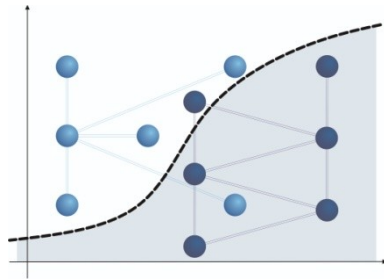


# Maschinelles Lernen und Data Mining: Methoden und Anwendungen

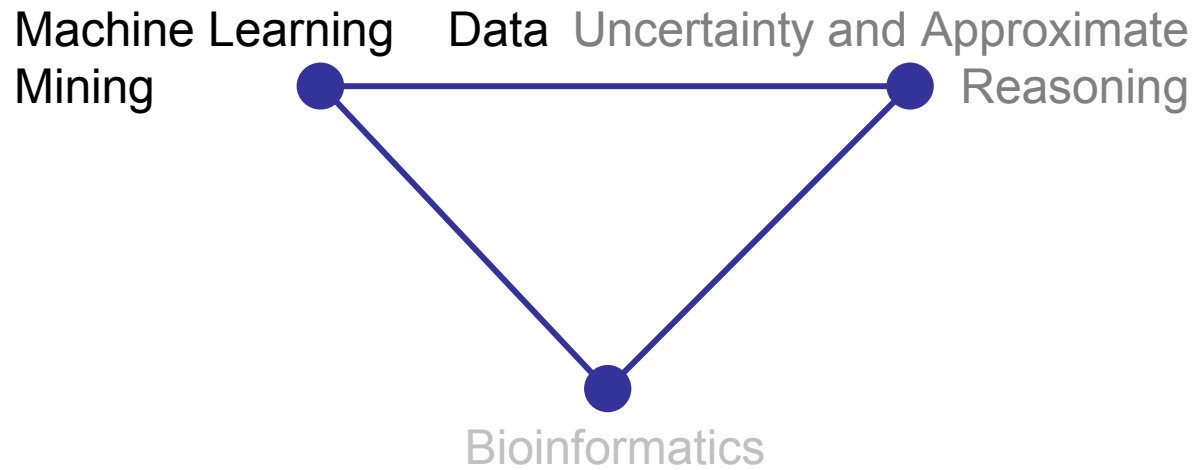
Eyke Hüllermeier  
Knowledge Engineering & Bioinformatics  
Fachbereich Mathematik und Informatik





Knowledge Engineering  
& Bioinformatics

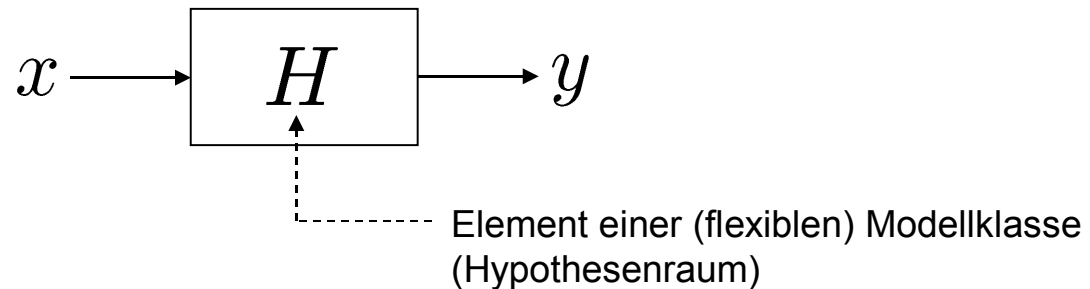
an der  
Philipps-Universität Marburg



## AGENDA:

1. Maschinelles Lernen und Data Mining
2. Forschungsschwerpunkte
3. Anwendungen

# Maschinelles Lernen: Adaption des (Ein-/Ausgabe) Verhaltens eines Systems auf der Basis von Beobachtungen und/oder Feedback



Suchen & Sortieren  
Kürzeste Pfade  
Lineare Programmierung  
...

Algorithmische Lösung:  
Wie löst man das Problem ?

Mustererkennung  
Fußballroboter  
Diagnose  
...

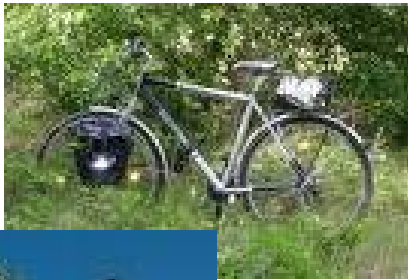
Automatische Adaption:  
Wie lernt man, das Problem  
zu lösen ?

