

Künstliche Intelligenz mit lernenden Maschinen:

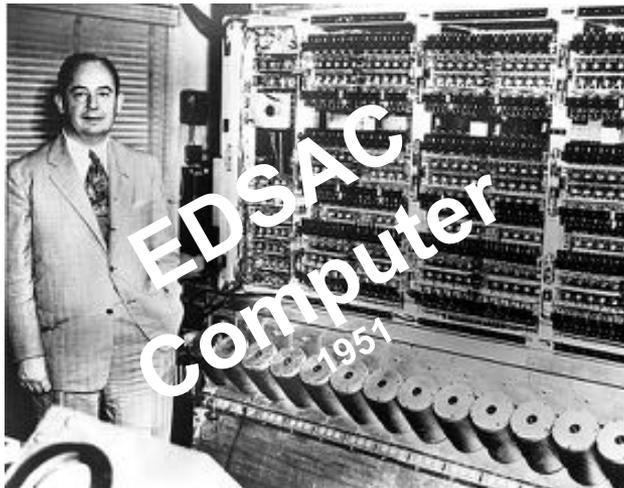
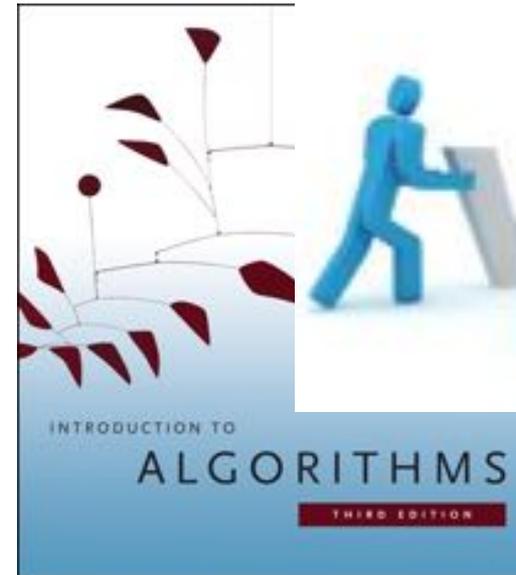
Wo stehen wir mit Big Data und Digitalisierung?

Joachim M. Buhmann

Institut für Maschinelles Lernen, ETH Zürich



Die Welt, in der wir leben!



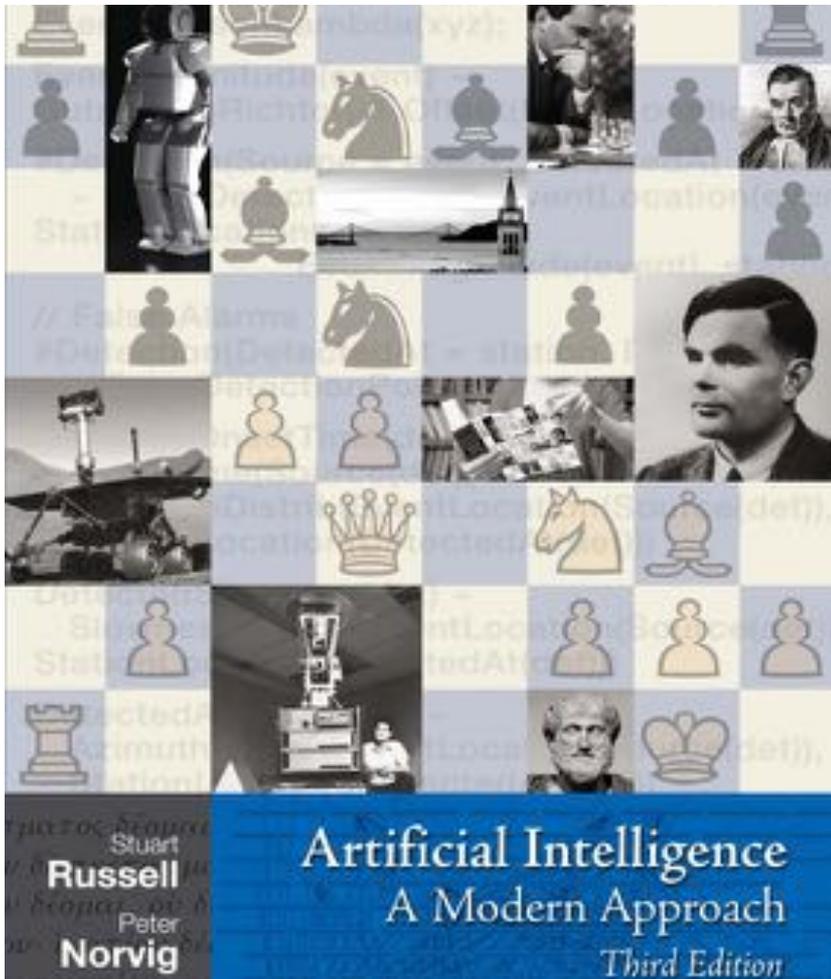
Die Welt, in der wir leben!



Werte aus Daten zu schöpfen
erfordert Intelligenz!
(zumindest “künstliche”)

Google
Data Center
2016

Was ist künstliche Intelligenz?



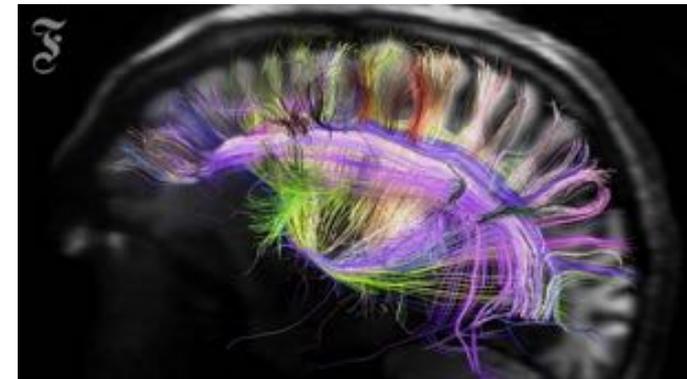
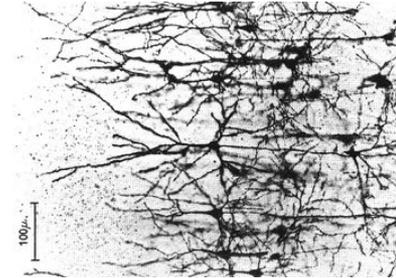
„Intelligence is mind implemented by any patternable kind of matter.“

Newell & Simon; Dartmouth Conference 1956

- Automatisierung intelligenten Verhaltens (Wikipedia)
- **Historie**
 - Kybernetik (Wiener) 1945-1960
 - Künstliche Intelligenz, (McCarthy, et al. 60s & 70s)
 - Neuronale Netze (1985)
 - Machinelles Lernen (Modern AI, 1990-today)
 - Deep learning (2010-today)

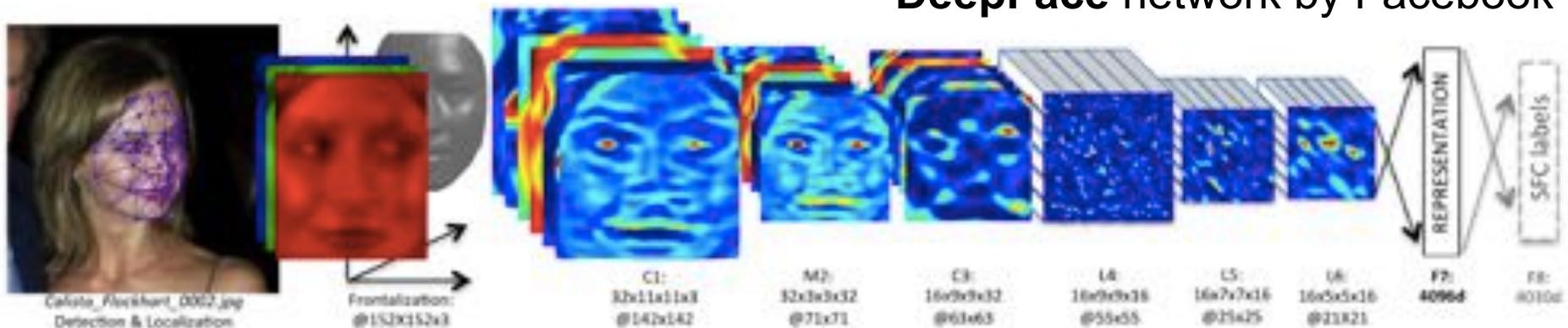
Lernende Maschinen beherrschen algorithmische Induktion

- **Biologische Nervennetze** sind adaptiv und lernfähig.
- **Künstliche neuronale Netze** bilden diese Lernfähigkeit nach.



Nervenzell-Netze mit Hirnscans sichtbar gemacht. © VAN WEDEEN

DeepFace network by Facebook



Disruption durch Digitalisierung

- Taxifahrer und **UBER**
- Hoteliers und *airbnb*
- Einzelhandel und **amazon**
- Bankiers und  **bitcoin**
ACCEPTED HERE
- **Ärzte** Ihres Vertrauens und...

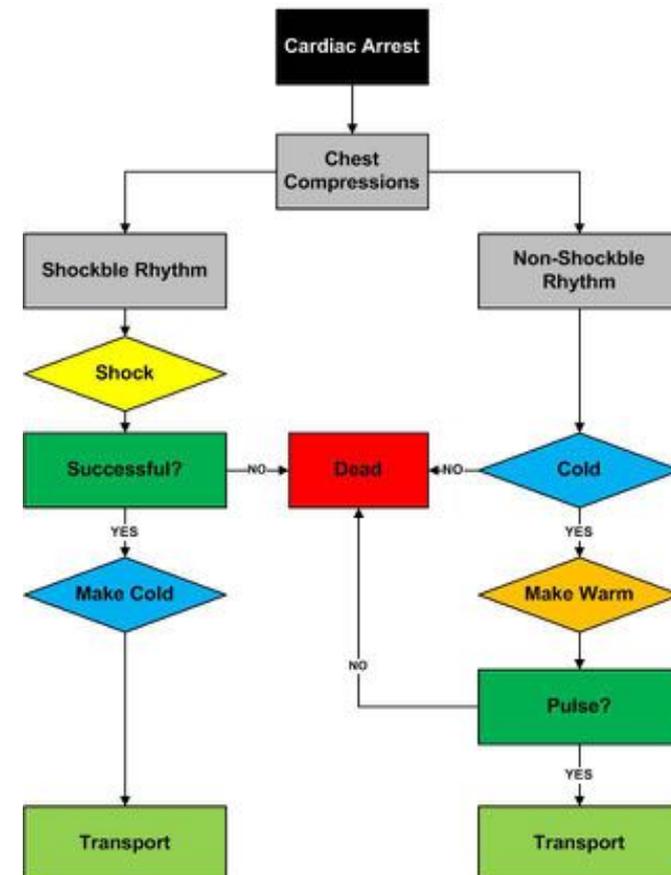


The Algorithm: Idiom of Modern Science

(Bernard Chazelle)

- Ein **Algorithmus** ist eine wohldefinierte **Rechenprozedur**, die **Eingabewerte** liest und **Ausgabewerte** berechnet. (CLRS)
- Analyse von Algorithmen**
 - ✓ Laufzeit, Speicherbedarf
 - ✗ **Robustheit, Generalisierung!**

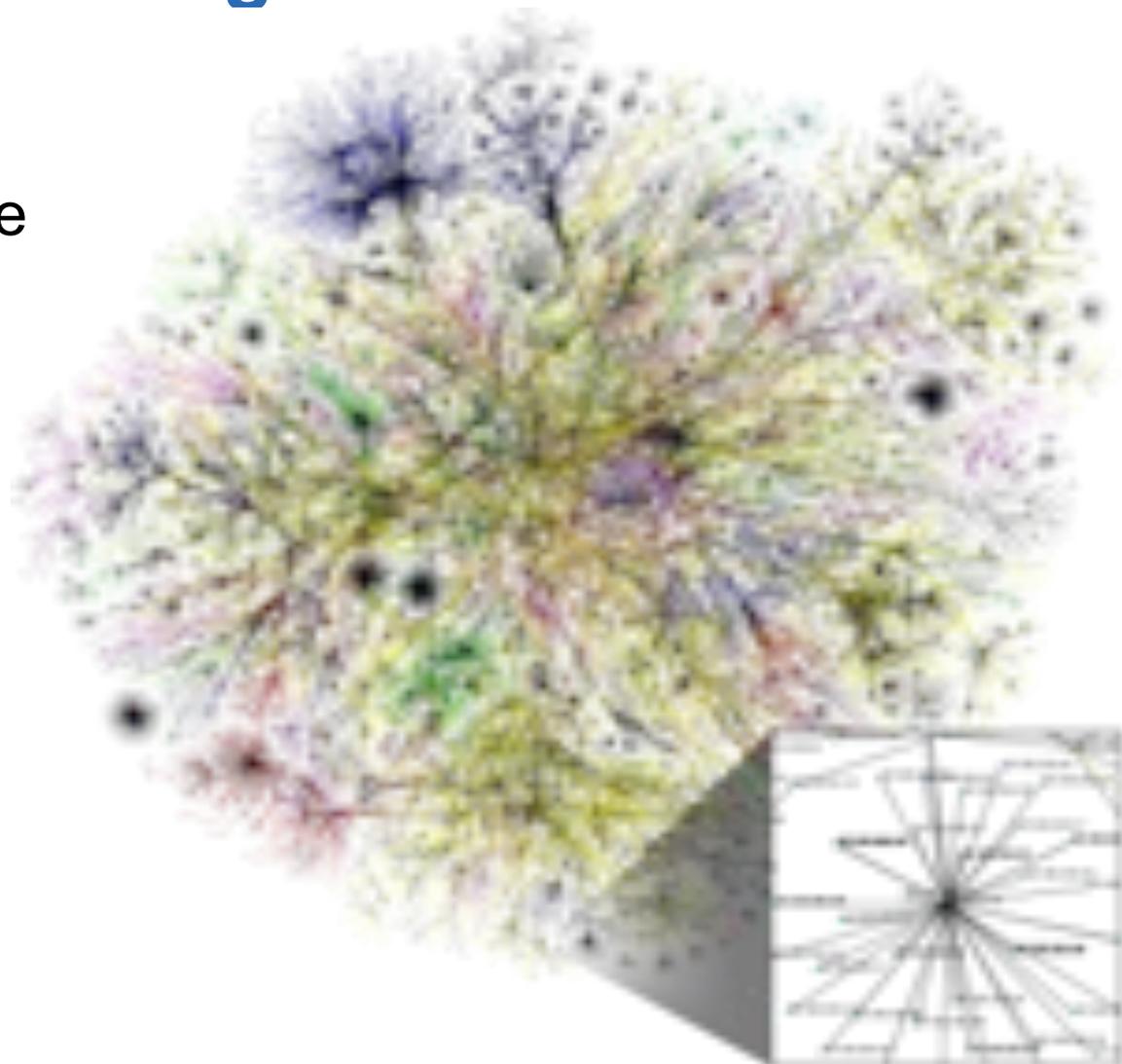
2015 ACLS CARDIAC ARREST ALGORITHM



Lernende Algorithmen „erforschen“
die Komplexität der Wirklichkeit

Network Science & Big Data

- Biologische Netzwerke
X-omics
- Kommunikations-
netzwerke
- Soziale Netze =>
*Computational
Humanities*
- **Finanznetzwerke**
- ...



