

Vorlesung

**Maschinelles Lernen und Datamining: 2011**

Volker Tresp, Siemens AG

## **Gliederung**

I. Einführung

II. Nichttechnische Perspektiven

» Logik und Philosophie

» Psychologie

» Neurobiologie

III. Technisches Lernen

IV. Einzelheiten zur Vorlesung

# I. Einführung

## *Warum ist Maschinelles Lernen spannend?*

These: Lernen ist eine der drei grundlegenden Methoden zum Entwurf und zur Verbesserung autonomer (intelligenter) Systeme

## 1: *Intelligent Design*

- » Ingenieur: der kenntnisreiche „Uhrmacher“
- » Programmierer (Informatiker)
- » Praktisch alle technische Lösungen basieren auf intelligentem Design
- » „Schlussfolgern“

### **Vorteile:**

- » Explizites Wissen: System ist gut verstanden und kann analysiert werden und analytisch verbessert werden
- » Zeitkonstante: mehrere Jahre

### **Nachteile:**

- » Man braucht den (teuren) intelligenten Designer

## 2: *Evolution*

- » Verbesserung durch Trial and Error
- » Biologische Evolution
  - » der blinde „Uhrmacher“
- » Technische Evolution
  - » Evolutionäre Verbesserung technischer Lösungen

### **Vorteile:**

- » Dumm und einfach
- » Selbst-Optimierend

### **Nachteile:**

- » Zeitkonstante: je nach Organismus Jahre, Jahrhunderte, Jahrtausende
- » Verschwenderisch

## 3: Lernen

### Biologisches Lernen:

- » Lebenslange Optimierung des Verhaltens eines Individuums in der Interaktion mit der Umgebung
- » Schon primitive Tiere sind lernfähig
- » Grundeigenschaft von Tieren („Naturgesetz“)
- » Rückkopplung des Lernerfolges (*Reinforcement*)
- » Zeitkonstante eines Individuums: Tage, Jahre

### Höhere Lebewesen, der Mensch

- » Lernen durch einen Lehrer
- » Langzeitvererbung des Gelernten durch Kultur, Literatur
- » Lernen durch Informationsaufnahme und interner Informationsverarbeitung

## 3: Lernen (Forts.)

### Maschinelles Lernen

- » Im weitesten Sinne: Versuch, biologisches Lernen technisch nutzbar zu machen
- » Selbstständige Optimierung eines technischen Systems durch Interaktion mit der Realität, durch realitätsnahe Daten
- » „Lernen anstatt von Programmieren“



## Charakterisierung des Lernens

- » **Lernfähigkeit ist eine exklusive Eigenschaft von Lebewesen** (und Computern?) (Ferromagnete zeigen Gedächtnis-Effekte?)
  - » Selbst recht primitive Lebewesen zeigen Lernfähigkeit (Adaption)
  
- » Versuch einer Definition des biologischen Lernens:
  - » (Nutzbringende? Permanente?) **Veränderung im** (zentralen?) **Nervensystem** (basierend auf Interaktionen mit der Umwelt?)
  
- » Versuch einer Definition des Maschinellen Lernens:
  - » **Nutzbringende Veränderung in einem technischen System basierend auf der Analyse von Daten oder der Interaktionen mit der Umwelt unter Mithilfe von Lernalgorithmen**

## *Etymologischer Ursprung*

- » Etymologisch ist das Wort "lernen" mit den Wörtern "lehren" und "List" verwandt und gehört zur Wortgruppe von "leisten", das ursprünglich "einer Spur nachgehen, nachspüren" bedeutet
- » Im Gotischen heißt "lais" "ich weiß", bzw. genauer "ich habe nachgespürt"
- » Schon von der Herkunft her hat Lernen etwas mit Spuren hinterlassen zu tun
- » Lernen soll im Gedächtnis ebenso Spuren hinterlassen (siehe letzte Folie: Veränderung im Nervensystem basierend auf Interaktionen mit der Umwelt)

## II. Nichttechnische Perspektiven

- 1: Logik und Philosophie*
- 2: Psychologie*
- 3: Neurobiologie*

# 1: Logik und Philosophie

Wesentliche Prozesse: Deduktion (Schließen) und Induktion (Lernen)

## Deduktion:

- » arbeitet vom Allgemeinen zum Besonderen (*top-down*)
- » Axiome sind gegeben und Sätze (Theoreme) werden abgeleitet
- » Axiome:
  - » Einfache Fakten (Jack ist 180 cm groß)
  - » Komplexe Axiome (Wenn etwas ein Hund ist, dann ist es auch ein Säugetier):
- » Grundlage der klassischen Künstlichen Intelligenz