**Eingabe:** Bild als Bitmap

**Ausgabe:** 1 falls ein Fahrrad vorhanden, 0 sonst

Positive Beispiele:

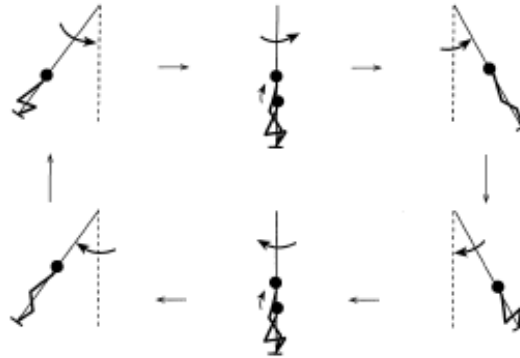
Negative Beispiele:



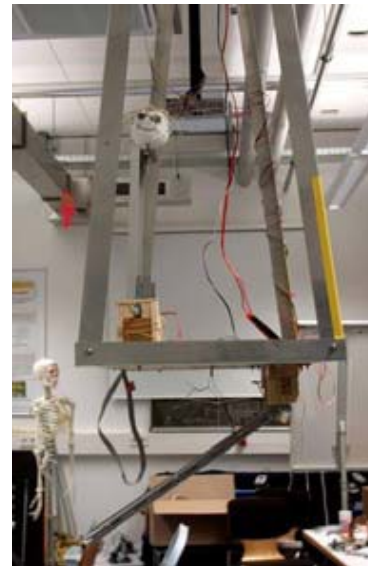
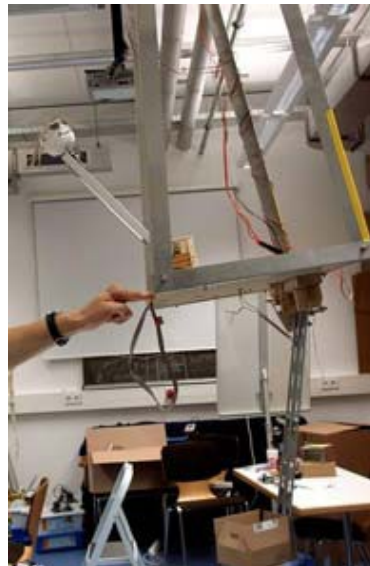
?



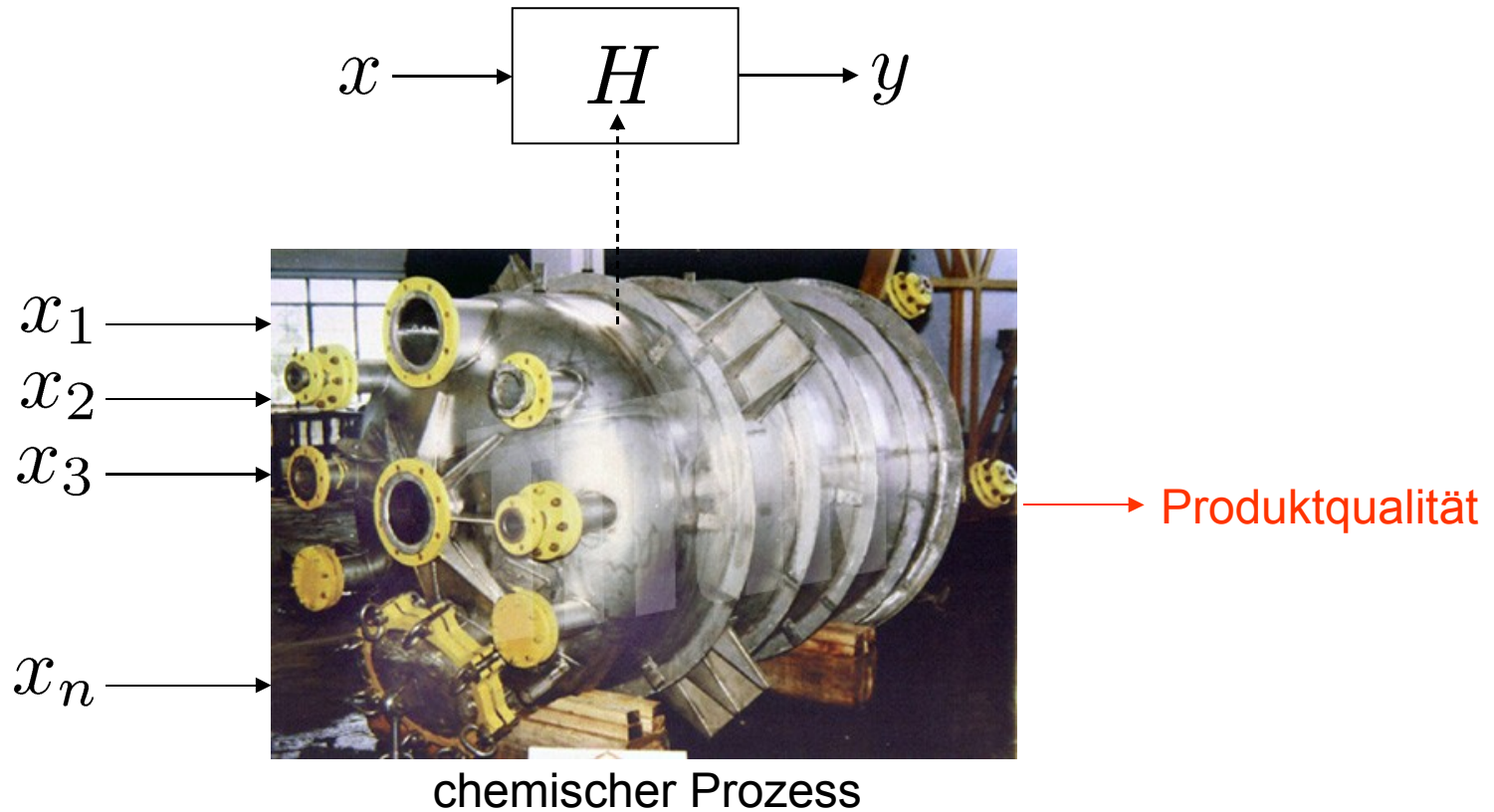
# Wie bringt man einem Roboter das Schaukeln bei?