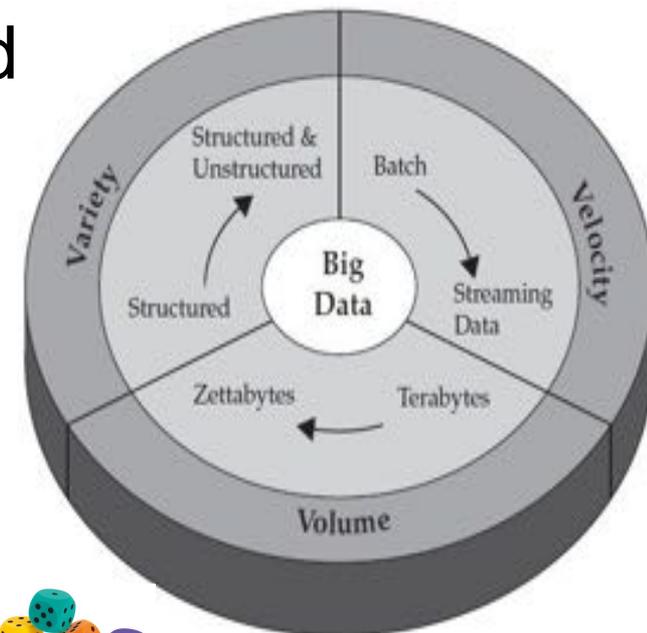
A blue-tinted photograph of a large, classical-style building with a prominent dome and arched windows, likely a part of the ETH Zurich campus, serving as a background for the top of the slide.

1. Data Science

die Wissenschaft des 21. Jahrhunderts

4 BIG V's Big Data Charakteristiken

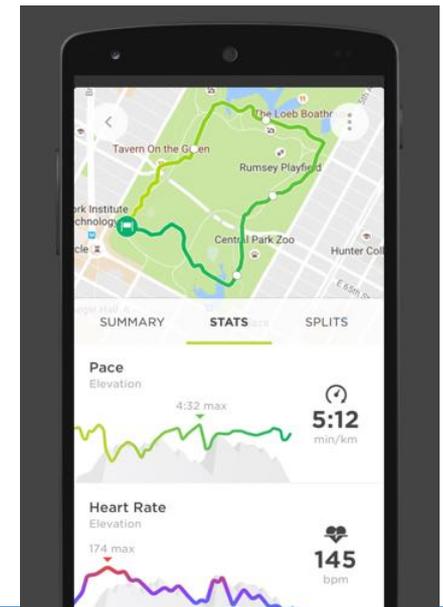
- **Volume** – Quantitatives Mehr an Daten
- **Velocity** – Datengenerierung und Nutzung in Echtzeit
- **Variety** – Unstrukturierte Daten, Beobachtungsdaten, Rohdaten (aus unterschiedlichen Quellen)
- **Veracity** – Unsicherheit in Daten



© wordpress.com

Big Data und Privacy

- Nicht personenbezogene Daten: CERN, Klimadaten, Prozess- und Sensordaten, Netz-Optimierung
- **Hochsensitive Daten:** medizinische Daten, Genomdaten, Finanzdaten, ...
- **“Aktivitätsdaten”:** Web Aktivität, Ortsdaten, Kauf- / Konsumverhalten, soziale Interaktionen ...



Google Websuche: New Kid, Big Data Veteran

- Websuche: existierende Webinhalte (das Web als **Big Data Repository**)
- **Intelligente Datenanalyse** (z.B. Ranking) bringt Nutzer und Inhalte zusammen
- **Nutzungsdaten**: Optimierung, z.B. Platzierung von Werbung



- Sammlung & Verwendung (persönlicher) Daten
- Gerechte Verteilung des Mehrwerts?

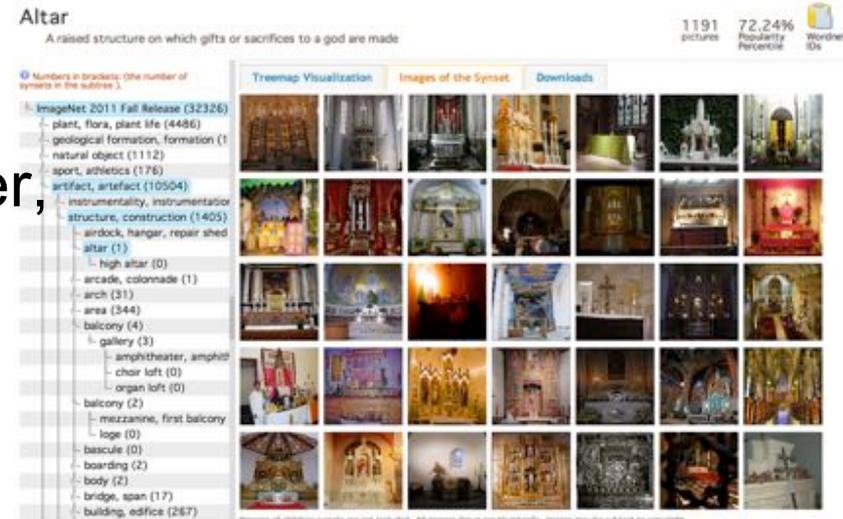


2. KI & ML

retten uns vor der Programmierkrise (!?)

Bildererkennung, Gesichtserkennung

- **Grosse Bilddatenmengen** als Trainingsdaten (2004: 1000 Bilder, heute: 10^7 Bilder aus **ImageNet**)
- Objektklassifikation mit einigen 10'000 visuellen Kategorien
- Anwendungsszenarien: Gesichtserkennung, Überwachung, aber auch: autonome Fahrzeuge
- **DeepMind: Solve Intelligence. Use it to make the world a better place.**



“Deep Network” Halluzinationen

(von Sebastian Nowozin)



Automatische Übersetzung

- **Statistical machine translation** mit Textbeispielen aus dem Internet, z.B. EU/UN Dokumenten.
- **Daten getriebener Ansatz** anstatt linguistische Analyse mit Wörterbuch/Grammatik. *“It has been said before!”*



Google Translate interface showing a German text box on the left and an English translation box on the right. The German text is: "Ohne Sinnlichkeit würde uns kein Gegenstand gegeben, und ohne Verstand keiner gedacht werden. Gedanken ohne Inhalt sind leer, Anschauungen ohne Begriffe sind blind. (Immanuel Kant: Kritik der reinen Vernunft.)". The English translation is: "Without sensibility no object would given us, and not be thought without understanding. Thoughts without content are empty, intuitions without concepts are blind. (Immanuel Kant: Critique of Pure Reason.)".

Jeopardy – IBM *Watson's* Claim to Fame:

- **IBM's Watson**, der Computer mit Sprachverständnis, **schlägt 2011 die besten menschlichen Spieler in der US TV show Jeopardy!**
- **Watson's Erfolgsrate** für Lungenkrebsdiagnose übertrifft mit 90% deutlich die 50% Rate von Onkologen.

[Samuel Nessbaum, Wellpoint]





Google DeepMind Challenge Match

8 - 15 March 2016



Relevanzvorhersage

Relevanz ist ein Schlüsselproblem

- **Amazon:** Welches Produkt ist relevant für mich?
- **Facebook:** Welches Posting ist relevant für mich?
- **Web Marketing:** Welche Werbung könnte mich interessieren?

Customers Who Bought This Item Also Bought



Customers Who Bought This Item Also Bought

Scala for the Impatient
Cay S. Horstmann
★★★★★ (50)
Paperback
\$30.57 ✓Prime

Scala in Depth
Joshua D. Suereth
★★★★★ (14)
Paperback
\$28.56 ✓Prime

Scala Cookbook: Recipes for Object-Oriented ...
Alvin Alexander
★★★★★ (16)
Paperback
\$37.03 ✓Prime

Ads You May Be Interested In



Are You A Business Owner?

Apply Now to the Worldwide Who's Who Registry of Distinguished Individuals. >



Publish with Springer

Have a textbook aimed primarily at undergraduates? Submit to Springer now. >



#1 Ticketing System

Heaps of customer queries giving headaches? Try this ticketing tool! >

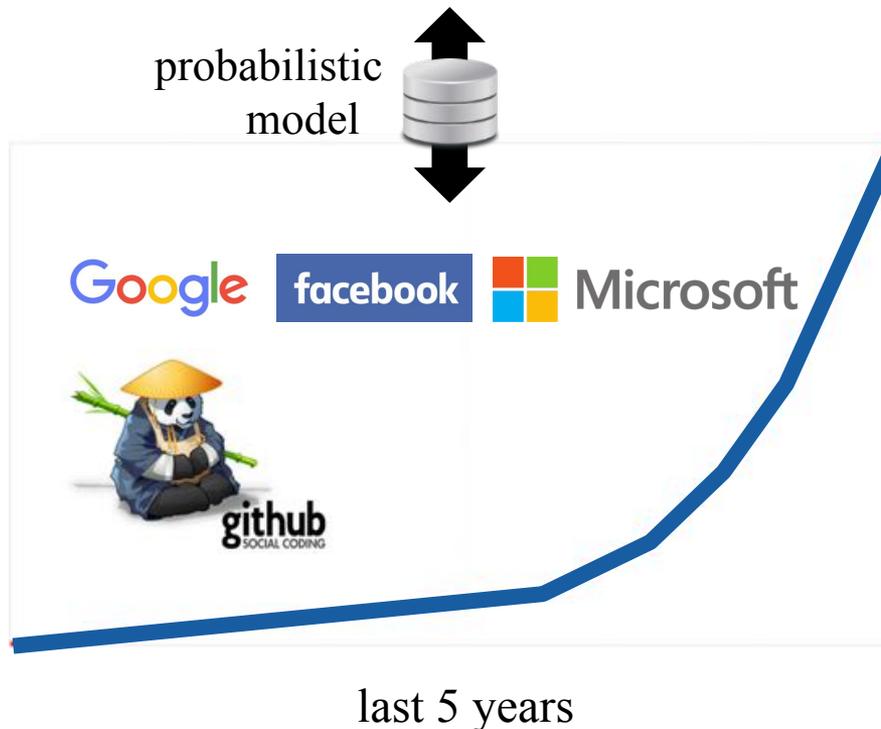
The stories that appear in your News Feed are influenced by your connections and activity on Facebook.

Learning from large datasets of programs ("Big Code")



number of
repositories

[Big Code]



~20 million repositories

Billions of lines of code

High quality, tested,
maintained programs

Computer lernen, Programme zu “entschleiern”

Verschleiertes Programm

```
function FZ(e, t) { var n =  
[]; var r = e.length; var i =  
0; for (; i < r; i += t) if (i  
+ t < r)  
n.push(e.substring(i, i + t));  
else n.push(e.substring(i,  
r)); return n;  
}
```



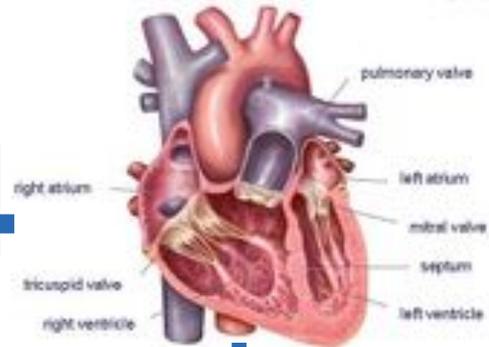
JSNice.org

Lesbares Programm

```
function chunkData(str, step) {  
  var colNames = [];  
  var len = str.length;  
  var i = 0;  
  for (; i < len; i += step)  
    if (i + step < len)  
      colNames.push(str.substring(i, i +  
step));  
    else  
      colNames.push(str.substring(i, len));  
  return colNames;  
}
```

3. ML 4

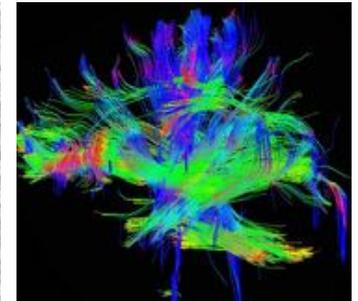
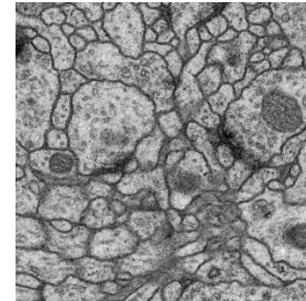
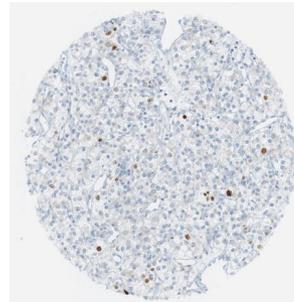
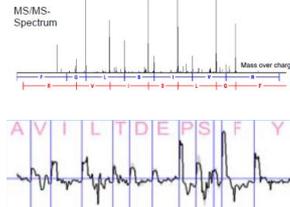
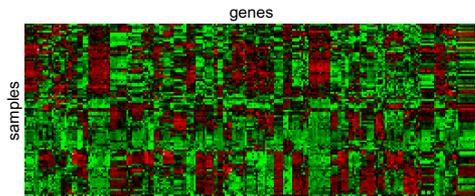
Machine Learning for the Life Sciences



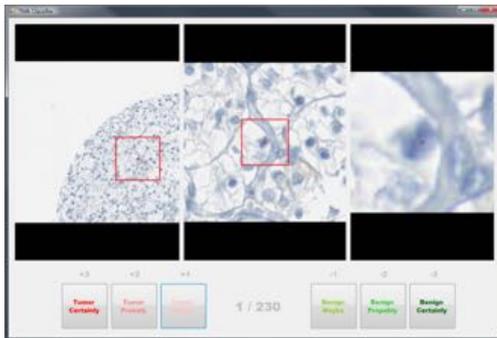
Big Data in der Medizin Worüber reden wir eigentlich?