## Induktion:

- » generalisiert Beobachtungen (*bottom-up*), um allgemeine Schlussfolgerungen zu ziehen oder Theorien zu begründen
- » Schließen über die Wahrheit einer Hypothese aufgrund von Erfahrungen und Beobachtungen
  - » Einfache Fakten (Jack ist 180 cm groß) wie in der Deduktion
  - » **Gelernte Abhängigkeiten ersetzen die in der Deduktion angenommenen komplexen Axiome!**
- » Grundlage des Maschinellen Lernen

## *Philosophie: Rationalismus <-> Empirismus*

**Rationalismus** (von lateinisch: *ratio* = „Vernunft“)

- » Priorität des rationalen Denken
  - » insbesondere in der Wissensfindung, und zwar gegenüber anderen Erkenntnisquellen wie etwa Sinneserfahrung oder religiöser Überlieferung
- » Vertreter: Socrates (ca 470–399B.C.E.), René Descartes (1596–1650), Baruch Spinoza (1632–1677), Gottfried Leibniz (1646–1716), Immanuel Kant (1724–1804)

**Empirismus** (von lateinisch: *empiricus* = der Erfahrung folgend)

- » Alle Erkenntnisse kommen aus der Sinneserfahrung, der Beobachtung oder dem Experiment
- » Der Empirismus steht im Gegensatz zum Rationalismus, der die Vernunft als für den Erkenntnisprozess wesentlich hervorhebt. Der Empirismus hingegen legt seinen Schwerpunkt auf die Erfahrung und sinnliche Wahrnehmung
- » „Nothing is in the understanding, which was not first in the senses“
- » Vertreter: Francis Bacon (1562-1626), John Locke (1632-1704), George Berkeley (1685-1753 ), David Hume (1711-1776)

Karl Popper

## *Philosophie: Scientific Discovery*

- » Ist Induktion schlüssig?
  - » Beispiel: Im Sommer hat es noch nie geschneit -> Im Sommer schneit es nie!?
- » Karl Popper (Sir, 1902-1994). *The Logic of Scientific Discovery*
- » Alles Wissen über die Welt beruht auf Beobachtungen
- » Nach Karl Popper (wie Hume vor ihm) ist induktive Inferenz logisch ungültig und daher als wissenschaftliche Methode nicht haltbar
- » **Wenn keine endliche Menge von Beobachtungen eine Theorie logisch beweisen kann, wie können wir überhaupt eine wissenschaftliche Theorie als wahr akzeptieren?**
- » Popper akzeptiert Empirismus als gültiges Mittel, um Wissen zu vermehren, wenn man akzeptiert, dass Hypothesen nur getestet aber nie bewiesen werden können
- » Falls eine Theorie den Prüfungen standhält, ist sie untermauert

## 2. *Psychologie:* *Psychologie als empirische Wissenschaft*

### **Beginn der wissenschaftlichen Psychologie:**

- » Herrmann von Helmholtz (1821-1894)
- » Wilhelm Wundt (1832-1920) (Assistent von Helmholtz)
  - » Wundt gilt als Begründer der Psychologie als eigenständiger Wissenschaft
  - » Von 1858 bis 1863 arbeitete er als Assistent bei Hermann von Helmholtz. Veröffentlichte fünf Abhandlungen zur Theorie der Sinneswahrnehmungen
- » Gustav Theodor Fechner (1801–1887): Begründer der Psychophysik
  - » Psychophysik bezieht sich auf die gesetzmäßigen Wechselbeziehungen zwischen subjektivem psychischen (mentalen) Erleben und quantitativ messbaren, also objektiven physikalischen Reizen als den auslösenden Prozessen

# Verhaltenswissenschaft

## Behaviorismus (1920-1960)

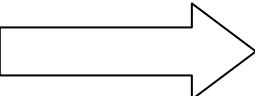
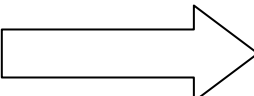
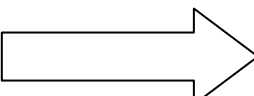
- » „*Belief in the existence of consciousness goes back to the ancient days of superstition and magic*“
- » Ablehnung einer Theorie, die mentale Zustände annehmen muss
- » Lebewesen als Maschine ("*black box*")
  - » Innere Struktur ist unwichtig
- » Die Funktionsweise kann nur aus dem Input (Reize) und dem Output (Reaktion) erschlossen werden
- » Vertreter: Iwan Pawlow (1849-1936), John Watson (1878-1958)