Schaukeln durch Schwerpunktverlagerung



Wie hängen Eingaben und Ausgaben zusammen?  
(Abhängigkeit ist vorhanden, aber unbekannt → Analyse)



**Wissensentdeckung und Data Mining:** Suche nach Mustern und Strukturen in typischerweise sehr großen Datenmengen.

“Knowledge Discovery in Databases is the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in large databases.” [Fayyad et al., 1996]

**Etwas anderer Fokus als beim maschinellen Lernen:** eher explorativ und deskriptiv als induktiv, lokale Muster statt globale Modelle, ...

## **Beispiel Warenkorbanalyse:**

Daten sind Warenkörbe ( $\{P1, P7\}$ ,  $\{P6, P18, P22\}$ ,  $\{P1, P8\}$ , ...)

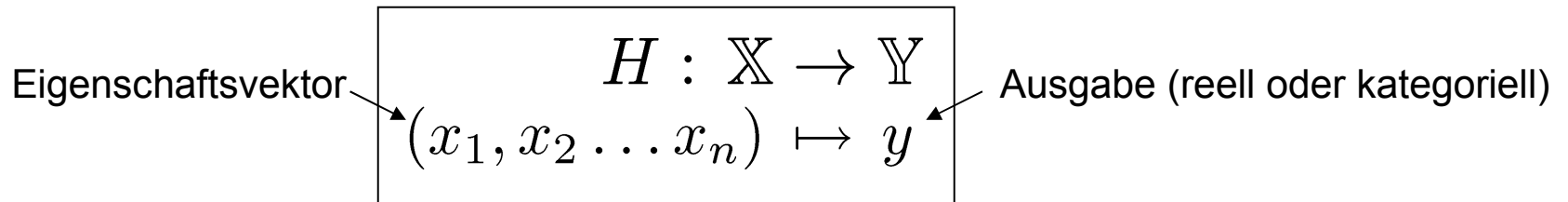
„Muster“ sind Assoziationsregeln, z.B. „Kunden, die Produkte P1 und P6 kaufen, kaufen häufig auch Produkte P5 und P12“.

## Anwendungen in vielen Gebieten aus Wissenschaft, Gesellschaft, Industrie und Wirtschaft:

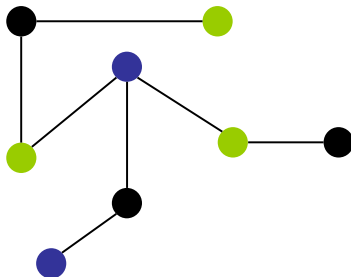
- Business (CRM, response prediction, ...)
- Banken und Finanzen (Aktienprognose, fraud detection, ...)
- Technische Systeme (Diagnose, Steuerung, Monitoring, ...)
- Smart Environments
- Bioinformatik, Analyse genomischer Daten
- Medizin (Diagnose, Prothetik, ...)
- Biometrics (Personenerkennung, ...)
- Medien (Spracherkennung, Bildererkennung, Video-Mining, ...)
- Internet (E-Mail Klassifikation, Personalisierung, WWW-log mining, ...)
- Games (z.B. Poker)
- ...

# Aktuelle Forschungsschwerpunkte

# Lernen mit strukturierten Ein-/Ausgaberräumen



Strukturierte Eingaben,  
z.B. Graphen:



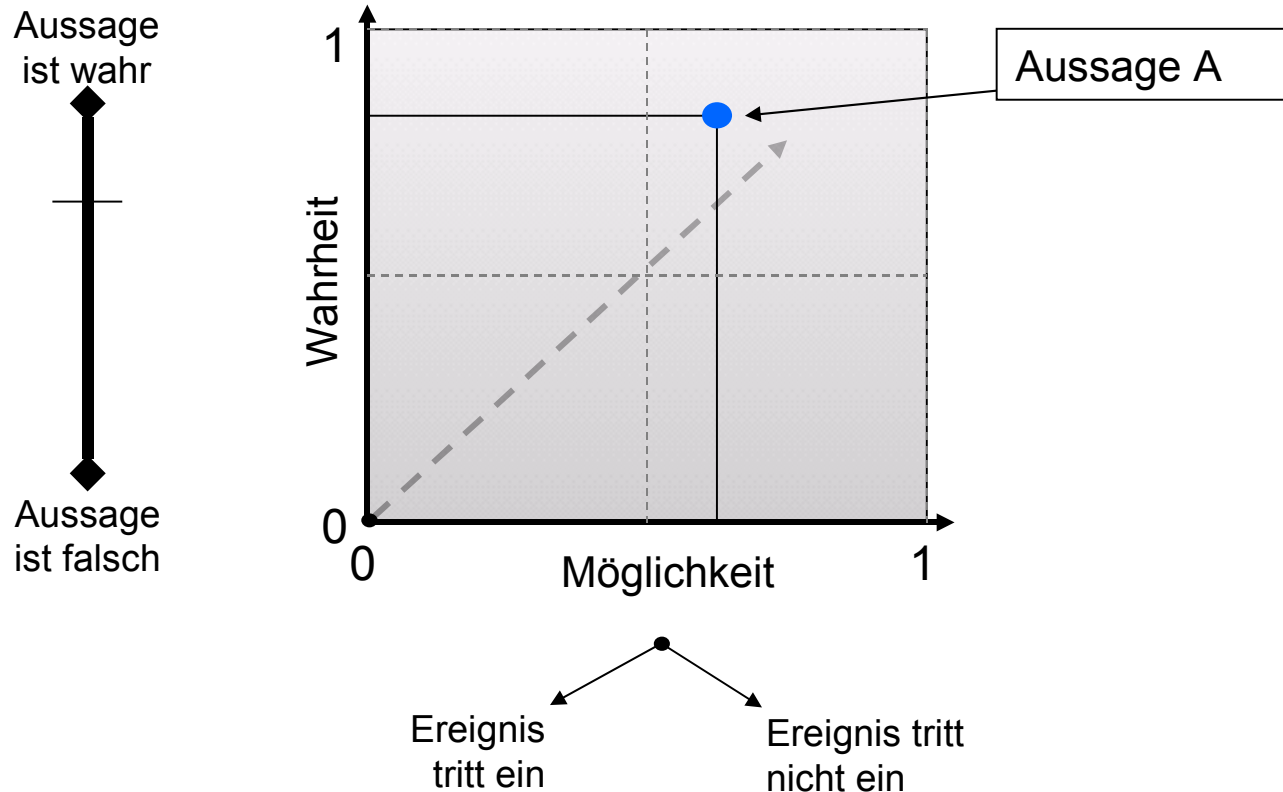
Strukturierte Ausgaben,  
z.B. Präferenzrelationen:

$$y_3 \succ y_1 \succ y_4 \succ y_2$$

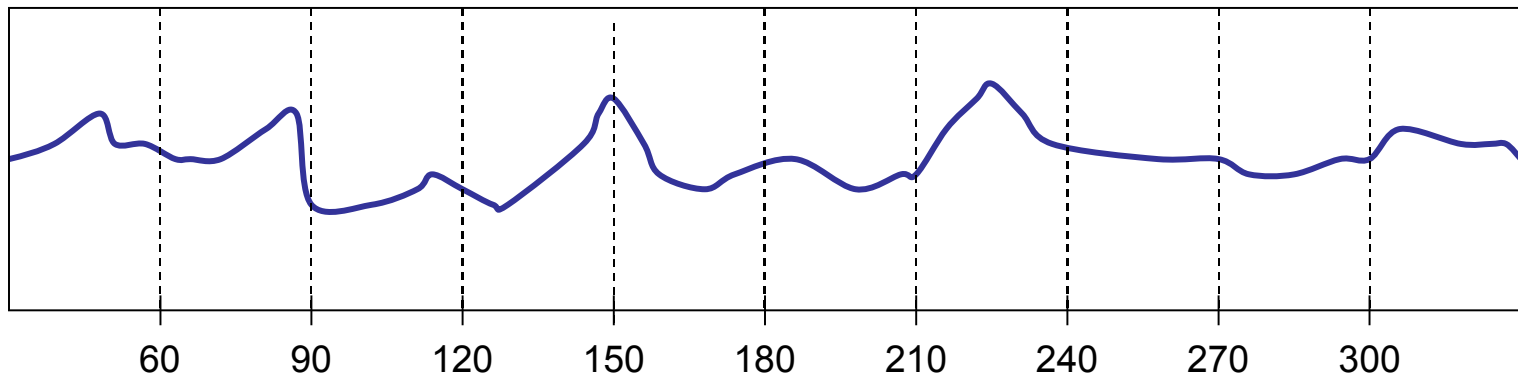
# Unsicherheit, Fuzzy Machine Learning & Data Mining

Unsicherheit ist ein zentraler Aspekt lernfähiger Systeme, der nicht nur die Eingabedaten und die Ausgaben (Prädiktionen) betrifft, sondern auch die interne Repräsentation von Modellen und Mustern.

Unsicherheit tritt in unterschiedlichen Formen auf:

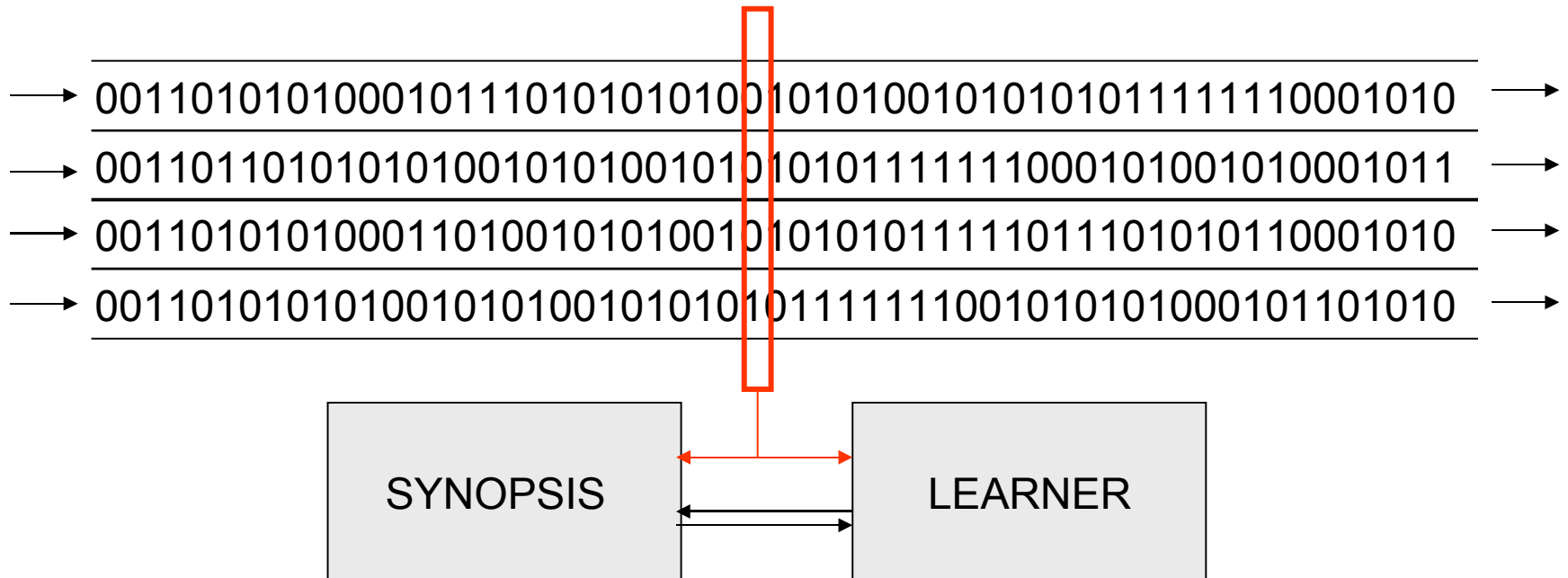


## Beispiel: „Unscharfe“ Muster in Daten



„Etwa 60 Minuten nach einem starken Abfall des Signals beobachtet man häufig einen ungefähr 10-minütigen Anstieg.“