Heterogene Datenquellen

Patientendaten



Diagnose-, Prognose- und Therapiedaten der Mediziner



Bestimmung	Einheit	0.000	0.000	10.00	10.00	10.00	10.00
WBC	10 ⁹ /L	10	10	10	10	10	10
Hemoglobin	g/L	12	12	12	12	12	12
Hematokrit	L/L	37	37	37	37	37	37
Erythrozyten	10 ¹² /L	4.5	4.5	4.5	4.5	4.5	4.5
SDH	U/L	27	27	27	27	27	27
MDV	U/L	30	30	30	30	30	30
LDH	U/L	30	30	30	30	30	30
Leukozyten	10 ⁹ /L	8	8	8	8	8	8
Thrombozyten	10 ⁹ /L	180	180	180	180	180	180
Neutrophil	%	78	78	78	78	78	78
Lymphozyt	%	18	18	18	18	18	18
Monozyt	%	2	2	2	2	2	2
Basophyl	%	0	0	0	0	0	0
Platylet	10 ⁹ /L	150	150	150	150	150	150
CRP	mg/L	10	10	10	10	10	10
Urea	mmol/L	5.5	5.5	5.5	5.5	5.5	5.5
Kreatinin	mmol/L	0.1	0.1	0.1	0.1	0.1	0.1
Glukose	mmol/L	5.5	5.5	5.5	5.5	5.5	5.5
Alkalische Phosphatase	U/L	120	120	120	120	120	120
Aspartataminotransferase	U/L	15	15	15	15	15	15
Alaninaminotransferase	U/L	15	15	15	15	15	15
Gamma-GT	U/L	15	15	15	15	15	15
Laktatdehydrogenase	U/L	150	150	150	150	150	150
Cholesterin	mmol/L	5.0	5.0	5.0	5.0	5.0	5.0
HDL-Cholesterin	mmol/L	1.0	1.0	1.0	1.0	1.0	1.0
LDL-Cholesterin	mmol/L	3.5	3.5	3.5	3.5	3.5	3.5
Triglyceride	mmol/L	1.5	1.5	1.5	1.5	1.5	1.5
Urea-Cholesterin	mmol/L	2.0	2.0	2.0	2.0	2.0	2.0
Urea	mmol/L	5.5	5.5	5.5	5.5	5.5	5.5
Cholesterin	mmol/L	5.0	5.0	5.0	5.0	5.0	5.0



Quantified self



Big Data Management & Analyse

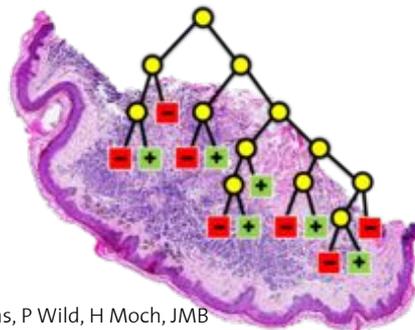
Datenmanagement

Speicherung, Informationssystem, Benutzerschnittstellen, *cloud computing*

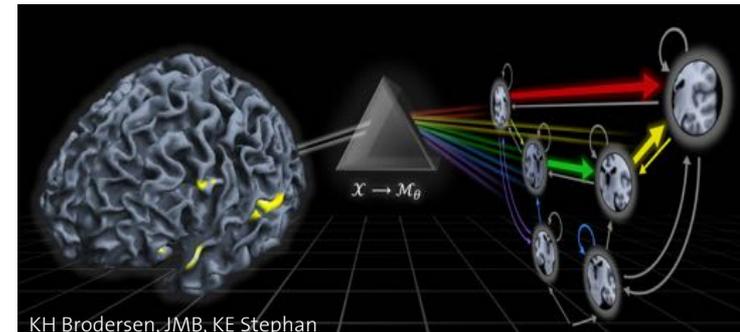


Datenanalytik & Maschinelles Lernen

Knowledge Discovery & Data Mining

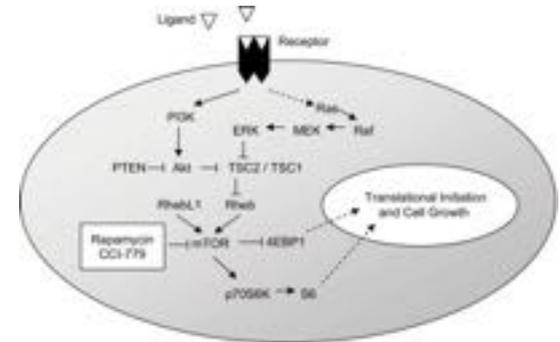
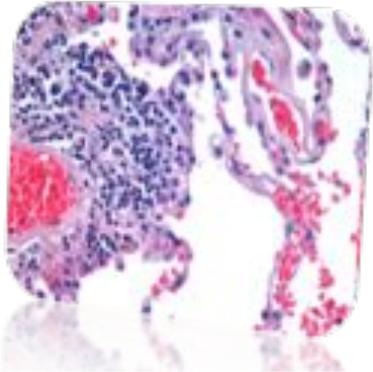


T Fuchs, P Wild, H Moch, JMB



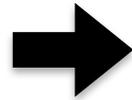
KH Brodersen, JMB, KE Stephan

Informationstechnologie für Personalisierte Medizin

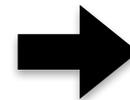


Activation of the mTOR Signaling Pathway in Renal Clear Cell Carcinoma. Robb et al., J Urology 177:346 (2007)

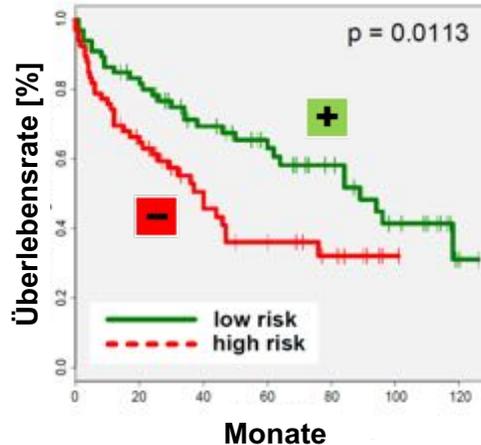
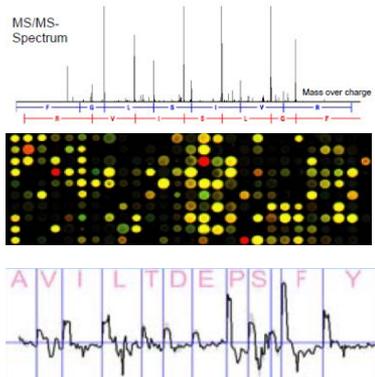
meine Daten



meine Information



unser Wissen

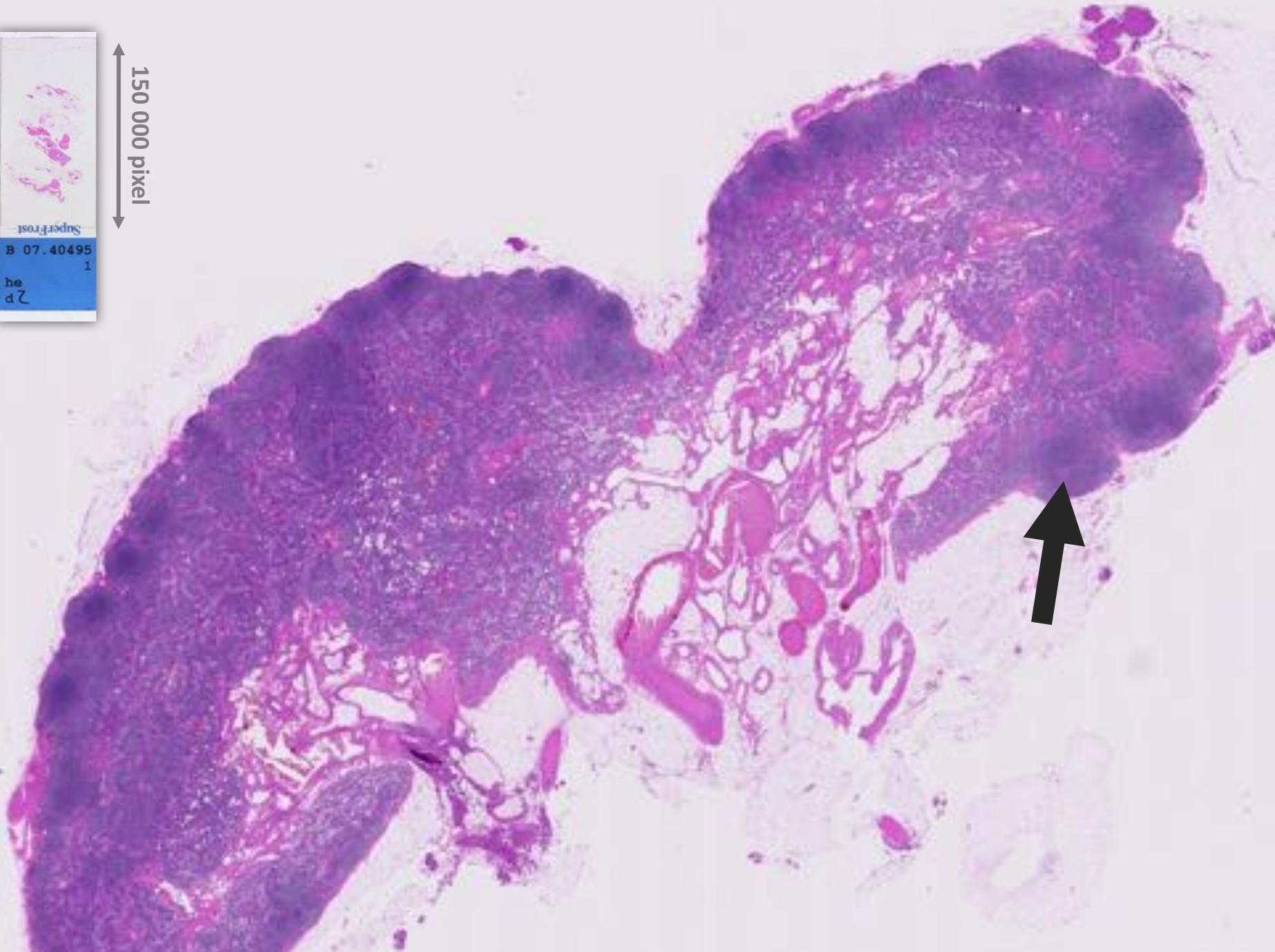


mein/unser Nutzen

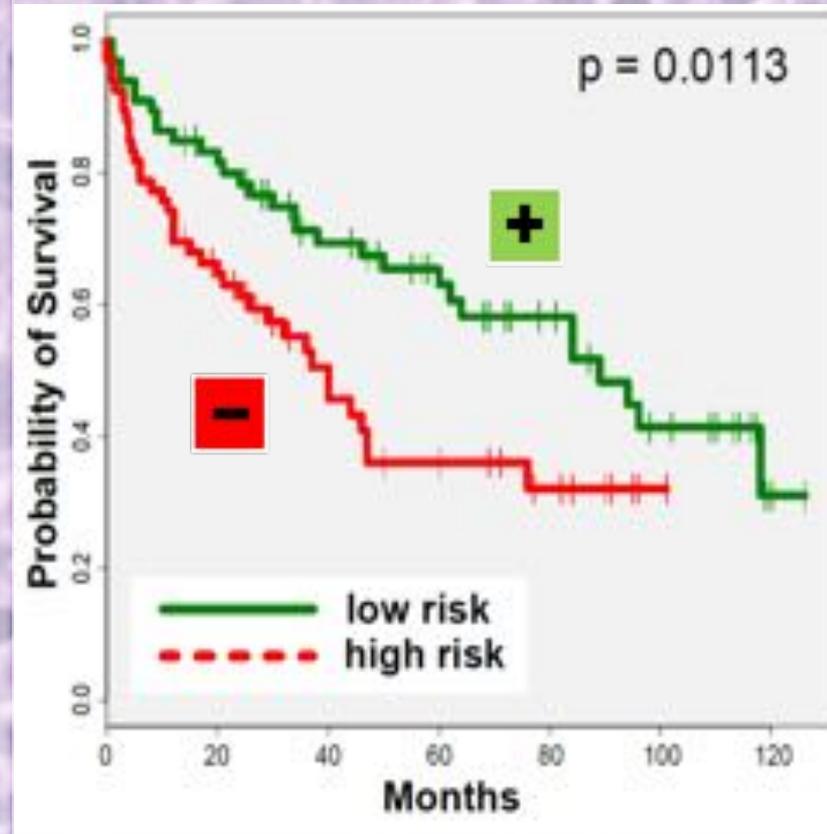
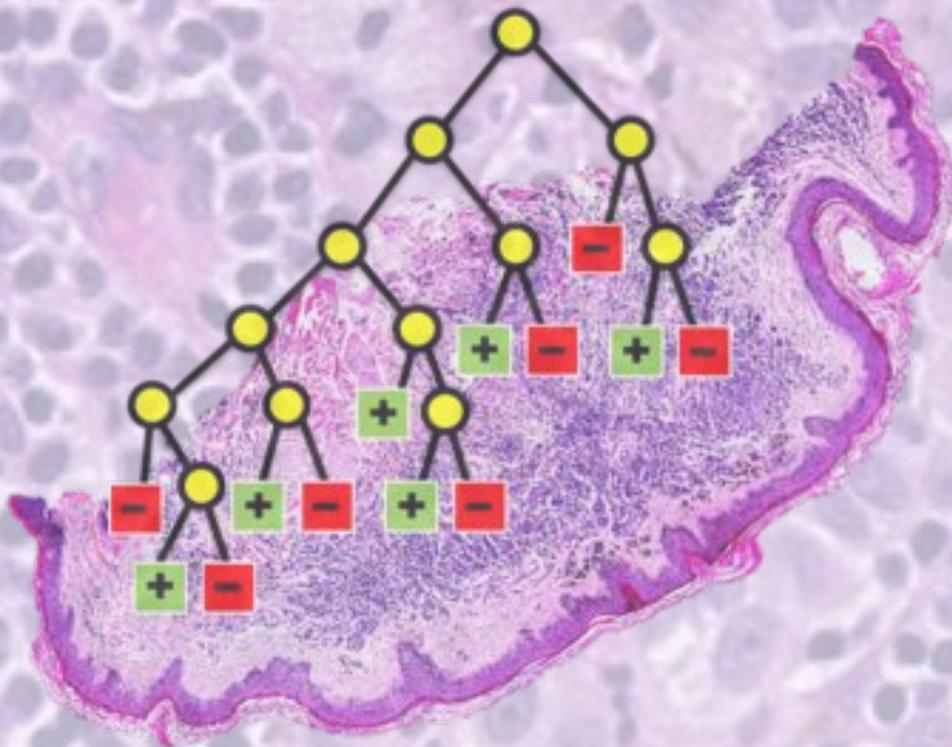




150 000 pixel



Überlebenswahrscheinlichkeit: Wo ist die Information verborgen?



Renal Cell Carcinoma (Nierenkrebs)

Clear cell Renal Cell Carcinoma ist einer der zehn häufigsten Krebsarten in westlichen Gesellschaften.

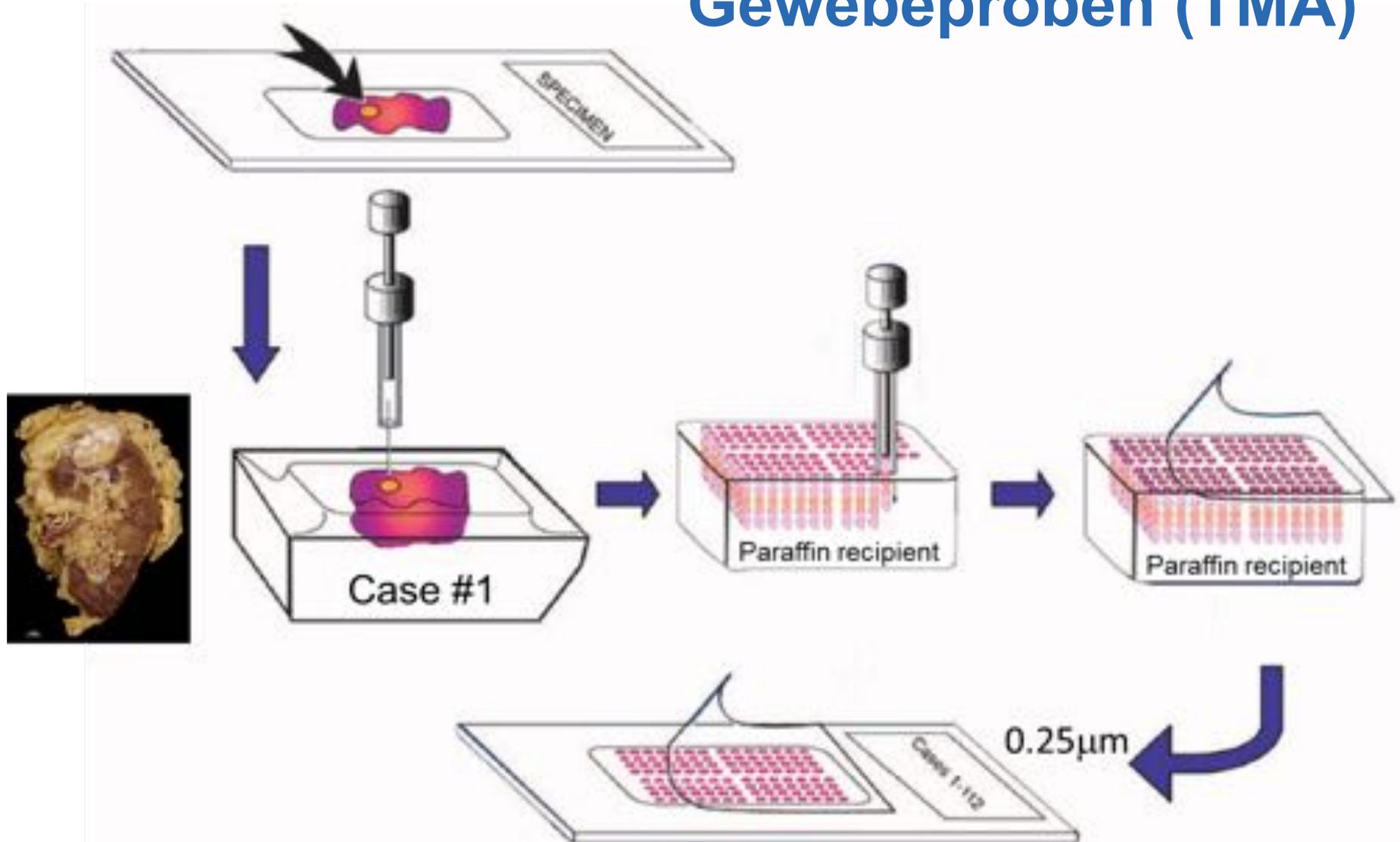
Nierenkrebs Prognose ist schlecht.