## *Lernen in der Psychologie*

- » **Definition: dauerhafte Änderung des Verhaltens**, die **durch Übung** (im Gegensatz etwa Reifung, Prägung oder Krankheit) **erfolgt**
- » **Habituation**: Lernen, einen Reiz zu ignorieren (vgl. Sensitivierung)
- » **Klassische Konditionierung**: Nach dem Lernvorgang mit dem Reiz Futter erfolgt der Speichelfluss auf einen Ton auch ohne den Reiz
- » **Operative Konditionierung**: Lernen, dass einer Aktion eine Konsequenz folgt
  - » Eine Ratte lernt es, ihren Käfig zu öffnen, wenn es nach Erfolg durch Futter belohnt wird
- » **Komplexes Lernen**: Lernen, das über das Bilden von Assoziationen hinausgeht, z.B. die Anwendung einer Strategie zur Problemlösung oder die Ausbildung der geistigen Landkarte einer Umgebung

## *Klassisches Konditionieren (Pawlow)*

Unkonditionierter Stimulus (Futter)		Unkonditionierte Antwort (Speichelfluss)		
Unkonditionierter Stimulus (Futter) mit Konditionierten Stimulus (Glocke)		Unkonditionierte Antwort (Speichelfluss)		
Konditionierter Stimulus (Glocke)		Konditionierter Antwort (Speichelfluss)		

## *Kognition und Kognitivisten*

- » Versuch, die inneren Vorgänge in der Black Box zu verstehen
- » Gegenpol zum Behaviorismus
- » Menschliches Verhalten ist **mehr als eine Reiz-Reaktions-Kette** (Passivität)
- » Entwicklung ist ein aktiver Prozess eines Subjektes, das mit **Erkenntnisfunktionen** ausgestattet ist
- » Dieses Subjekt baut Erkenntnis durch die **aktive Auseinandersetzung mit der Umwelt** auf Handeln wird von **Denkvorgängen** bestimmt und damit nicht nur von den äußeren Gegebenheiten (z. B. können Kinder Verhaltensweisen entwickeln, die in ihrer Umwelt gar nicht vorkommen)
- » Das Bindeglied zwischen Reiz/Umwelt und Reaktion/Verhalten ist die **kognitive Repräsentation**
- » Williams James (1842-1910), Herrmann von Helmholtz (1821-1894), Frederik Bartlett (1886-1969), George Miller (*The magic number seven*, 1956), Noam Chomsky (*Three Models of Language*, 1956)

### 3. Neurobiologie

- » Lernvorgang: „Neuronale Plastizität“
- » Es muss sich tatsächlich etwas physikalisch verändern, wenn etwas gelernt / abgespeichert wird!
- » Zentraler Lernmechanismus: Synapsen ändern ihre Effizienz
  - » Es gilt inzwischen als gesichert, dass die synaptische Plastizität die Voraussetzung für Lernen und Gedächtnis ist
- » *Short-term plasticity*: Die Änderung der Übertragungsstärke hält einige Millisekunden bis höchstens einige Minuten an
- » *Long-term plasticity*: Die Stärke der Übertragung ändert sich für viele Minuten bis einige Stunden, möglicherweise lebenslang

## Neuron

- » Ruhemembranpotential: -70 mV.
- » Depolarisation:  $> -50\text{mV}$ 
  - » -> Öffnung der Natrium-Kanäle, Auslösung eines Aktionspotentials
- » Refraktärphase: Während dieser Zeit kann kein Aktionspotenzial ausgelöst werden, unabhängig von der Reizstärke (ca. 2ms)
- » Systemtheorie: *leaky integrator*

# Synapse

**Präsynaptisch:** Eine in der signalgebenden Nervenzelle erzeugte elektrische Entladung, das Aktionspotential, führt dazu, dass an der präsynaptischen Membran Neurotransmitter ausgeschüttet werden

**Postsynaptisch:** Öffnung von Ionenkanälen und damit Änderung des Membranpotentials der postsynaptischen Nervenzelle

## *Beispiel: Lernvorgänge bei Aplysia*

- » Eric Richard Kandel (\* 7. November 1929 in Wien): US-amerikanischer Neurowissenschaftler österreichischer Herkunft
- » Studien am Kalifornischen Seehasen (*Aplysia californica*), einer Meeresschnecke
- » Nobelpreis 2000
- » Kiemen-Rückzugsreflex mit 24 Sensor-Neuronen und 6 Motor-Neuronen
- » Habituation (Gewöhnung)
  - » Verringerung der Neurotransmitter auf wiederholtem Reiz
- » Sensitization (Sensibilisierung):
  - » Erhöhung der Neurotransmitter auf schädlichem Reiz
- » Association:
  - » Licht/Elektrischer Schock

Die Synapse (links) beeinflusst eine zweite Synapse (unten)

### **Kurzzeitgedächtnis (Minuten):**

- » ein schwacher Reiz bewirkt die Phosphorylierung von Proteinen der Ionenkanäle (schwacher Pfeil, links), was zur verstärkten Ausschüttung von Neurotransmittern führt