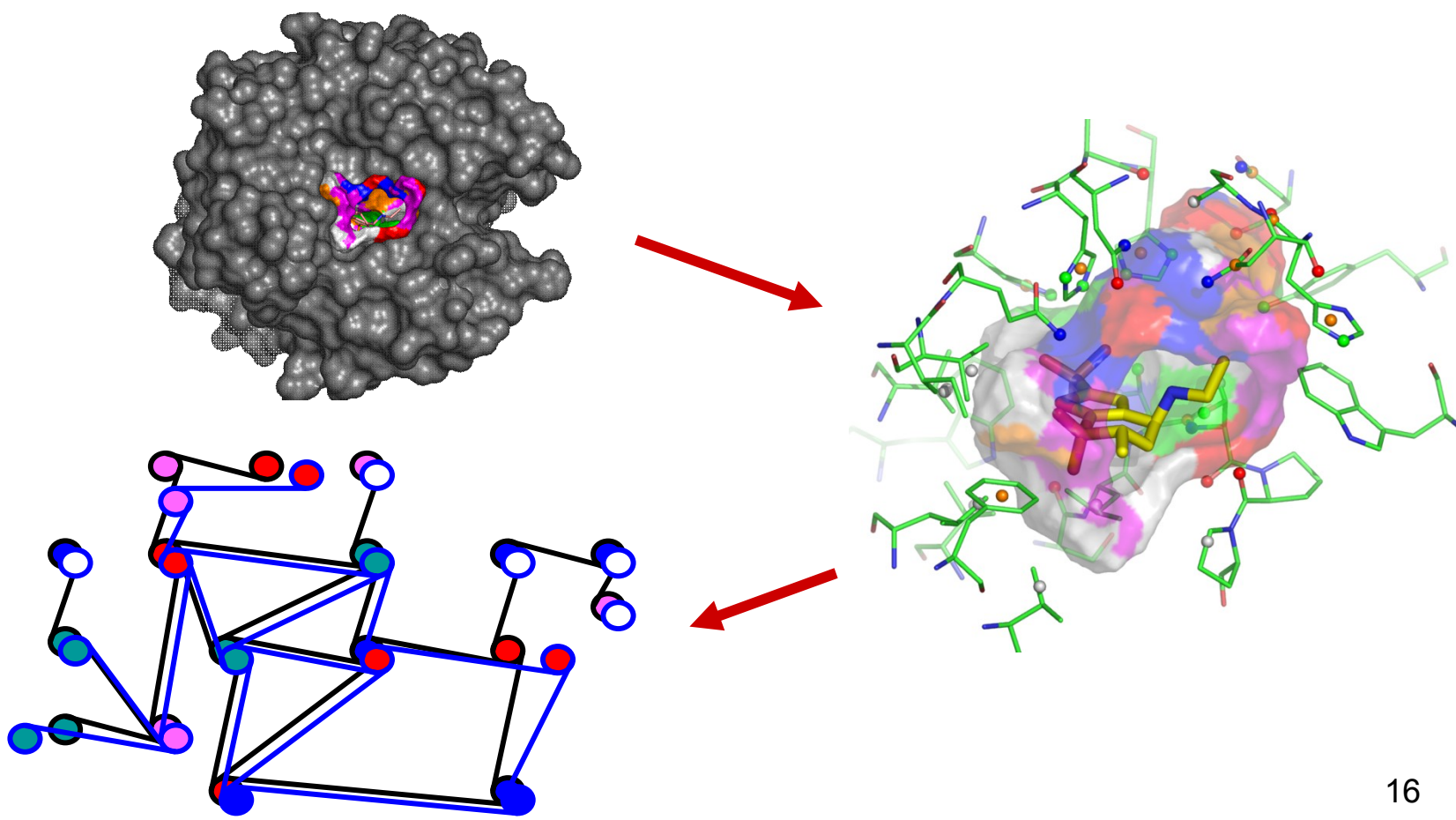
# Lernen auf Datenströmen



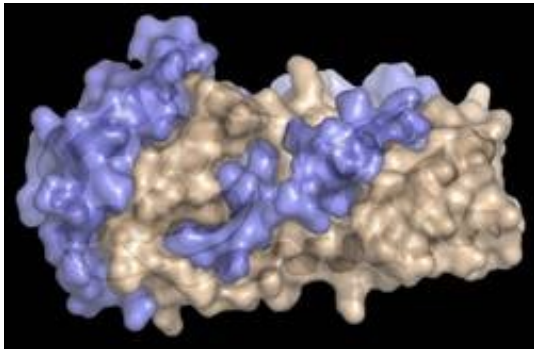
- Vollständiges Speichern der Daten ist nicht möglich
- Lernverfahren müssen inkrementell sein, effizient, skalierbar
- Wichtig: Adaption an Veränderungen des datengenerierenden Prozesses (concept drift/shift)

# Anwendungen

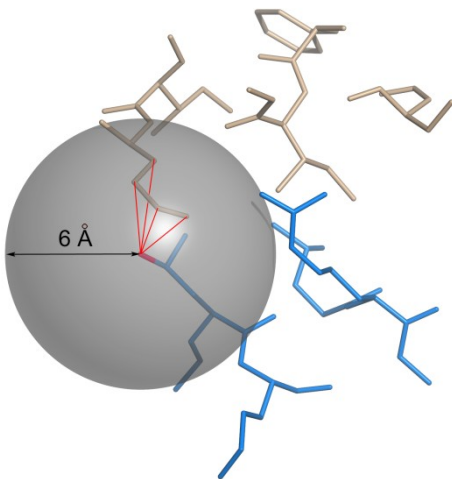
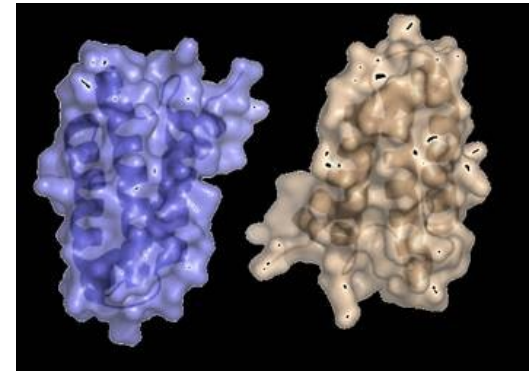
# Graph-based data mining for the structural analysis of protein active sites



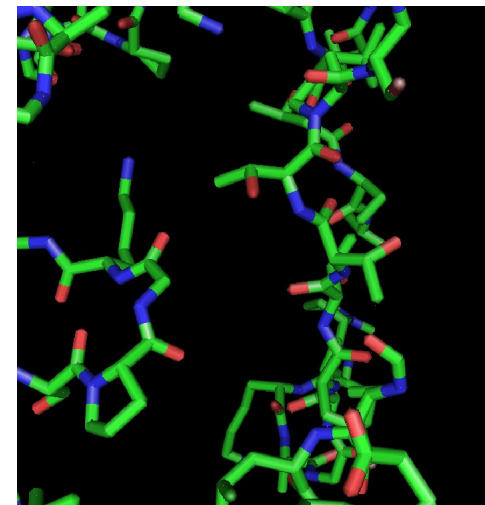
# Machine learning methods for the discrimination of permanent and transient protein complexes