Häufig leiden Patienten bereits an Metastasen, wenn die erste Diagnose erstellt wird.

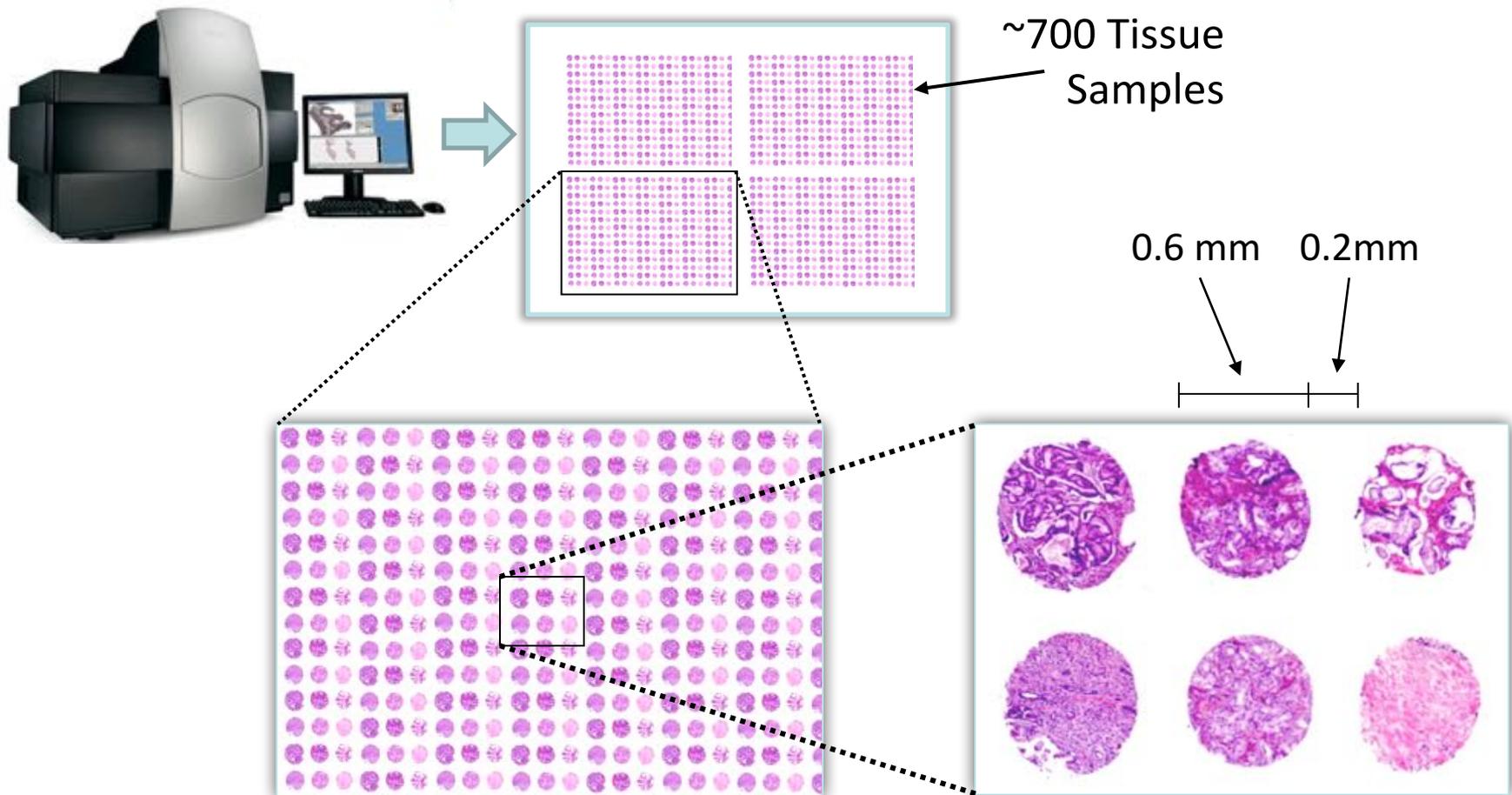
Identifikation von Biomarkern für Prognose (**prognostic marker**) oder Antwort auf Therapie (**predictive marker**) ist von höchster Bedeutung, um die Patientenprognose zu verbessern.



Gewebeproben (TMA)



Digitale Gewebeproben einer Kohorte



Expertenwissen als Gold Standard für Pathologie



TMA Annotator



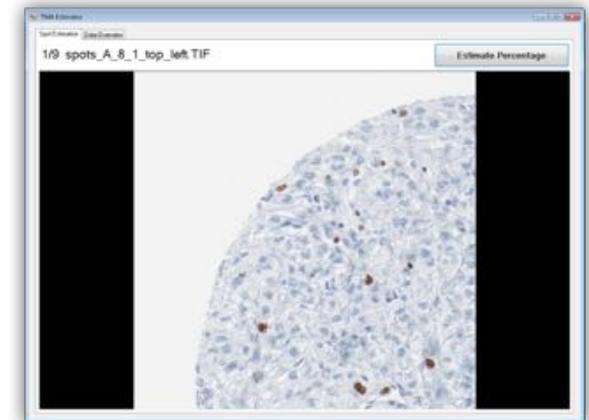
Detektion von Zellkernen im histologischen Schnitt (20% Fehler).

TMA Classifier



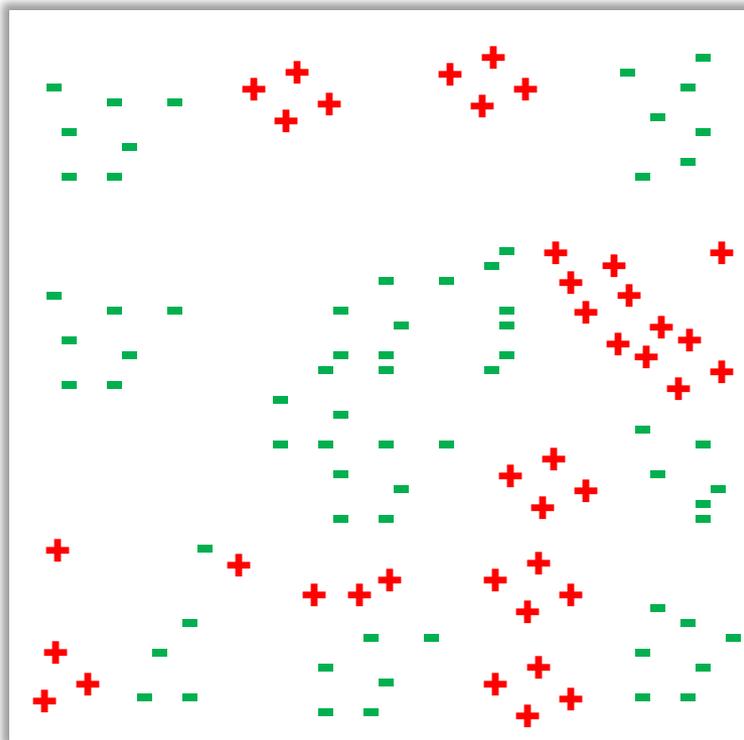
Klassifikation einzelner Zellkerne in bösartig, gutartig, gefärbt und nicht gefärbt (20% Fehler).

TMA Estimator



Schätzung, wie stark ein histologischer Schnitt gefärbt ist (+- 2 Ränge).

Lernende Algorithmen: Krebsdiagnose mit Entscheidungsbäumen



■ Gutartige Stichproben



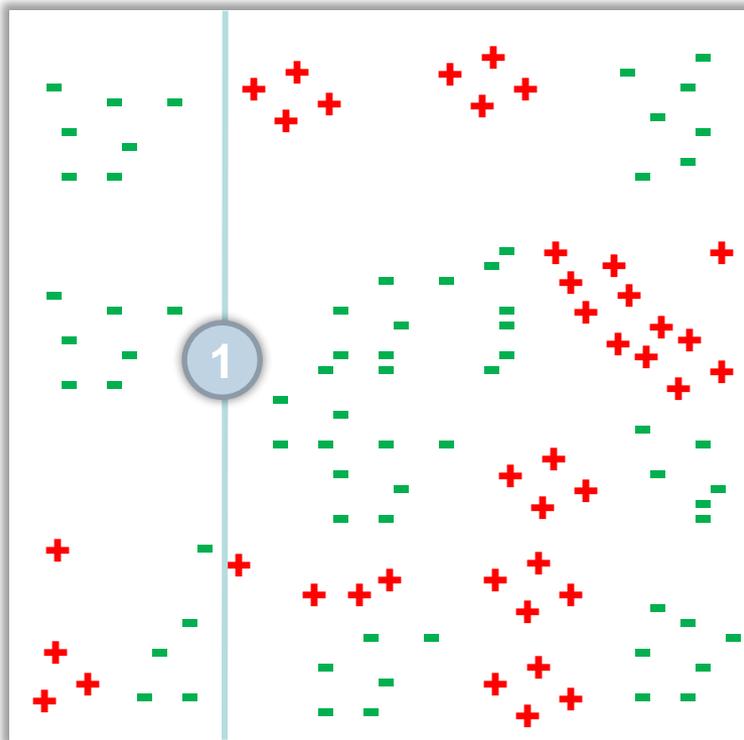
⊕ Bösartige Stichproben



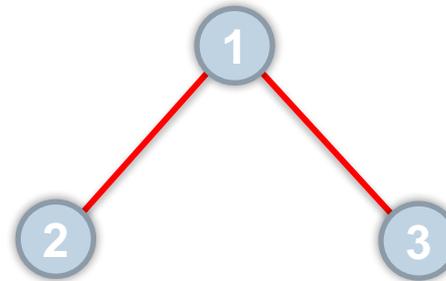
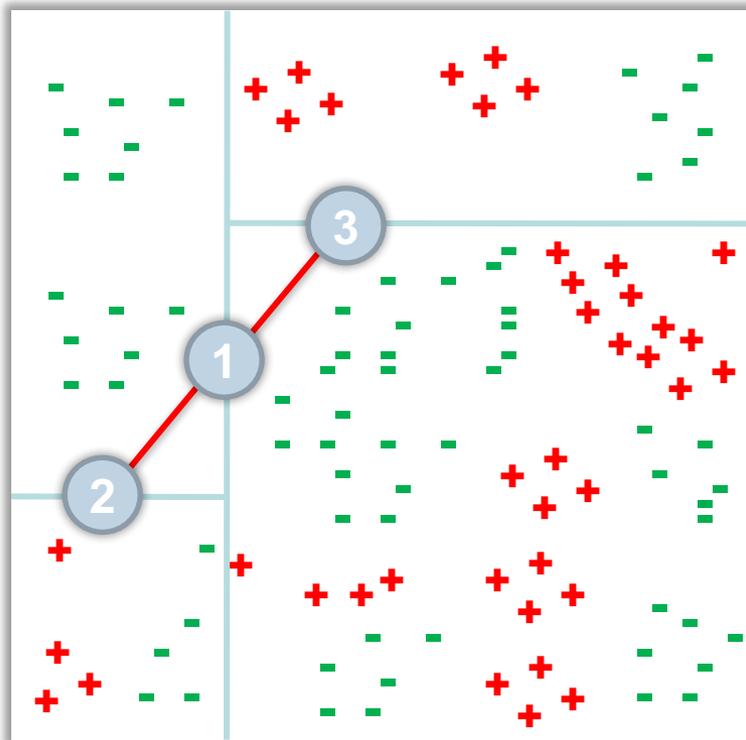
Lernen eines Entscheidungsbaumes

1

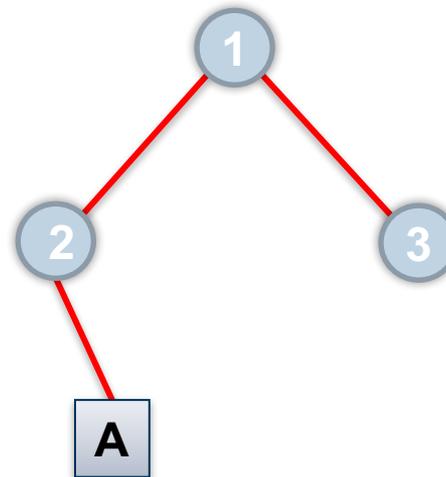
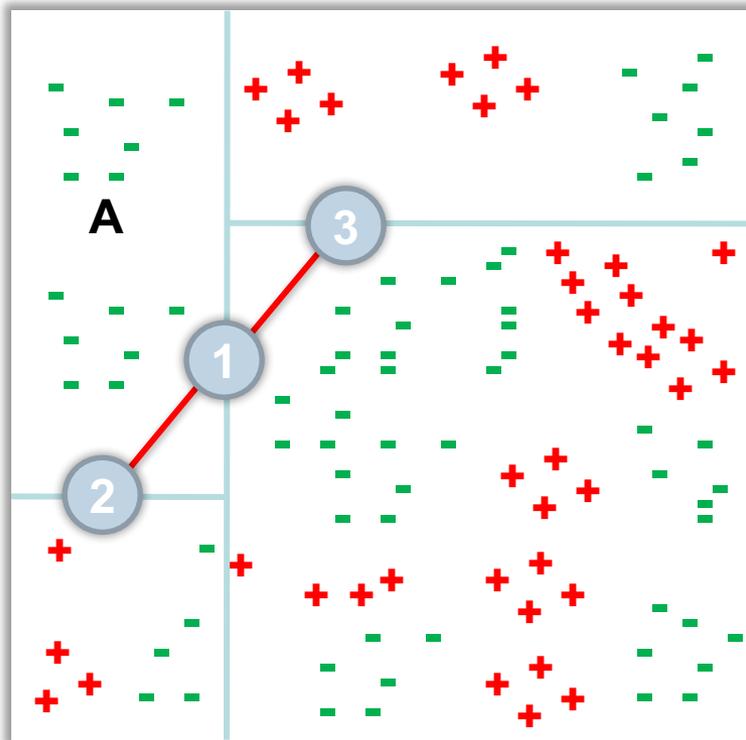
1