### **Langzeitgedächtnis (Wochen):**

- » ein starker Reiz (dicker Pfeil, links) führt zu einem erhöhten Anteil des Botenstoffs cAMP (Second Messenger bei der zellulären Signaltransduktion), was zur Verstärkung der Proteinkinase führt
- » Dies beeinflusst den Zellkern, wodurch neue Proteine erzeugt werden
- » Dies kann zur Veränderung der Form und Funktion der Synapse führen: die Effizienz ist erhöht und mehr Neurotransmitter werden erzeugt



## *Hebb'sches Lernen in Psychologie und Neurophysiologie*

- » Die Ergebnisse von Kandel liefern weitere Evidenz für die Theorie des Hebb'sches Lernen:
  - » Wiederholte Aktivierung (Feuern) eines präsynaptischen Neurons und des postsynaptischen Neurons über eine bestimmte Synapse erhöht deren Effizienz
- » Hebb versuchte damit das klassische Konditionieren neurologisch zu erklären
- » Hebb'sches Lernen nachgewiesen: Neuronen im Hippocampus von Säugetieren
- » Bei Hebb wird Lernen auf einer viel abstrakteren Ebene beschrieben als bei Kandel
  - » Offene Frage: wie viel der biologischen Einzelheiten kann man ignorieren, ohne die Essenz des Lernens zu verlieren?

## *Zusammenfassung: Bezug des Nichttechnischen Lernens zum Maschinellen Lernen*

- » Logik: Formalisierung der empirischen Sichtweise (Popper)
- » Philosophie: Tradition des Empirismus
- » Psychologie
  - » In der empirische Tradition der Behavioristen
  - » Aber auch: Wechselseitiger Austausch mit den Kognitivisten in der Formulierung von internen Modellen
- » Biologie
  - » Lernen in der Neurophysiologie noch wenig verstanden
  - » Neuroinformatik: Inspiration durch die Biologie
  - » Maschinelles Lernen produziert Lernalgorithmen und Lernmodelle, die biologisch interessant sind

## III. Maschinelles Lernen

- » Vision: ein System, welches sich selbst durch Interaktion mit der Umwelt verbessert
- » Oder wie vorher definiert: Nutzbringende Veränderung in einem technischen System basierend auf der Analyse von Daten oder der Interaktionen mit der Umwelt unter Mithilfe von Lernalgorithmen

## *Technisches Lernen*

### **Vor dem Computerzeitalter:**

- » Statistik: Ableiten von Wissen aus Beobachtungen

### **Mit dem Computerzeitalter:**

1. Statistik
2. Neuroinformatik; Adaptive Systeme; Kybernetik (Mustererkennung, Spracherkennung)
3. Symbolisches Maschinelles Lernen in der Künstlichen Intelligenz
4. Revival der Neuroinformatik
5. Statistisches Maschinelles Lernen
6. Datamining

# 1: Statistik

## Thomas Bayes (Rev., 1701 -1761)

- » Schließen über Hypothesen basierend auf Beobachtungen
- » Grad des Wahrheitsgehalt einer Hypothese
- »  $P(H=1)$ : Annahme über den Wahrheitsgehalt der Hypothese H (a priori Annahme)
- »  $P(D|H=1)$ : Wahrscheinlichkeit der Beobachtung (Data) D, falls Hypothese H wahr ist (Likelihood)
- » Satz von Bayes:
  - »  $P(H=1|D) \sim P(D|H=1) P(H=1)$   
(a posteriori Wahrscheinlichkeit der Hypothese)

## Subjektive Wahrscheinlichkeiten

- » Subjektive Wahrscheinlichkeit: Meiner Meinung nach liegt die Wahrscheinlichkeit, dass Partei X die nächsten Wahlen gewinnt bei 45%
  - » Beachte: die Wahl ist ein einmaliger Vorgang
- » Cox (1946): Wenn man seinen Überzeugungen (*Beliefs*) Zahlen zuordnen will, kommt man unter wenigen Konsistenzannahmen auf den Bayesschen Formalismus
  - » *Wenn 1 der Gewissheit entspricht, dass ein Ereignis eintritt, und 0 der Gewissheit, dass ein Ereignis nicht eintritt, und Werte dazwischen den entsprechenden (Un)Sicherheiten entsprechen, dann verhalten sich diese Zahlen genau wie Wahrscheinlichkeiten*

## *Kritik an der Bayesschen Statistik*

**Karl Pearson (1857 – 1936)** now considered the founder of modern statistics

» *"I felt like a buccaneer of Drake's days -... I interpreted that sentence of Galton to mean that there was a category broader than causation, namely correlation, of which causation was only the limit, and that this new conception of correlation brought psychology, anthropology, medicine, and sociology in large parts into the field of mathematical treatment."*