permanent vs. transient

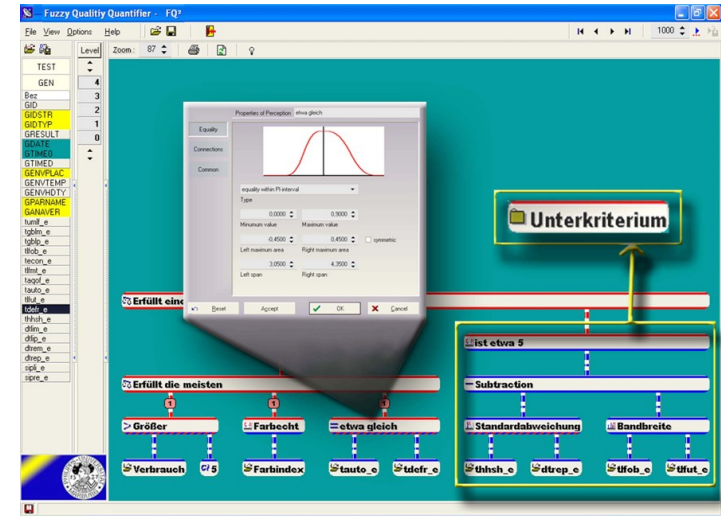


Modeling of protein complexes by means of atomic contact vectors



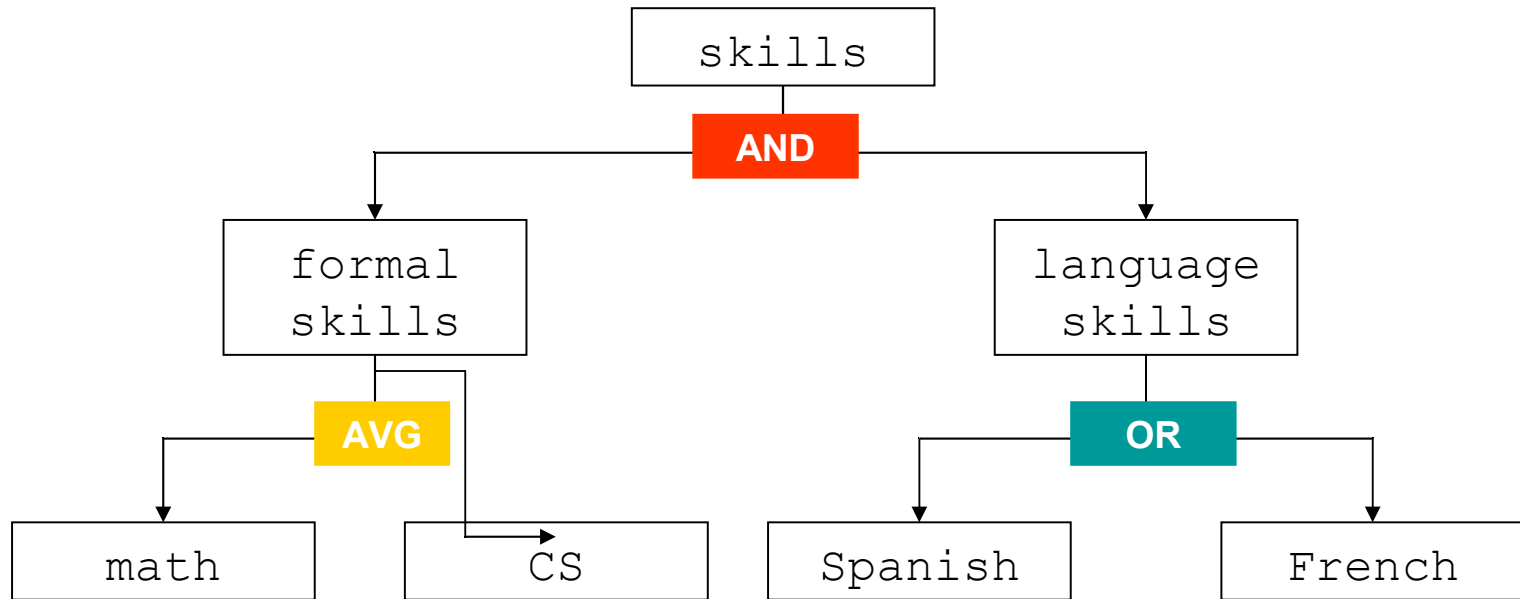
# Qualitätsbewertung von Produkten mithilfe von Fuzzy-Operatorbäumen

- Realisiert Qualitäts-/Nutzenfunktion mit Ausgaben zwischen 0 (schlecht) und 1 (gut)
- Rekursive Zerlegung von Kriterien in Subkriterien
- Aggregation von Subkriterien mithilfe von **parametrisierten** logischen und arithmetischen Operatoren



- Verwendung von „unscharfen“ Bewertungen für Basiskriterien („mehr oder weniger gut“)
- **Lernen** der Parameter von Operatoren, ggf. Topologie