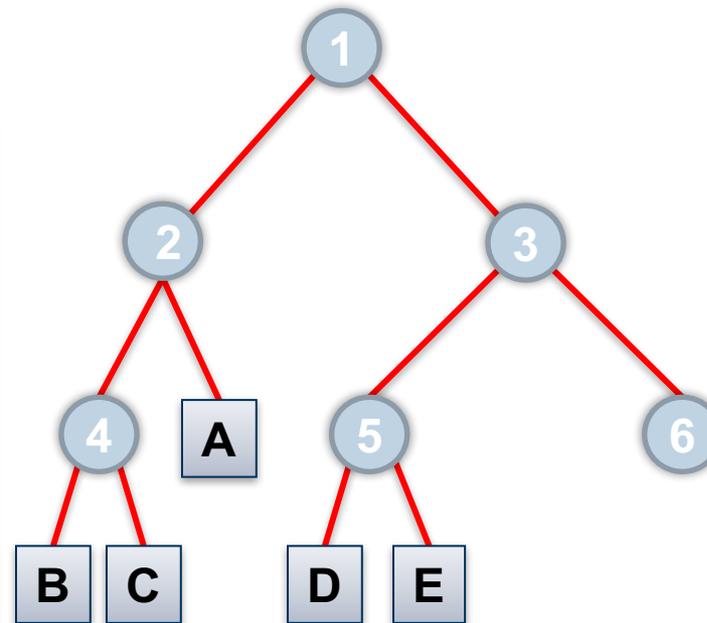
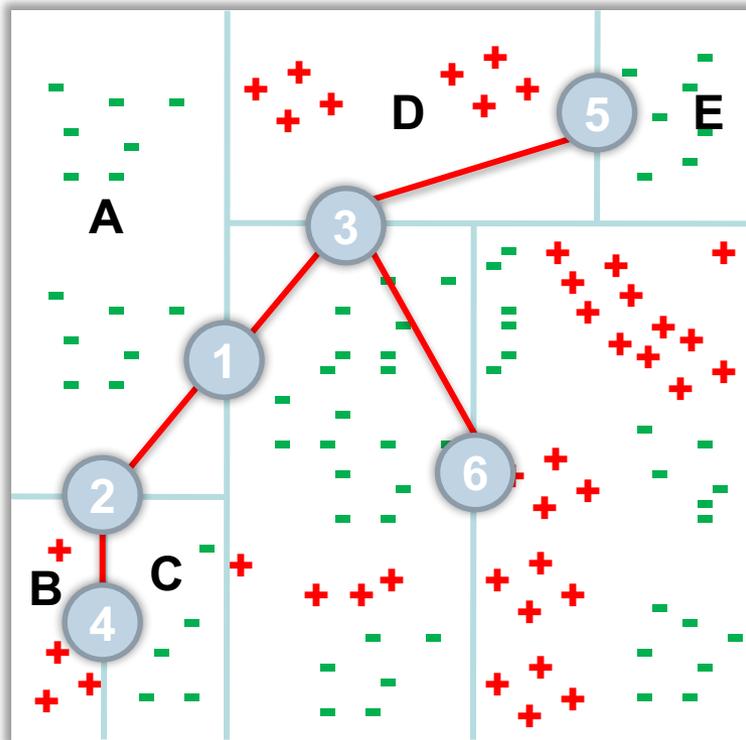
Lernen eines Entscheidungsbaumes



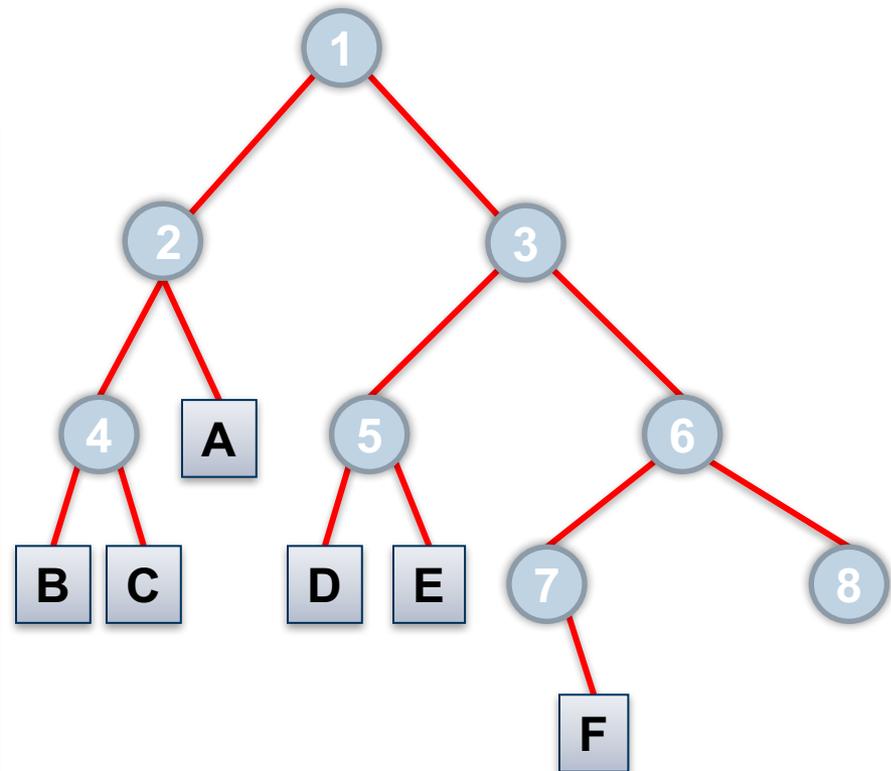
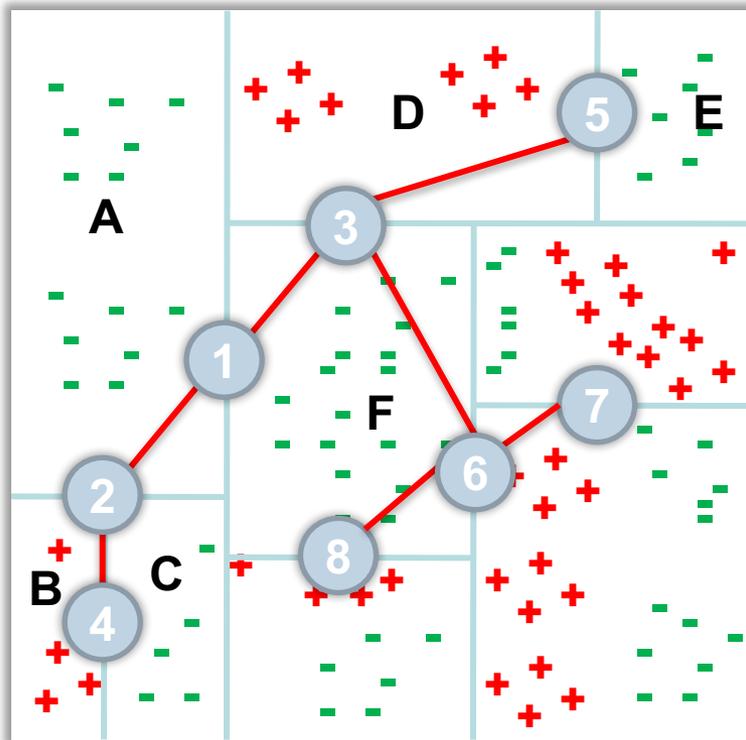
Lernen eines Entscheidungsbaumes



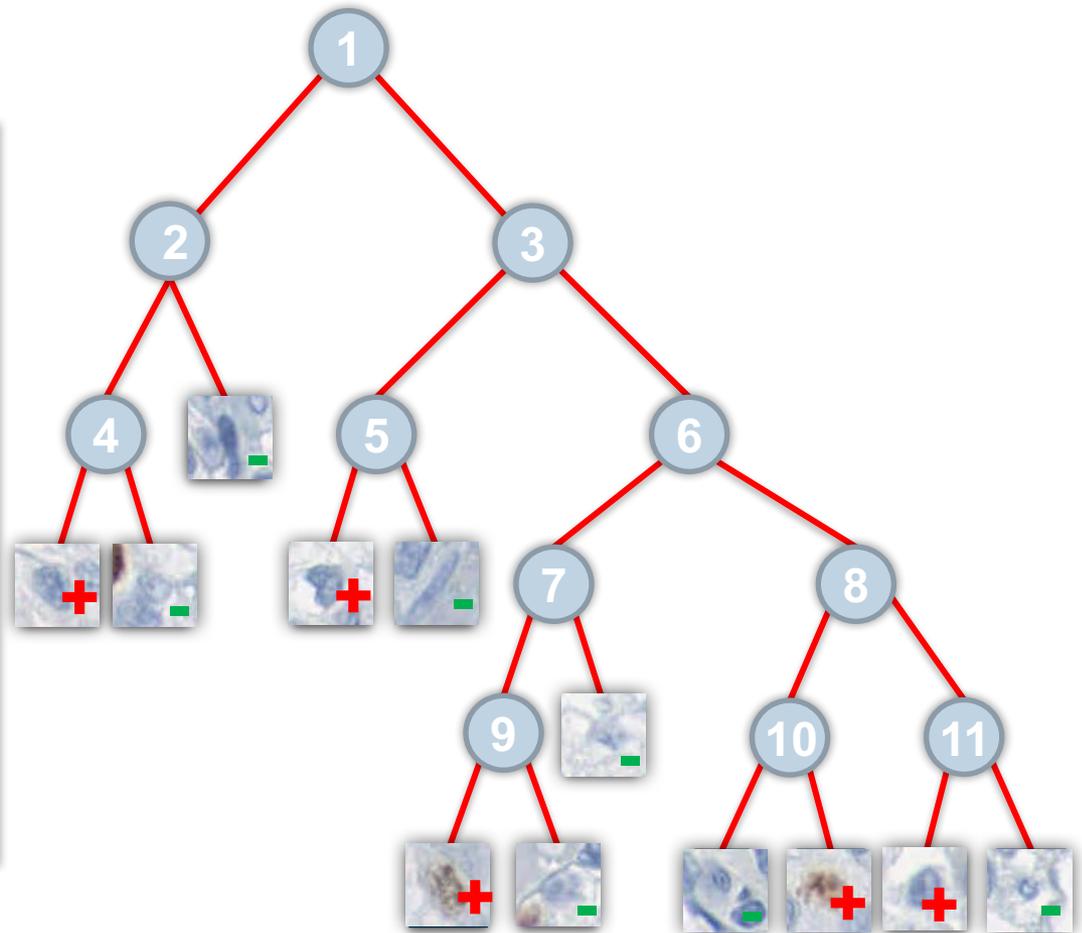
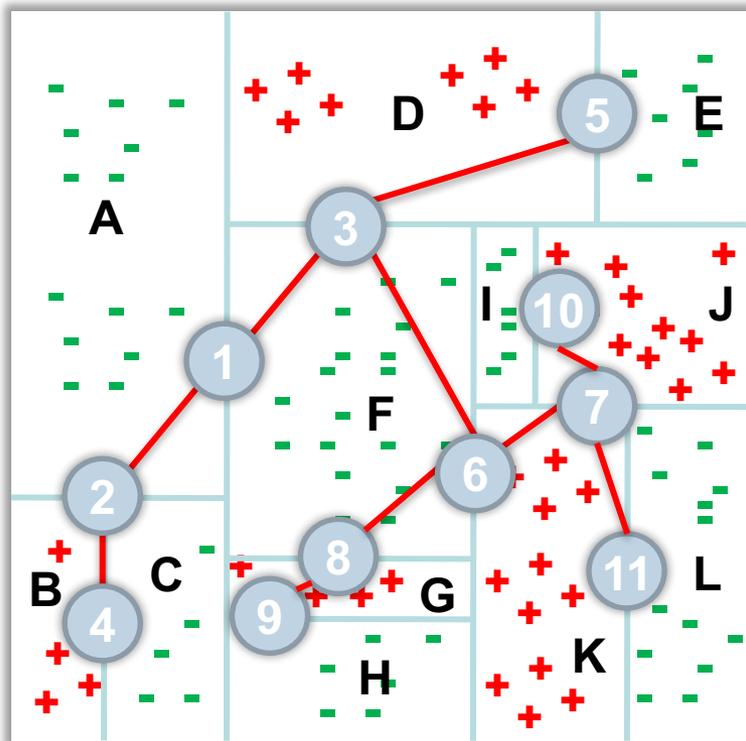
Lernen eines Entscheidungsbaumes



Lernen eines Entscheidungsbaumes

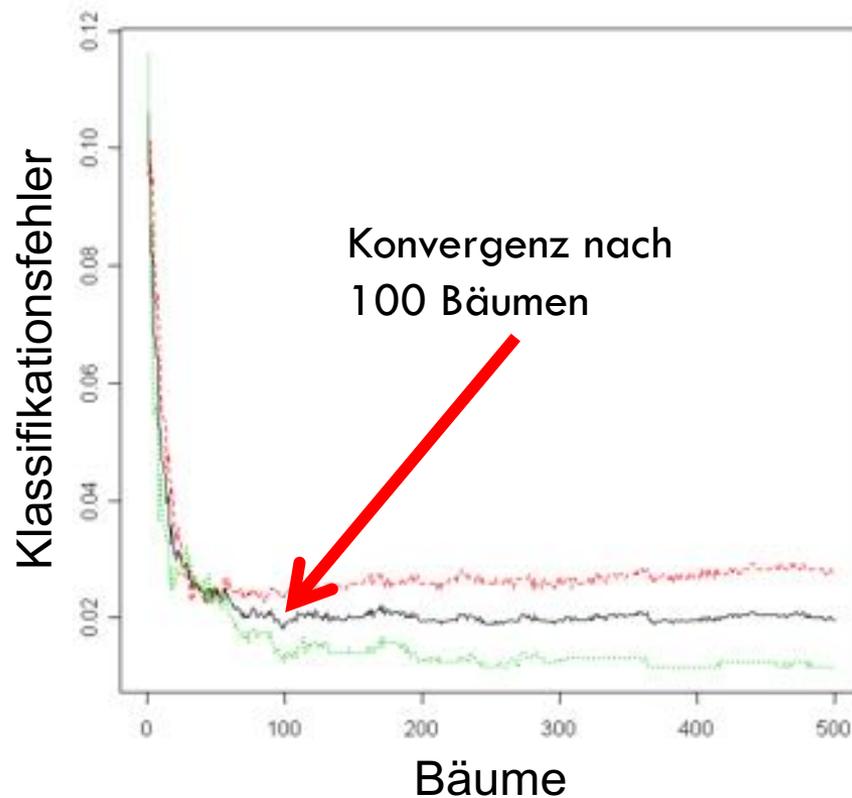


Lernen eines Entscheidungsbaumes



Klassifikation mit Entscheidungsbäumen

Neue Stichprobe =



Gesamt-
vorhersage:
Mehrheit über
alle Bäume



Baum T



Datenerzeugung

Image Aquisition: X

Sample preparation and TMA construction. MIB-1 staining on clear cell renal cell carcinoma

Label Aquisition: Y

Slide scanning and tiling of TMA into spots

Gold standard: samples of nuclei via labeling experiments

Background objects through Voronoi sampling

Training samples: cancerous nuclei

background

Bildanalyse

Learning

Relational Detection Forest

cancerous nuclei + background

Prediction

Classifying every pixel of a spot results in a probability map

Nuclei centers are found by applying mean shift clustering

Several hundred nuclei are detected on each image of a TMA spot

Estimation

Overall staining per patient is Calculated by assessing the staining of all detected nuclei

x%

Überlebensstatistik

Patient Cohort

Application to TMA spots of 133 RCC patients.