**Sir Ronald Aylmer Fisher** (Sir, 1890-1962)

- » Kritik an der Rolle subjektiver Wahrscheinlichkeiten: Frequentisten beschränken sich auf Aussagen über wiederholbare Experimente
- » Es wird bewertet, ob die vorliegende Beobachtung einer Hypothese widerspricht, aber es wird nicht die Wahrscheinlichkeit eine Hypothese bewertet

**Egon Pearson** (1895-1980) Sohn von K.P., **Jerzy Neyman** (1894-1981)

## 2: Neuroinformatik

- » Zunächst: Fokus auf Ausdruckskraft neuronaler Strukturen und nicht deren Lernfähigkeit
- » *McCulloch and Pitts (1943)*: erster Versuch, Gehirnmechanismen über einfache Recheneinheiten zu formalisieren (Netzwerke von einfachen logischen Einheiten)



## *Ausdruckskraft neuronaler Strukturen*

- *John v. Neumann (1956): untersuchte die Fehlertoleranz neuronaler Netze (“reliable computing with unreliable elements”)*
- *John v. Neumann (1958): Computer and the Brain*
- *John von Neumann concludes that the brain operates in part digitally, in part analogically, but uses a peculiar statistical language unlike that employed in the operation of man-made computers*

## *Lernfähigkeit neuronaler Strukturen*

- » *Hebb (1949): Repeated activation of one neuron by another, across a particular synapse, increases its conductance (Hebb'sche Regel); "Neurons that fire together wire together"*
- » Hebb versuchte damit das klassische Konditionieren neurologisch zu erklären.
- *Wiener (1949): Cybernetics, or control and communications in the animal and the machine*  
*The whole world -- even the universe -- could be seen as **one big feedback system subject to the relentless advance of entropy**, which subverts the exchange of messages that is essential to continued existence (Wiener, 1954). Buch: *Cybernetics or Control and Communication in the Animal and the Machine* (1948)*

## *Lernfähigkeit neuronaler Strukturen: Assoziativspeicher*

- » W. K. Taylor (1950er), Karl Steinbuch (1961)
- » Assoziativspeicher, Lernmatrix
- » Bezug zu: Hebb'schem Lernen

## *Perzeptron und ADALINE*

- » Minsky entwickelte 1954 im Rahmen seiner Dissertation den ersten neuronalen Computer, den er "Snark" nannte
- » Rosenblatt entwickelte 1958 die Perzeptron Lernregel und formulierte einen Konvergenzbeweis; Mark I Perceptron
- » Widrow und Hoff entwickelten 1960 das ADALINE (Adaptives Lineares Element) (Modem)
- » Minsky und Papert veröffentlichten 1969 das Buch „Perzeptrons“ und zeigten Grenzen des Perzeptrons und des Adalines auf (Exclusive-Or Problem)

## *Die (erste) Blütezeit der Künstlichen Intelligenz: Neuroinformatik und Mustererkennung in ihrer Eiszeit*

- » Unter anderem wegen des Buchs von Minsky und Papert wurden (im Westen) Fördergelder fast ausschließlich in das erblühende Gebiet der Künstlichen Intelligenz (KI) gesteckt
- » Zum Studium von lernenden Systemen standen keine Mittel mehr zur Verfügung
- » Es folgt ein kurzer Überblick über die KI
  - » Die KI formuliert oft interessante Probleme, die dem Maschinellen Lernen zugänglich sind
- » Wenn wir in den nächsten Folien über KI reden, so meinen wir die KI der 60er, 70er und 80er Jahre und nicht die späteren und heutigen Entwicklungen

### *3: Künstliche Intelligenz und Symbolisches Maschinelles Lernen in der Künstlichen Intelligenz*

Etwas vereinfacht:

- » Maschinelles Lernen ist „Intelligenz durch Lernen“
  - » Induktion
  - » Empirismus
  
- » (Klassische) KI ist „Intelligenz durch Design“: Fakten und Axiome erlauben über logische Inferenz die Ableitung (den Beweis) neuer Aussagen
  - » Deduktion

## *Die Vier Ansätze der KI*

- » Ununterscheidbarkeit vom menschlichen Agieren (Turing Test)
  - » Sprache, Wissensrepräsentation, Schlussfolgern, Lernen, Vision, Robotik
- » Menschliches Denken verstehen: Kognitives Modellieren
- » Rationales (nichtnotwendig menschliches) Denken: Logik
- » Rationales (nichtnotwendig menschliches) Handeln: Agenten

## *Künstliche Intelligenz: Vorgeschichte*

### **Philosophische Basis:**

- » Wiener Schule (Rudolf Carnap (1891-1970))
  - » **Logischer Positivismus: Das gesamte Wissen kann charakterisiert werden durch logische Theorien ...**
- » Ludwig Wittgenstein (1889-1951), Bertrand Russel (1872-1970)



