Probleme:

- Gratis T-Shirts die am Flughafen als Waffen erkannt werden
- Plakate neben der Autobahn die als Stoppschilder erkannt werden
- Lars
- ...



Ein paar offensichtliche Probleme:

- Gratis T-Shirts die am Flughafen als Waffen erkannt werden
- Plakate neben der Autobahn die als Stoppschilder erkannt werden
- LARs
- ...



Wat lernt misch datt?

Maschinelles Lernen ist nicht unfehlbar und darf in kritischen Systemen nie unüberprüft wichtige Entscheidungen treffen.

Zusammenfassung

Die guten Nachrichten:

Zusammenfassung

Die guten Nachrichten:

- Roboterapokalypse: Erstmal unwahrscheinlich

Zusammenfassung

Die guten Nachrichten:

- Roboterapokalypse: Erstmal unwahrscheinlich
- Maschinelles Lernen kann uns das Leben sehr vereinfachen

Aber:

Zusammenfassung

Die guten Nachrichten:

- Roboterapokalypse: Erstmal unwahrscheinlich
- Maschinelles Lernen kann uns das Leben sehr vereinfachen

Aber:

- Maschinelles Lernen ist nicht fehlerfrei

Zusammenfassung

Die guten Nachrichten:

- Roboterapokalypse: Erstmal unwahrscheinlich
- Maschinelles Lernen kann uns das Leben sehr vereinfachen

Aber:

- Maschinelles Lernen ist nicht fehlerfrei
- Meist nur ein Ergebnis, keine Begründung

Zusammenfassung

Die guten Nachrichten:

- Roboterapokalypse: Erstmal unwahrscheinlich
- Maschinelles Lernen kann uns das Leben sehr vereinfachen

Aber:

- Maschinelles Lernen ist nicht fehlerfrei
- Meist nur ein Ergebnis, keine Begründung
- Vorsicht vor Vorurteilen in der Datengrundlage

Zusammenfassung

Die guten Nachrichten:

- Roboterapokalypse: Erstmal unwahrscheinlich
- Maschinelles Lernen kann uns das Leben sehr vereinfachen

Aber:

- Maschinelles Lernen ist nicht fehlerfrei
- Meist nur ein Ergebnis, keine Begründung
- Vorsicht vor Vorurteilen in der Datengrundlage
- Kritische Entscheidungen brauchen Menschen in der Schleife