Staining Estimation

MIB-1 estimations

- 0.4%
- 0.8%
- 1.5%
- ...
- 12.1%
- 17.4%
- 23.6%

Estimation from the domain expert and prediction from the algorithm for each patient in the cohort

Subgroup Analysis

Kaplan-Meier estimators for subgroups of patient with high and low MIB-1 expression.

Holistischer Blick auf rechnergestützte Pathologie

Medizinische Technologien

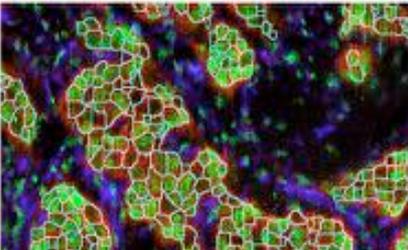
Hochdurchsatz Scanning



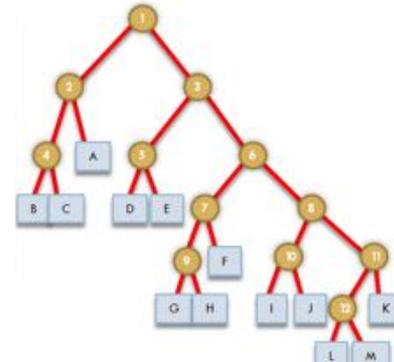
Digitalisierung von Bildarchiven



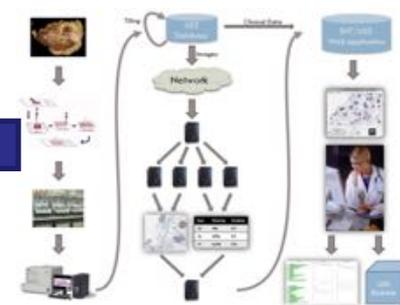
Neue bildgebende Verfahren



Maschinelles Lernen



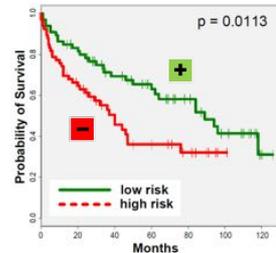
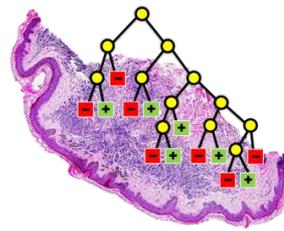
Big Data Science



Expert/Crowd Sourcing



Rechnergestützte Pathologie



A blue-tinted photograph of a large, classical-style building with a prominent dome and arched windows, likely a part of the ETH Zurich campus, serving as a background for the top of the slide.

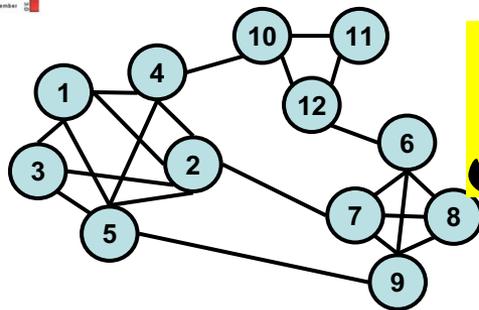
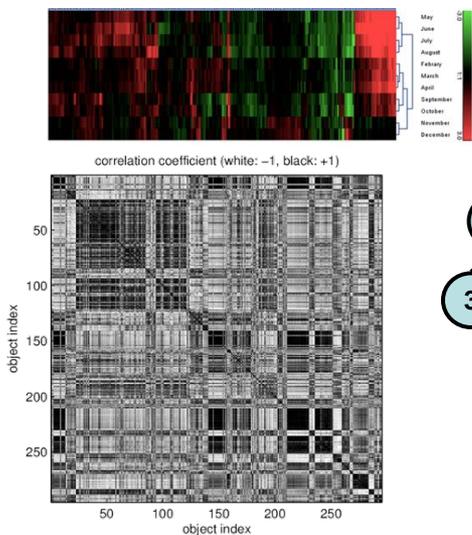
4. Holy Grail of ML

Validierung von Algorithmen zur Datenanalyse

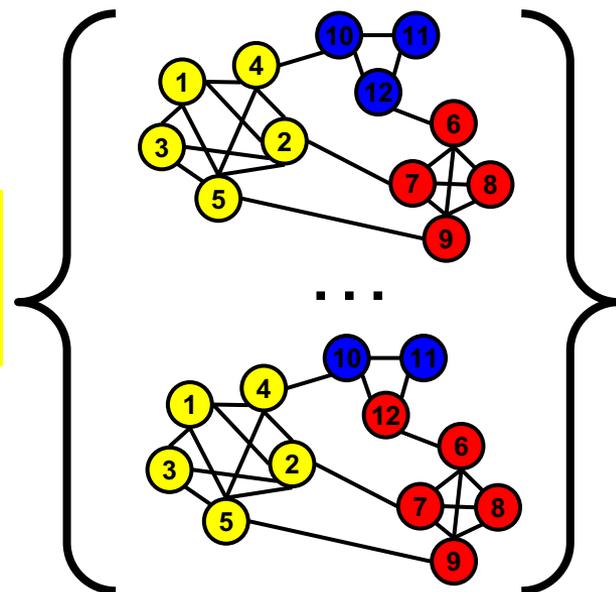
Algorithmik für Datenanalyse - was ist das Problem?

- Zufällige Eingabe bedingt zufällige Ausgabe

$$\underbrace{\text{Eingabe } \mathbf{X} \sim P(\mathbf{X})}_{\text{gegeben}} \implies \underbrace{\mathcal{A}}_{\text{Algorithmus}} \implies \underbrace{\text{Ausgabe } c \sim P(c|\mathbf{X})}_{\text{Entwurf}}$$

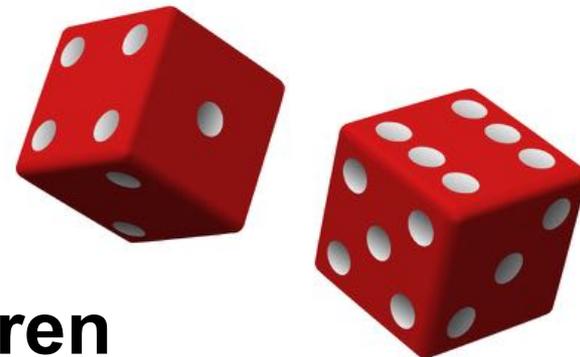


A?



Algorithmik für Datenanalyse: Gedanken

- **Zufall in Daten reduziert die Auflösung im Lösungsraum!**
- **Resiliente Algorithmen lokalisieren Lösungen/Hypothesen für verrauschte Daten mit starkem Signal.**
- **Ziel: Bewerte Algorithmen nach *Sensitivität für Signal* und *Robustheit gegen Rauschen***



Untersuche $(P(\mathbf{X}), \mathcal{A}(\mathbf{X})) \Rightarrow$ **Informationstheorie**

Algorithmen zur Datenanalyse

- gegeben: Datenraum \mathcal{X} , Lösungsraum \mathcal{C} ,
Algorithmus $\mathcal{A} : \mathcal{X} \rightarrow \mathcal{C}, \mathbf{X} \mapsto c^\perp$
- Daten \mathbf{X} sind Zufallsvariablen, die eine Verteilung über Lösungen induzieren

$$P(c^\perp) = \int P(\mathbf{X}) \delta(c^\perp - \mathcal{A}(\mathbf{X})) d\mathbf{X}$$

- **Regularisierung eines Algorithmus**

ersetze $\mathcal{A}(\mathbf{X}) = c^\perp$ durch Posterior $P^{\mathcal{A}}(c|\mathbf{X})$

$$P^{\mathcal{A}}(c) = \int P(\mathbf{X}) P^{\mathcal{A}}(c|\mathbf{X}) d\mathbf{X}$$

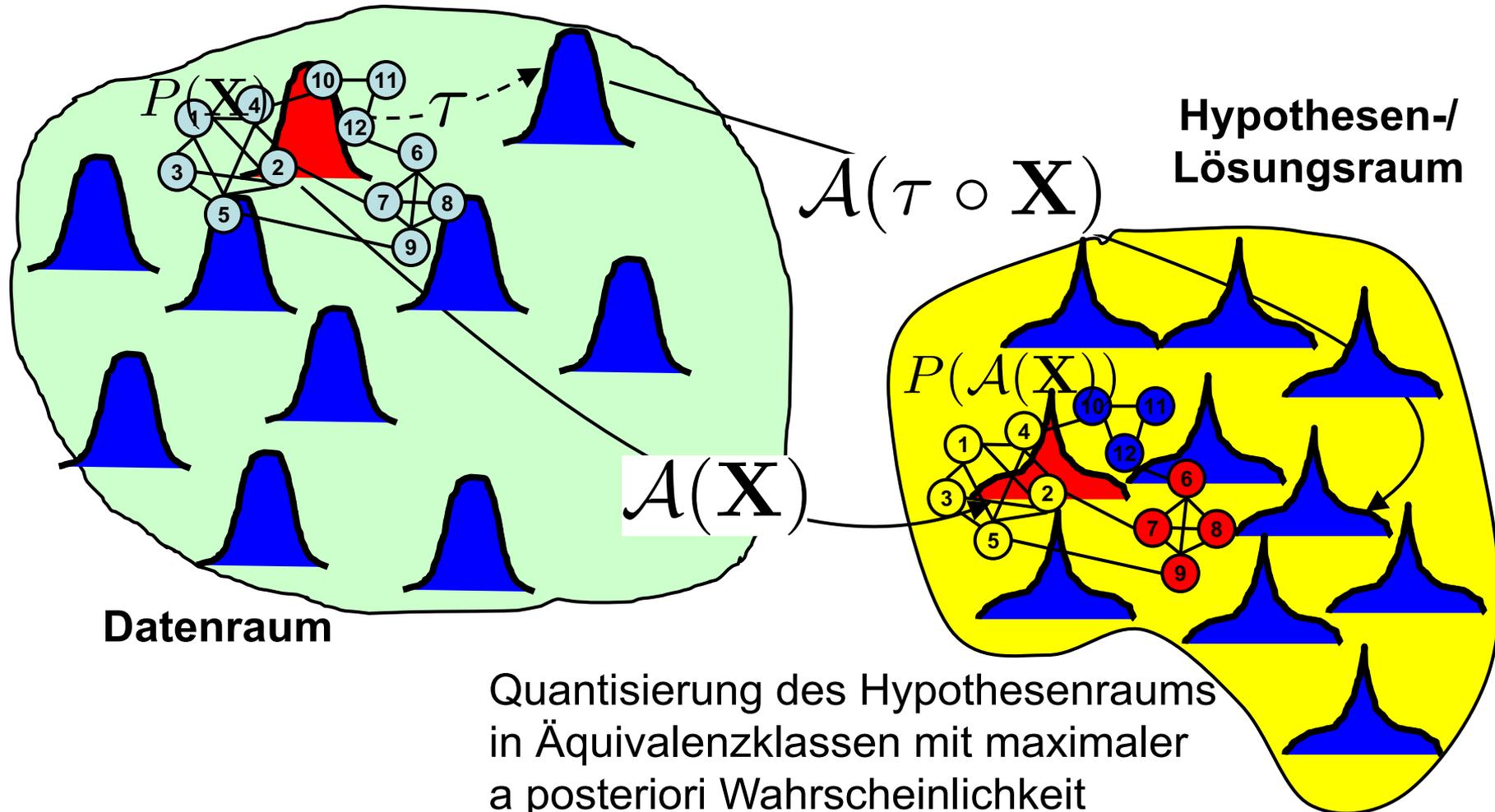
Gibbs Verteilungen für Optimierung

- **Gegeben sind Kosten** $R : \mathcal{C} \times \mathcal{X} \rightarrow \mathbb{R}$
- **Gibbs Posterior maximiert Entropie** für erwartete Kosten $\mathbb{E}_{c|\mathbf{X}} [R(c, \mathbf{X})]$!

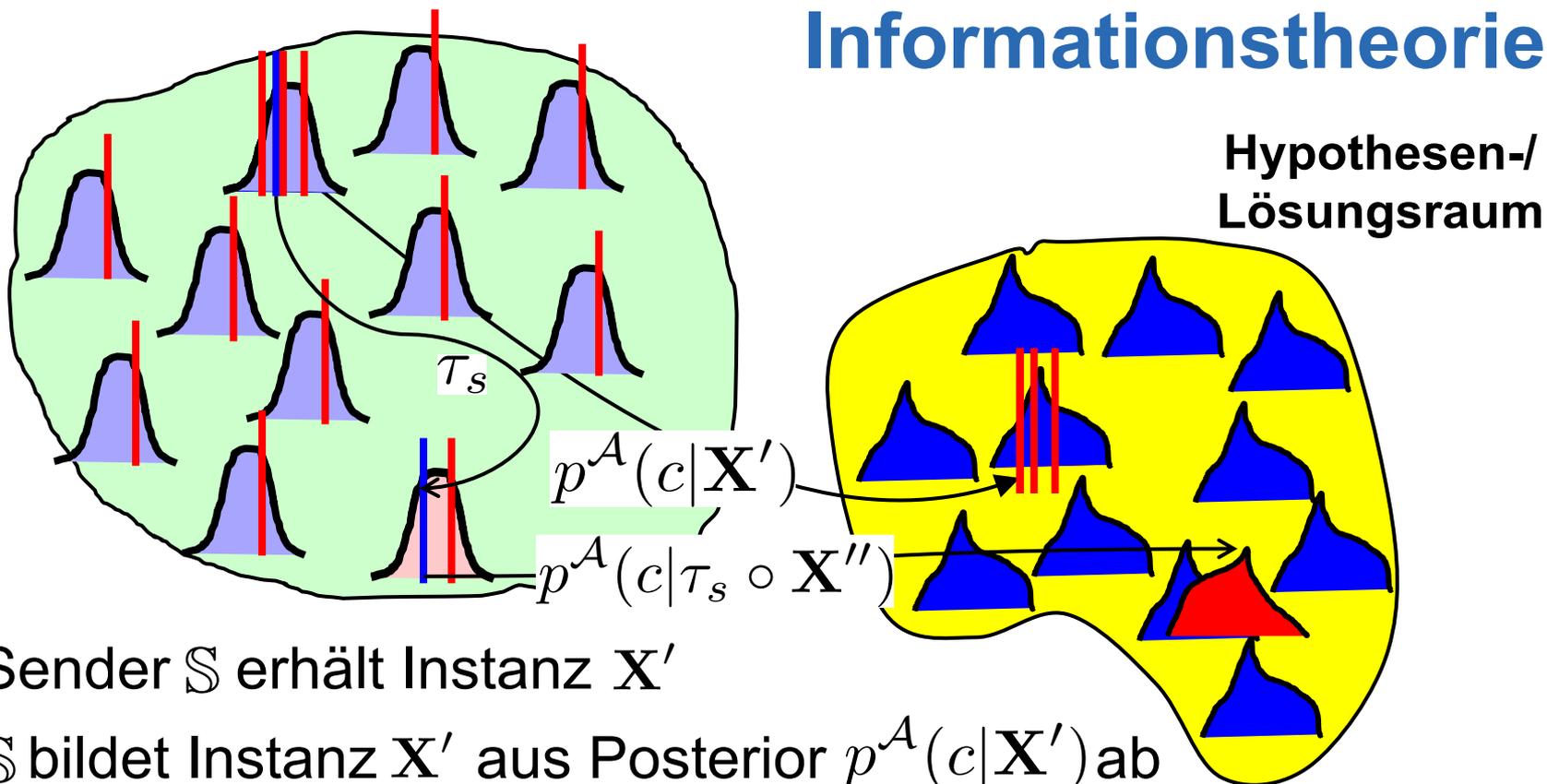
$$P_t(c|\mathbf{X}) = \frac{\exp(-\beta_t R(c, \mathbf{X}))}{\sum_{c' \in \mathcal{C}} \exp(-\beta_t R(c', \mathbf{X}))}$$