## *Geburt der KI: Dartmouth Workshop (1956)*

**John McCarthy** (Dartmouth, dann MIT) (1927-)

» Namensgeber: AI/KI; Erfinder von LISP

**Marvin Minsky** (1927-) (MIT)

» SAINT (calculus integration); ANALOGY (geometric analogy); STUDENT (algebra); Blocks World; The Society of Mind (1985); eher Anti-Logik und Anti-Neuro

**Claude Shannon** (1916-2001) (Bell Labs) Erfinder der Informationstheorie

**Arthur Samuel** (1901-1990) IBM; checkers program

**Ray Solomonoff** (1926-2009) (MIT) Begründer der Algorithmic Probability

**John von Neumann** Institute for Advanced Study; Begründer der Spieltheorie

**Allen Newell** (1927-1992)(CMU), **Herbert Simon** (1916-2001) (CMU) (Nobel):

» General Problem Solver (GPS): ein Programm, das mit einfachen Methoden beliebige Probleme lösen können sollte (nach zehnjähriger Entw. eingestellt)

» Ihre Auffassung, dass Intelligenz unabhängig von der Trägersubstanz ist, wird von den Vertretern der starken KI-These geteilt

**Nathaniel Rochester (IBM), Trenchard More (Princeton), Oliver Selfridge (MIT), Cliff Shaw**

## *Kritik an der Künstliche Intelligenz*

### **Früher Enthusiasmus (1952-1969)**

- » Insbesondere die Anfangsphase der KI war geprägt durch eine fast grenzenlose Erwartungshaltung im Hinblick auf die Fähigkeit von Computern, "Aufgaben zu lösen, zu deren Lösung Intelligenz notwendig ist, wenn sie vom Menschen durchgeführt werden" (Minsky).
- » Herbert Simon (1957)
  - » innerhalb der nächsten 10 Jahre wird ein Computer Schachweltmeister werden und einen wichtigen mathematischen Satz entdecken beweisen würde
  - » In don't want to chock you ... There are now in the world machines that think ... in a visible future the range of problems they can handle will be coextensive with the range to which the human mind has been applied...
- » McCarthy schlug 1958 vor, das gesamte menschliche Wissen in eine homogene, formale Darstellungsform, die Prädikatenlogik 1. Stufe, zu bringen. Die Idee war, Theorem-Beweiser zu konstruieren, die symbolische Ausdrücke zusammensetzen, um über das Wissen der Welt zu diskutieren.

### **Erste Realitäts-Dosis (1966-1973)**

- » Translation of Russian into English was stopped: "the spirit is willing but the flesh is weak" became "the wodka is good but the meat is rotten"
- » Reasoning did not scale up

## Wissensbasierte Systeme

### Wissensbasierte Systeme (1969-1979)

- » Expertensystemen: In einem Expertensystem wird Expertenwissen formal, beispielsweise in Form von Regeln, repräsentiert und auf gegebene Fakten angewandt, um auf neue bis dahin nicht bekannte Fakten zu schließen.
- » Bruce Buchanan: *DENDRAL* (1969); inferring molecular structure from mass spectroscopy data; first knowledge intensive system
- » Ed Feigenbaum (Stanford): *Heuristic Programming Project* (HPP)
- » Feigenbaum, Buchanan, Edward Shortliffe; *MYCIN*: Diagnose blood infections; extensive interviewing of experts; uncertainty factors
- » Progress in NLP: Eugene Charniak, Roger Shank
- » Prolog

### KI wird Industrie (1980- wenige Jahre später)

- » McDermott: *R1* (DEC, 1982); Konfiguration von neuen Computersystemen; jede große Firma hatte eine AI Gruppe
- » Japan (1981) Fifth Generation Project; 10-Jahres Projekt zum Bau Intelligenter Computer basierend auf PROLOG

**Kollaps (1984)** vieler Silicon Valley Start-Ups (Anfang des KI Winters)

## Maschinelles Lernen in der KI

- » Maschinelles Lernen in der KI eher wurde eher stiefmütterlich behandelt (*only deductive inference is sound*)
- » Man wollte sich sehen im Kontrast zu Statistik und Wahrscheinlichkeitslehre
- » Fokussierung auf Symbolisches Maschinelles Lernen
- » Aus dieser Tradition entstanden
  - » Fallbasiertes Schließen (*case-based reasoning*, CBR) (Schank, 1977)
  - » Lernen von Entscheidungsbäumen (Ross Quinlan's ID3, 1979, Rivest)
  - » *Inductive Logic Programming* (Stephen Muggleton, 1991)
    - » Intuitiver Reiz: Man versucht (einfache) logische Regeln zu extrahieren
    - » Mächtigkeit: Man kann (*first-order*) Prolog Regeln erlernen (Turing-äquivalent)