# Qualitätsbewertung von Produkten mithilfe von Fuzzy-Operatorbäumen

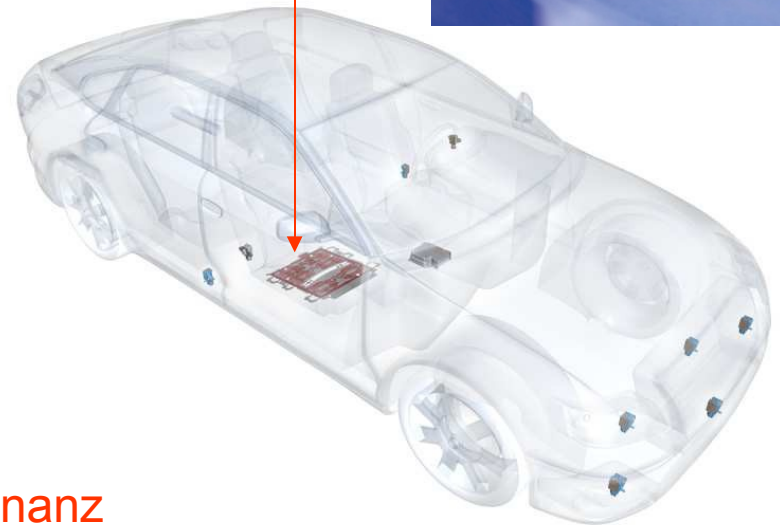


Beispiel: Bewertung von Kandidaten

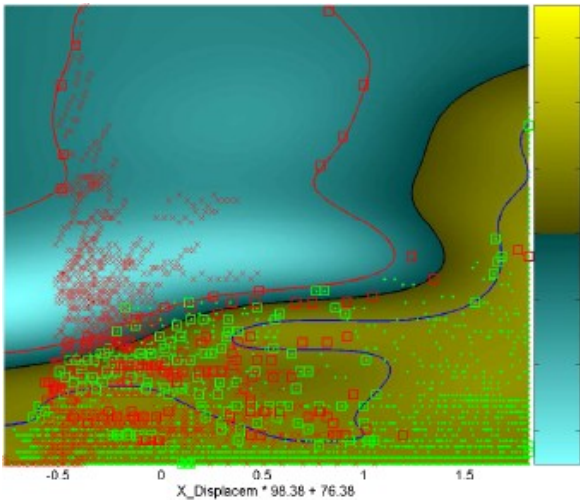
# Maschinelle Lernverfahren zur Erkennung und Diskriminierung von Unfallereignissen (Kalibrierung von Airbag-Steuergeräten)



Sensorik

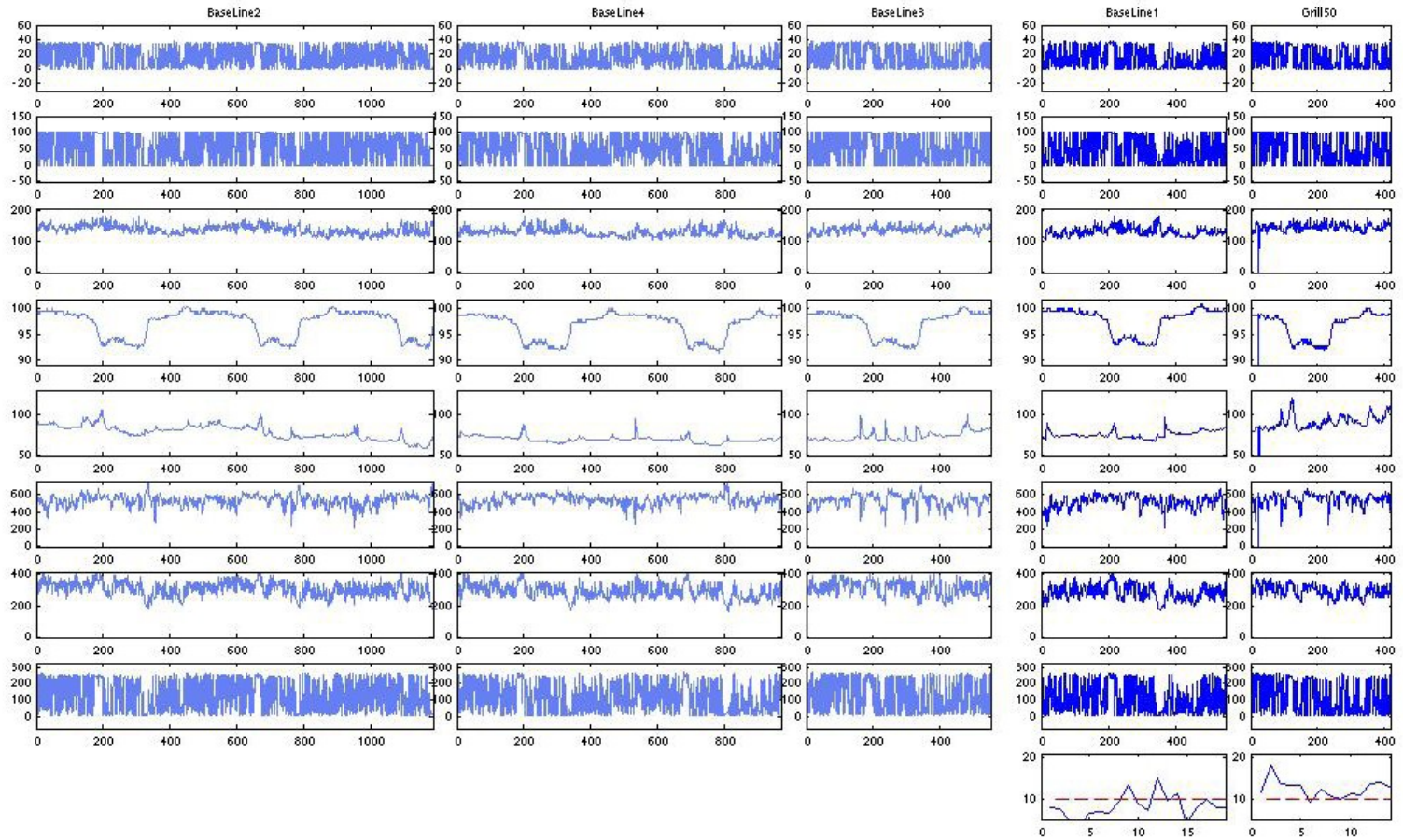


Daten



Diskriminanz  
zwischen „fire“ und  
„no-fire“ Events

# Online Fault Detection in Trucks



## FAZIT:

Das maschinelle Lernen hat einen Reifegrad erreicht, der es, insbesondere in Kombination mit Data Mining Methoden, sehr interessant für industrielle Anwendungen macht.

Das hohe Anwendungspotential wurde in der jüngeren Vergangenheit exemplarisch unter Beweis gestellt, ist aber bei weitem noch nicht ausgeschöpft.