- **Robustheit** durch Entropiemaximierung
- **Annealing**: erhöhe iterativ β_t während der Ausführung des Algorithmus

Rauscheingabe quantisiert Hypothesenraum

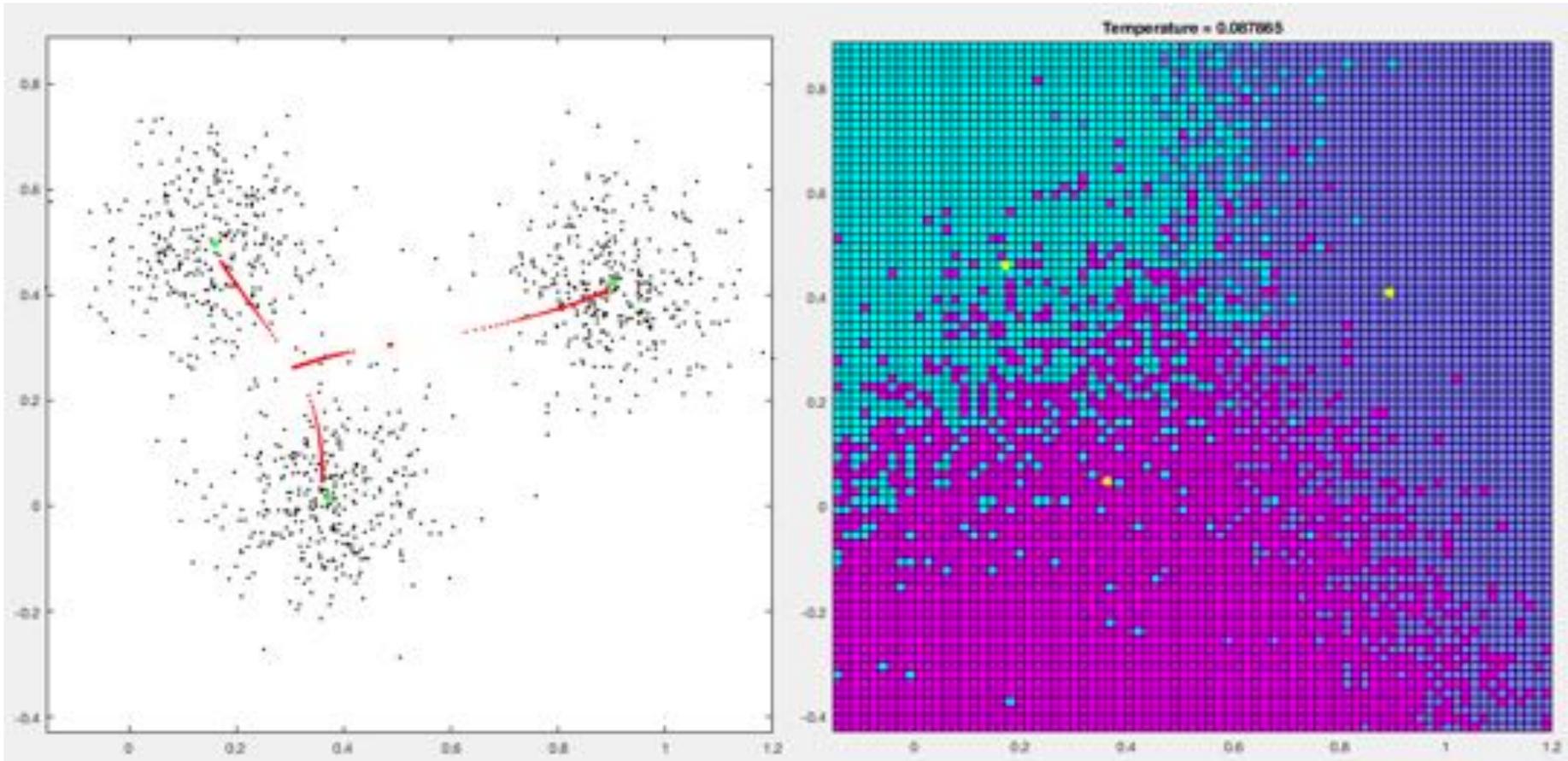


Algorithmenvalidierung mit Informationstheorie



- Sender \mathcal{S} erhält Instanz \mathbf{X}'
- \mathcal{S} bildet Instanz \mathbf{X}' aus Posterior $p^A(c|\mathbf{X}')$ ab
- Orakel zieht neue Instanz $\tilde{\mathbf{X}} = \tau_s \circ \mathbf{X}''$
- Empfänger schätzt Abbildung $\hat{\tau}$ **posterior agreement**

Annealing Dynamik für *k-means*



Lokalisierung von Lösungen

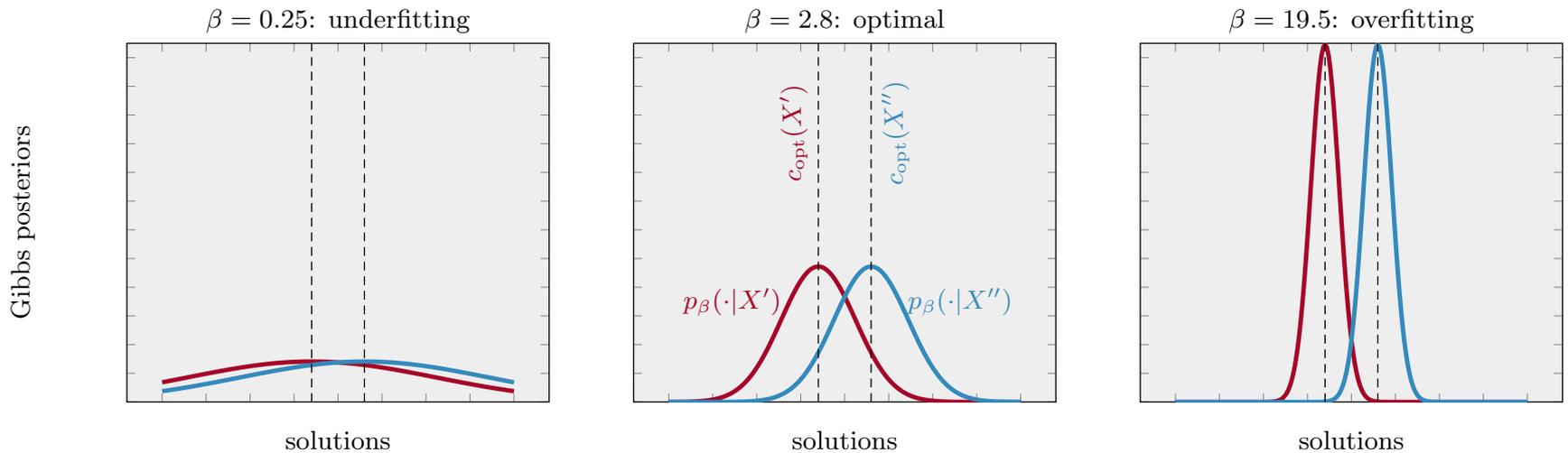
- **“Posteriors”** für **typische** Daten \mathbf{X}' , \mathbf{X}'' sollten übereinstimmen!

$$k^{\mathcal{A}}(\mathbf{X}', \mathbf{X}'') = \sum_{c \in \mathcal{C}} p^{\mathcal{A}}(c|\mathbf{X}') p^{\mathcal{A}}(c|\mathbf{X}'') \in [0, 1]$$

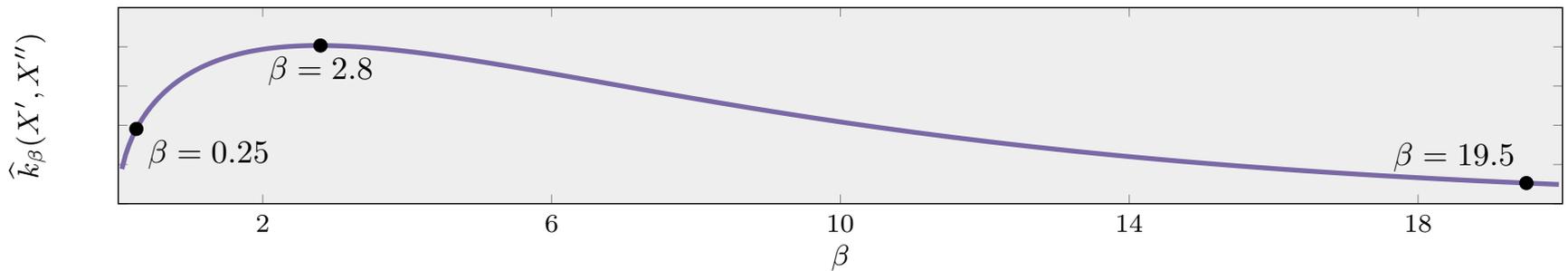
Zu breite oder zu schmale Posterior $p^{\mathcal{A}}(\cdot|\mathbf{X})$ erzeugen kleine Kernwerte $k^{\mathcal{A}}(\mathbf{X}', \mathbf{X}'')$! Optimiere Breite β_t

$$\mathbb{P}^* \in \arg \max_{\{\mathcal{A}\}} \max_t \mathbb{E}_{\mathbf{X}', \mathbf{X}''} \log(|\mathcal{C}| k_t^{\mathcal{A}}(\mathbf{X}', \mathbf{X}''))$$

Lokalisierung von Lösungen



(a)



(b)

Konsistentes Lernen eines Algorithmus

- **Lernen** bedingt, dass die empirische Kapazität gegen ihren Erwartungswert konvergiert

$$\mathbb{E}_{\mathbf{X}', \mathbf{X}''} \log(|\mathcal{C}| \hat{k}^{\mathcal{A}}(\mathbf{X}', \mathbf{X}'')) \geq$$

$$\frac{1}{L} \sum_{l \leq L} \log(|\mathcal{C}| \hat{k}^{\mathcal{A}}(\mathbf{X}'_l, \mathbf{X}''_l)) - \text{Strafterm}$$

mit optimalem empirischem Posterior \hat{P}^*

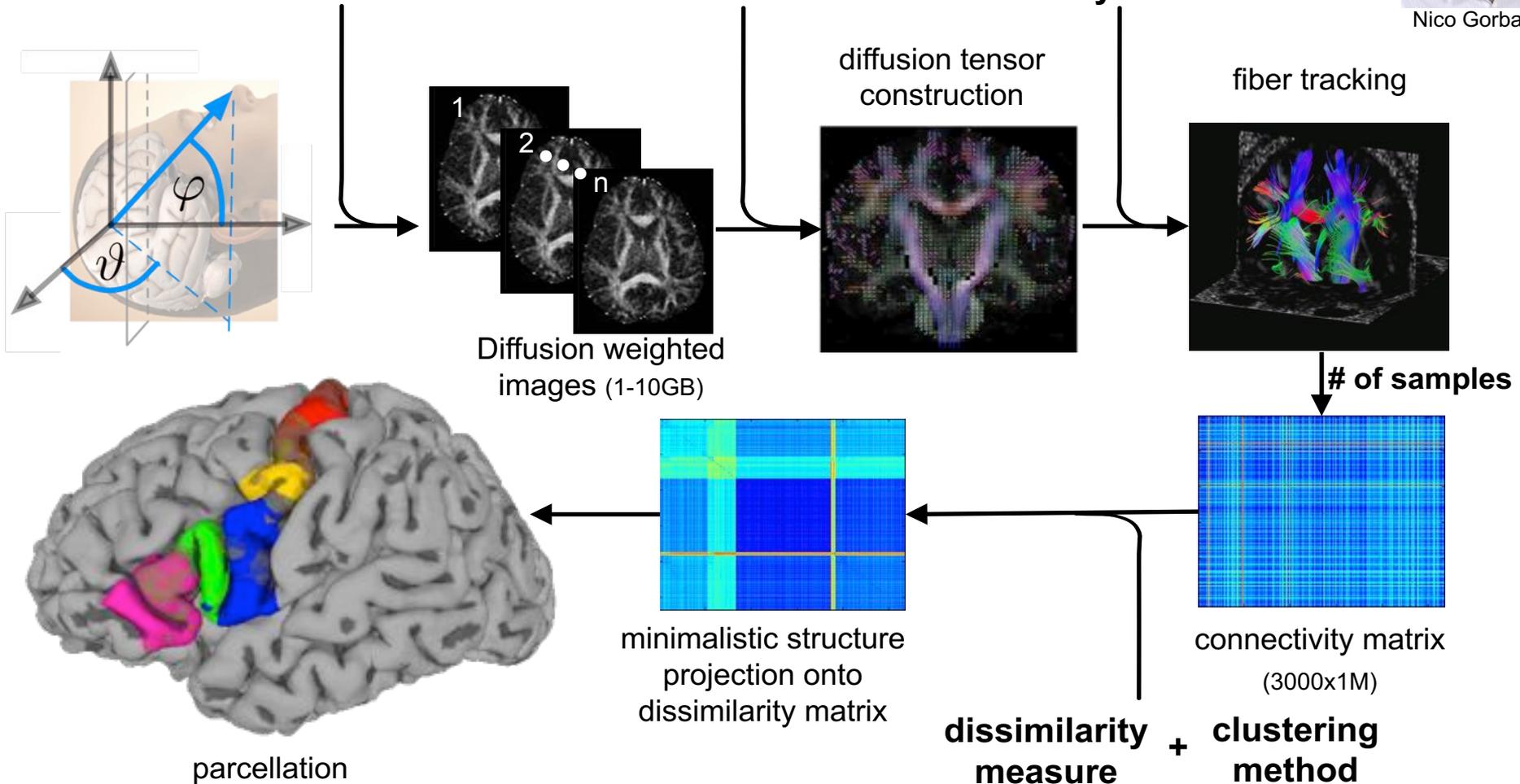
Diffusion weighted tensor imaging: pipeline