## 4: *Revival der Neuroinformatik*

- Die Aktivitäten in der Neuroinformatik fanden erst Mitte der 80er Jahre wieder mehr Interesse; Ende des Winters für Neuronale Algorithmen / Mustererkennung (1969-1982)
- Hype der KI war vorüber: vielleicht kommt es doch auf das Substrat an?
- Lernen steht im Vordergrund; der regelbasierte Zugang wird verlassen
- Faszination Gehirn: trotz der biologischen Komplexität sollte es doch ein einfaches Organisationsprinzip geben, welches über Lernen zu intelligenten Leistungen führt; vielleicht kann Intelligenz nur über Lernen erreicht werden?
- **Technisch konkurrenzfähige Lösungen konnten relativ leicht erreicht werden**

## *Revival der Assoziativspeicher*

- » **John Hopfield (1982, 1984):** *Neural networks and physical systems with emergent collective computational abilities*
- » Systemleistungen:
  - » Assoziativspeicher (Hebb'sches Lernen)
  - » kombinatorische Optimierung
- » Zugang von Seiten der statistischen Physik (Spin-Gläser)
- » Nichtlinear, parallel, fehlertolerant, rückgekoppelt
- » Implementierung als Optische Rechner?
- » Bezug zur Funktionsweise des Gehirn
- » Im Endeffekt: technisch nicht überlegen zu konkurrierenden Lösungen
- » Unabhängig und zum Teil früher: Stephen Grossberg, Teuvo Kohonen

## *Erweiterungen des Perzeptrons*

**Ackley, Hinton, Sejnowsky (1985): *Boltzmann Machine***

- » Diskriminatives Lernen
- » Theoretisch sehr interessant aber nicht so praktikabel wie das folgende MLP

**Rumelhart, Hinton, Williams (1986): Multi-layer Perzeptron (MLP)**

- » MLP: ein relativ leicht handhabbares mächtiges Werkzeug zur Modellierung hochdimensionaler nichtlinearer Abhängigkeiten
  - » Lösung des *exclusive-or-Problems*, Nettek
  - » MLP: überlegenes Modellierungstool für hochdimensionale Probleme
  - » *Durchbruch der Neuroinformatik*
  - » Interesse in der Statistik
- 
- » Ab 1988 ein gewisser Hype („Lernen anstatt Programmieren“)

## 4: *Statistisches Maschinelles Lernen*

- » Im Gegensatz zur klassischen KI gab es großen Teilen der Neuroinformatik das Interesse, die Arbeiten in einen statistischen Rahmen zu fassen
- » Daraus entstand gegen Ende der 90er Jahre das Statistische Maschinelle Lernen, welches bis heute das Maschinelle Lernen dominiert und ebenso einen wichtigen Platz in der KI einnimmt
- » Konvergenz von KI und Maschinellem Lernen: die KI formuliert die Langzeitziele, für die häufig Maschinelles Lernen Lösungen liefert



## 5: *Datamining*

### » **Statistik**

- » Statistische Analyse eines Experimentes steht im Fordergrund
- » Daten sind oft Mangelware
- » Typisches Betätigungsfeld: Biostatistik: verbessert mein neues Medikament die Heilungsaussichten signifikant?

### » **Maschinelles Lernen**

- » Traum von einer intelligenten lernenden Maschine

### » **Datamining**

- » Basis: große (Geschäfts-) Datenbanken
  - » Heute zunehmend auch das WWW
- » Auswertung der Daten zum Customer Relation Management (CRM), Betrugserkennung und für viele andere Zwecke
- » Extraktion oder „Mining“ von Wissen aus großen Datenmengen
- » Hohe Anzahl von Variablen

## *Datamining*

- » Datamining als Teil des KDD - Prozesses (**Knowledge Discovery in Databases** (KDD))
- » Wichtige Daten:
  - » 1989: IJCAI Workshop zu KDD
  - » 1995 KDD Konferenz
  - » 1998 SIG KDD Konferenz
- » Nützliche Methoden anderer Gebiete werden übernommen
  - » Statistik
  - » Maschinelles Lernen
  - » Neuroinformatik
- » Neue Methoden werden entwickelt
  - » Association Rules (Agrawal)
  - » DBSCAN (Ester, Kriegel, Sander, XU) ...