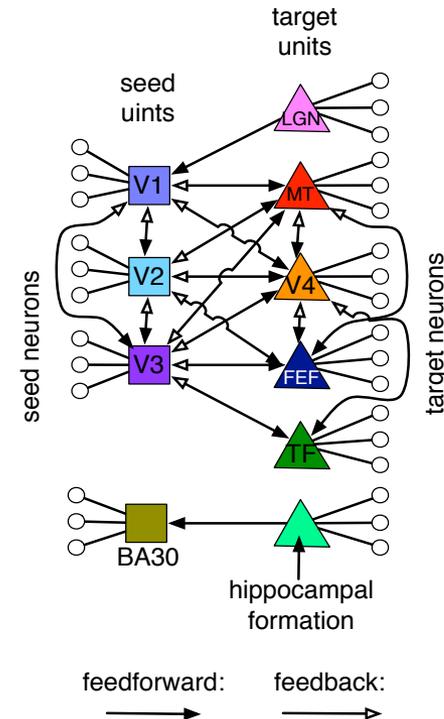
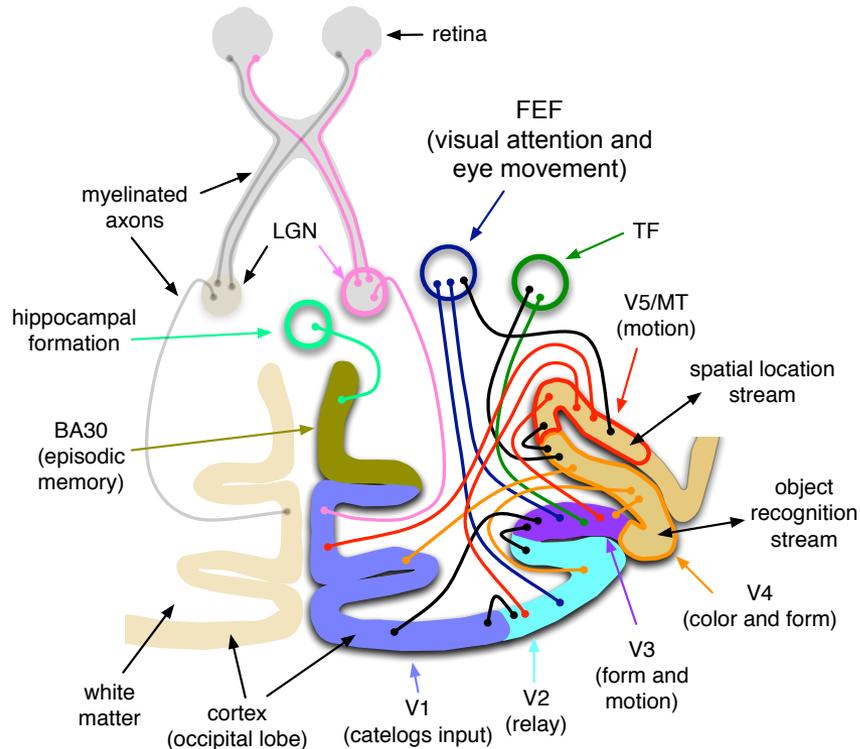
Nico Gorbach

scanner parameters

noise modeling

local fiber orientation
density function

Systems Neuroscience

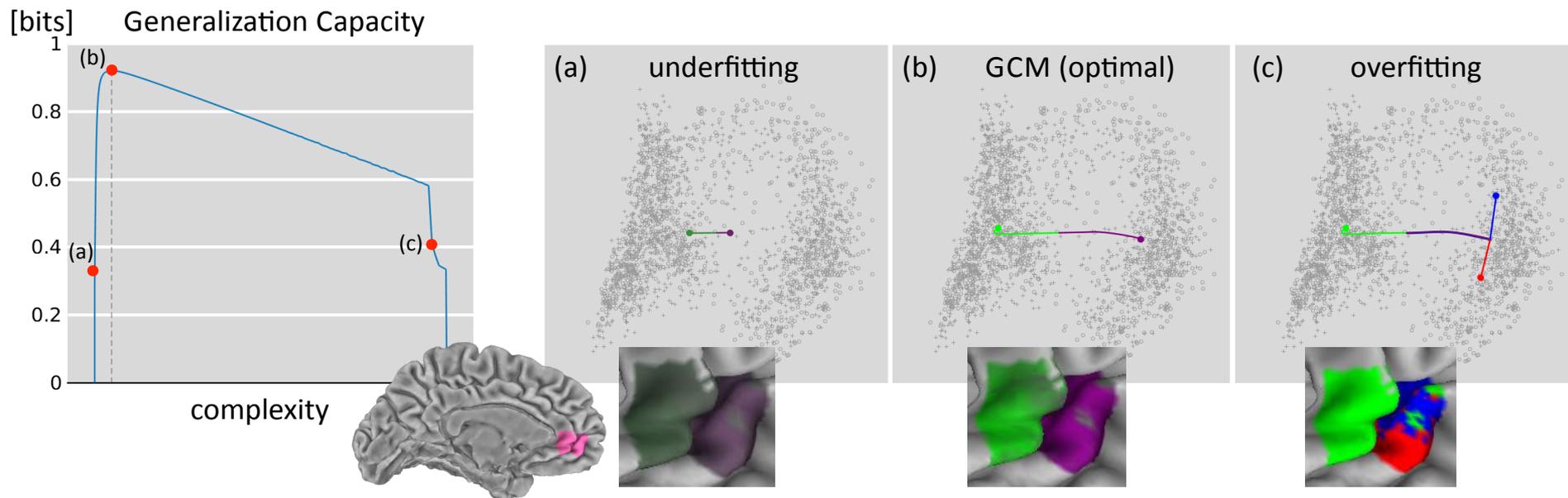


Subset of the visual system in the macaque monkey.
Target connections are limited for illustration purposes.

- The brain is considered as an ensemble of functionally specialized units coupled together in a modulatory fashion (Friston, 2002).

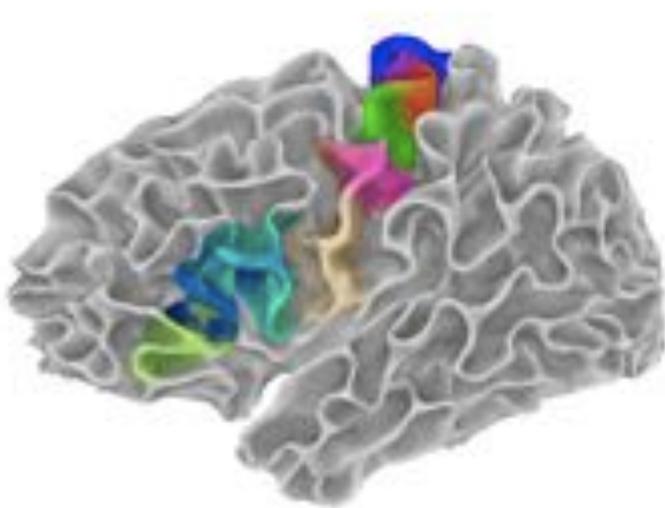
Under- and overfitting in parcellation

- Connectivity of two brain regions is analyzed
- Generalization capacity maximizer (GCM) outperforms empirical risk minimizer



Dynamics of cortex parcellation

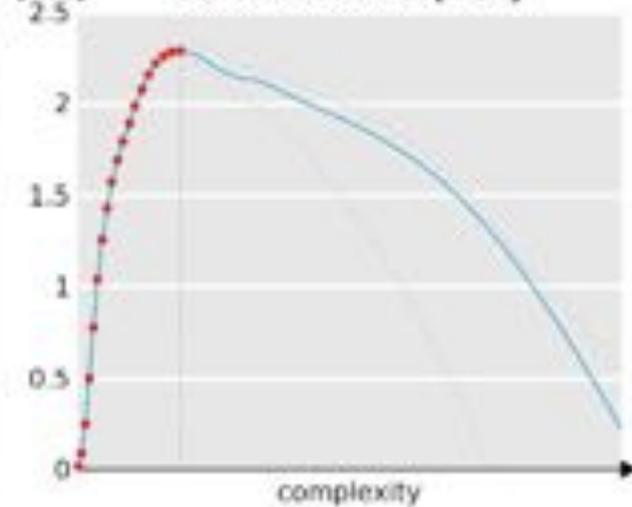
- Start at low resolution
- Estimate parcellations with higher resolution
- Stop at maximal generalization capacity



Dissimilarity between Centroids [bits]

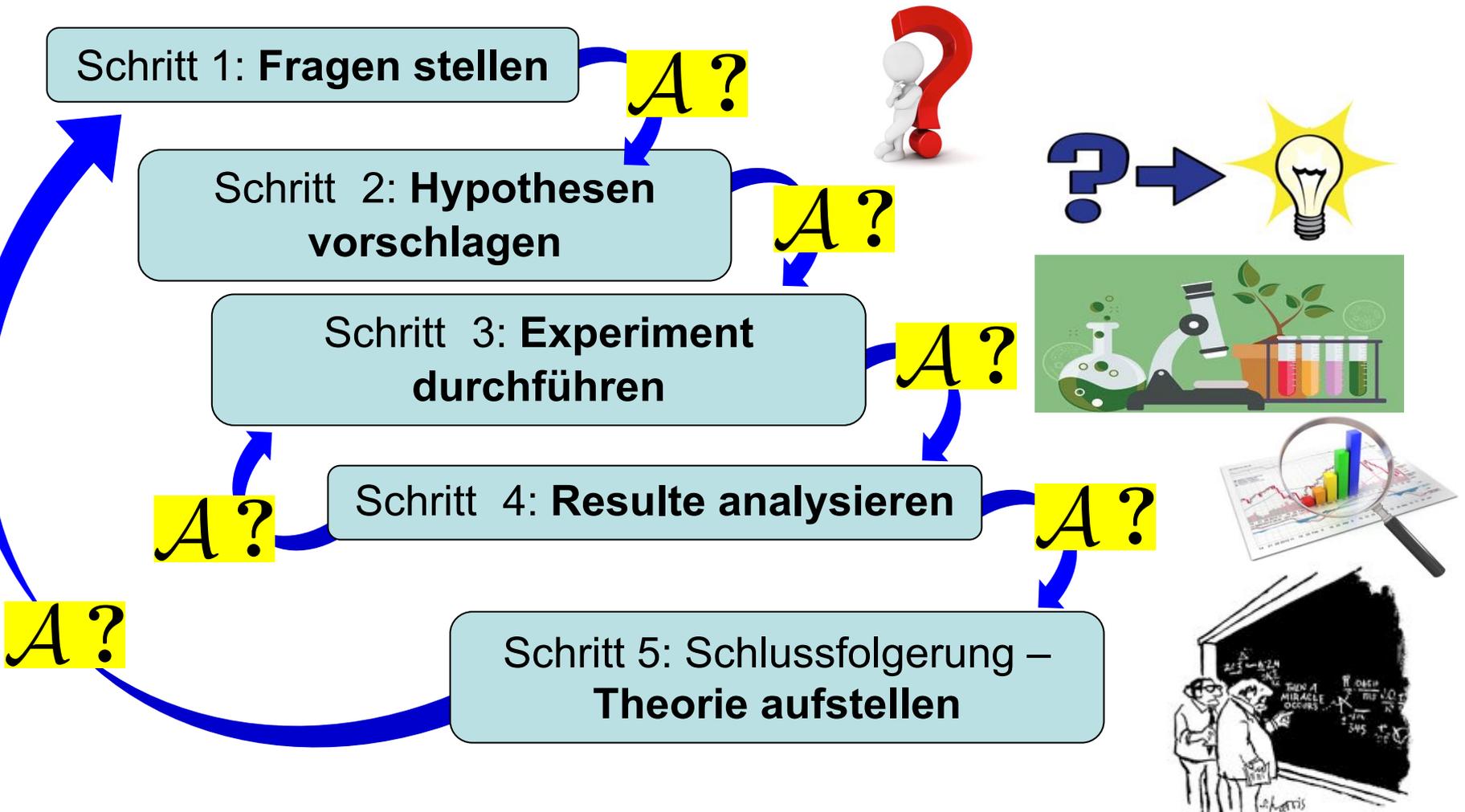


Generalization Capacity



Was fehlt noch?

Die wissenschaftliche Methode!



Machine Learning Magic

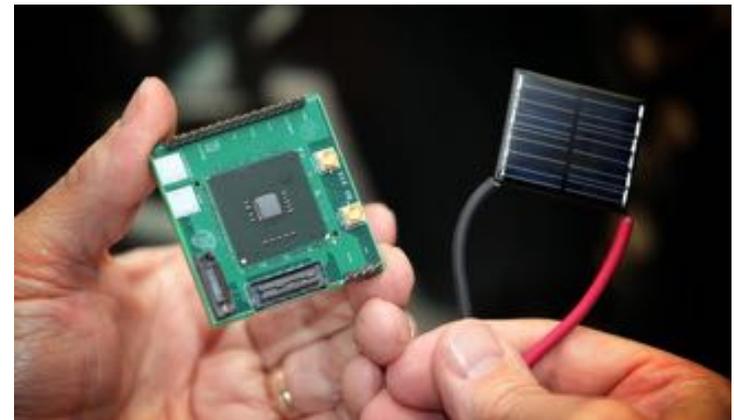


Low-Energy Architecture Trends

- Novel low-power architectures operate **near transistor threshold voltage (NTV)**

- e.g., Intel Claremont
- 1.5 mW @10 MHz (x86)

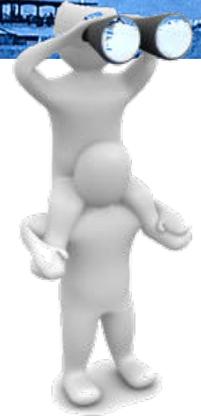
- NTV promises 10x more energy efficiency at 10x more parallelism!



source: Intel

- 10^5 times more soft errors (bits flip stochastically)
- Hard to correct in hardware → expose to programmer?

@ Thorsten Höfler



Thesen mit Ausblick

1. “Big Data” in der Medizin ist heute Realität

- Trotzdem: **Hoher Bedarf an medizinisch annotierten Daten!**

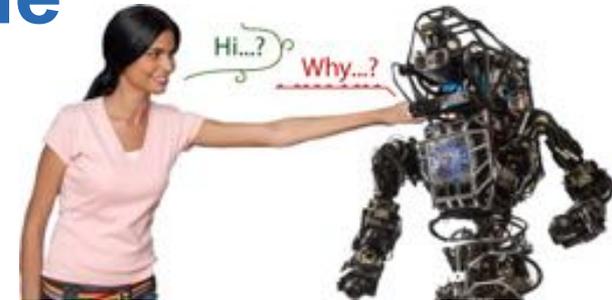
2. Herausforderungen für Informatik und Medizin

- *Knowledge Discovery*: komplexe Regelmässigkeiten in medizinischen Daten entdecken und nutzen
- Modelle, Algorithmik und Software für *Health Data Science*

3. “Big Data” Technologien werden heute erfunden.

- **Ethik**: Informative Selbstbestimmung mit gesellschaftlicher Solidarität
- **Begeisterung** fördert Kreativität!

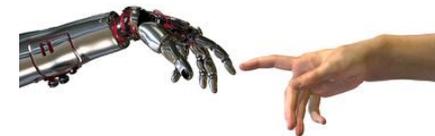
Ist der Mensch die unterlegene Maschine ?



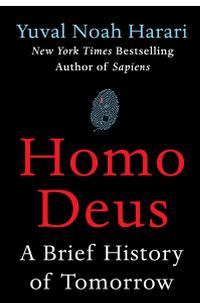
Wie ändert KI unser Weltbild ?

- Wir erleben durch KI gerade eine Zeitenwende wie z.B. die Entdeckung des Feuers oder das „Sesshaft-Werden“.
- Maschinen werden in 2-3 Generationen „klüger“ sein als wir!
- Vertreten wir ein rein materialistisches Weltbild, werden wir gegen die KI-Rechenmaschinen verlieren.

Theologisch Metaphysische Weltsicht



- Der Mensch ist ein komplettes Wesen aus Geist und Fleisch, Gefühlen, Bewusstsein, ... , d.h. mehr als eine Denkmachine!
- ***Gott schuf den Menschen ihm zum Bilde*** [1. Mose 1:27]



jbuhmann@inf.ethz.ch