## *Die Gegenwart*

- » Obwohl jedes Gebiet seine eigenen Schwerpunkte behalten hat, gibt es einen gegenseitigen intensiven Austausch
- » Statistik und Wahrscheinlichkeitslehre dominieren als technische Grundlage
- » Maschinelles Lernen ist ein gewisser Innovationstreiber

## *Aktuelle Aktivitäten*

- » Die großen Spieler im Informationszeitalter wie Microsoft, Google, Yahoo stellen im beträchtlichen Maße Personen mit einem Hintergrund in Maschinellern Lernen / Data Mining ein
- » Microsoft hat mindestens zwei starke Forschungsgruppen zu dem Thema; der Job-Markt ist international exzellent
- » ML hat zunehmend *Impact* auf andere Gebiete wie Vision, rechnerbasierte Spracherkennung, Information Retrieval, Bioinformatik, ...
- » Beispiel für eigene Aktivitäten: Maschinelles Lernen im Semantischen Web / Linked Data in zwei geförderten Projekten THESEUS (BMWi), LarKC (EU)

## *Die Vorlesung*

- » Technischen Grundlagen und Werkzeuge, die heute im Fokus stehen mit entsprechenden Anwendungen
  
- » Mathematische Voraussetzungen (werden eingeführt):
  - » Lineare Algebra (Vektoren, Matrizen, ...)
  - » Wahrscheinlichkeitslehre (Elementare Manipulationen mit Wahrscheinlichkeiten, ...)
  - » Statistik (Bayessche Statistik, Frequentistische Statistik)

# Literatur

## Vorlesung

- *Machine Learning*. Tom Mitchel: McGraw-Hill
- *The Elements of Statistical Learning: Data mining, Inference and Prediction*. Hastie, Tibshirani, Friedman: Springer (2nd Ed.) **Frei verfügbar unter: <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>**
- *Pattern Classification*. Duda, Hart, Storck: Wiley
- *Neural Networks for Pattern Recognition*. Bishop: Clarendon Press
- *Pattern Recognition and Machine Learning*. Bishop: Springer
- *Data Mining: Concepts and Techniques*. Han and Kamber: Morgan Kaufmann (2<sup>nd</sup> ed.)
- *Artificial Intelligence-a Modern Approach*. Russel and Norvig, Prentice Hall

## Literatur

### Zeitreihenanalyse

- *Time Series Analysis*. Hamilton

### Reinforcement Lernen und Spieltheorie

- *Reinforcement Learning: an Introduction*. Sutton and Barto: MIT Press
- *Fun and Games: A Text on game Theory*. Binmore and Linster, Houghton Mifflin

### **Statistik**

- *Bayesian Data Analysis*. Gelman, Carlin, Stern, Rubin: Chapman
- *Heckerman's Tutorial*:  
[http://research.microsoft.com/research/pubs/view.aspx?msr\\_tr\\_id=MSR-TR-95-06](http://research.microsoft.com/research/pubs/view.aspx?msr_tr_id=MSR-TR-95-06)
- *Statistik*. Fahrmeir, Kuenstler, Pigeot, Tutz: Springer
- *Introduction to Mathematical Statistics*. Hogg, Craig: Prentice Hall
- *Principles of Statistics*. Bulmer: Dover
- *Probability, Random Variables and Stochastic Processes*. Papoulis, McGraw, Hill

## Zeit und Ort

Veranstaltung	Zeit	Ort	Beginn
Vorlesung	Mi, 9.00 - 12.00 Uhr	Raum C 113 (Theresienstr. 41)	04.05.2011
Übung	Do, 12.00 - 14.00 Uhr	Raum 067 (Oettingenstr. 67)	12.05.2011
	Fr, 14.00 - 16.00 Uhr	Raum 067 (Oettingenstr. 67)	13.05.2011

## Planung

Vorlesung		Übung	
Datum	Thema	Datum	Blatt
04.05.11	Einführung	05./06.05.11	entfällt
11.05.11	Das Perceptron / Wiederholung: Lineare Algebra / Lineare Regression	12./13.05.11	optional: <span style="border: 1px dashed black; padding: 2px;">R</span> - Einführung ( <a href="#">Sourcen</a> )
18.05.11	Basisfunktionen / Neuronale Netze	19./20.05.11	<a href="#">Blatt 1</a>
25.05.10	Kerne / Wiederholung: Wahrscheinlichkeitslehre	26./27.05.11	Blatt 2
01.06.11	Rekurrente Neuronale Netze und Zeitreihen	02./03.06.11	Blatt 3
08.06.11	Frequentistische Statistik und Bayes'sche Statistik / Lineare Klassifikatoren	09./10.06.11	Blatt 4
15.06.11	Fortsetzung: Lineare Klassifikatoren / Support Vector Machines	16./17.06.11	Blatt 5
22.06.11	Modellvergleich	23./24.06.11	Blatt 6
29.06.11	Hauptkomponentenanalyse	30.06./01.07.11	Blatt 7
06.07.11	Instanzbasiertes Lernen	07./08.07.11	Blatt 8
13.07.11	Bayes'sche Netze	14./15.07.11	Blatt 9
20.07.11	Reinforcement Learning	21./22.07.11	Blatt 10
27.07.11	vor.: Klausur	28./29.07.11	entfällt