

Vorlesung

**Maschinelles Lernen und Datamining: 2009**

Volker Tresp, Siemens AG

# *Warum ist Maschinelles Lernen spannend?*

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

# 1: *Intelligent Design*

- Ingenieur: der kenntnisreiche „Uhrmacher“
- Programmierer (Informatiker)
- Im Grunde fast alle technische Lösungen basieren auf intelligentem Design

## **Vorteile:**

- Explizites Wissen
- System ist gut verstanden und kann analysiert werden und analytisch verbessert werden
- Schnell: industrielle Revolution

## **Nachteile:**

- Man braucht den (teuren) intelligenten Designer

## 2: *Evolution*

- Biologische Evolution; der blinde „Uhrmacher“
- Technische Evolution

### **Vorteile:**

- Dumm und einfach
- Selbst-optimierend

### **Nachteile:**

- Langwierig
- Teuer

## 3: *Lernen*

- **Biologisches Lernen:**
  - Vergleichsweise schnelle Anpassungen an die Umgebung
  - Vererben von Gelerntem durch Lehren
  - Rückkopplung des Erfolges
  - Schon primitive Tiere sind lernfähig
  - Grundeigenschaft von Tieren („Naturgesetz“)
- **Technisches Lernen**
  - Vorteile:
    - Schnell; realitätsnah
    - Rapider Austausch von Information (Lehren)
    - Lernen: Optimieren am Problem
  - Nachteile:
    - Man steht noch am Anfang
    - Es gibt noch viel zu tun!

# Charakterisierung des Lernens

- Lernfähigkeit ist eine **exklusive Eigenschaft von Lebewesen** (und Computern?) (Ferromagnete zeigen Memory-Effekte?). Selbst recht primitive Lebewesen zeigen Lernfähigkeit (Adaption)
- Versuch einer Definition:
  - (Nutzbringende?, Permanente?) **Veränderung im** (zentralen?) **Nervensystem** (basierend auf Interaktionen mit der Umwelt?) (Immunsystem?)

## *Charakterisierung des Lernens(2)*

- 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" und "laists" für "Spur". Die indogermanische Wurzel \*lais- bedeutet "Spur, Bahn, Furche" (vgl. <http://learn.idoneos.com/> und Duden/Etymologie).

(Wikipedia)

- Schon von der Herkunft her hat Lernen etwas mit Spuren hinterlassen zu tun. Lernen soll im Gedächtnis ebenso Spuren hinterlassen.

## *Nicht-technische Perspektiven:*

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



# 1: *Logik und Philosophie*

- Deduktion (Schliessen) und Induktion (Lernen)
- **Deduktion** arbeitet vom Allgemeinen zum Besonderen (top-down), von Axiomen zu Theoremen
- **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

# Philosophie: Empirismus

- **Francis Bacon** (1562-1626): Beginn des Empirismus; *Novum Organum*
  - Der Empirismus ist eine erkenntnistheoretische Richtung in der Philosophie und Psychologie, die alle Erkenntnis aus Sinneserfahrungen ableitet.
  - 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. (Wiki)
- **John Locke** (1632-1704): „Nothing is in the understanding, which was not first in the senses“
  - Locke gilt als ein Hauptvertreter des englischen Empirismus. Er bildet zusammen mit George Berkeley (1684–1753) und David Hume (1711–1776) das große Dreigestirn der englischen Aufklärung und des aufkommenden Empirismus.
  - Locke wendet sich gegen die Rechtfertigung der Naturwissenschaften aus dem bloßen Denken und sucht ihr Fundament stattdessen in der Erfahrung.
- **David Hume** (1711-1776): *A treatise on human nature*; Prinzip der Induktion: General rules are acquired by exposure to repeated associations between their elements
  - Während nach Locke die Wahrnehmungen Auskunft über die körperliche Welt geben, sind für Berkeley und Hume nur die Wahrnehmungen real. Eine gegenständliche Welt, die die Wahrnehmungen auslöst, ist nach ihnen nicht existent.

## *Philosophie: Scientific Discovery*

- Ist Induktion schlüssig? 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*

- **Definition: dauerhafte Änderung des Verhaltens, die durch Übung** (im Gegensatz etwa Reifung, Prägung oder Krankheit) **erfolgt**
- **Klassifizierung:**
  - **Habituation:** Lernen, einen Reiz zu ignorieren (vgl. Sensitivierung)
  - **Klassische Konditionierung:** Nach dem Lernvorgang erfolgt der Speichelfluss auf einen Ton auch ohne den Reiz: Futter.
  - **Operative Konditionierung:** Lernen, dass einer Aktion eine Konsequenz folgt. Ein Aktions-Konsequenz-Paar, dessen Konsequenz eine positive Motivationswirkung hat, kann man gezielt zum Antrainieren verwenden, entsprechend eines mit negativer Motivationswirkung zum Abgewöhnen einer Verhaltensweise (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

## Classical Conditioning (Pavlov)

Unconditioned Stimulus  
(*food*) → Unconditioned  
Response (*salivation*)

Unconditioned Stimulus  
(*food*) together with  
Conditioned Stimulus  
(*bell*) → Unconditioned  
Response

Conditioned Stimulus  
(*bell*) → Conditioned  
Response (*salivation*)

# *Psychologie als Wissenschaft*

- **Beginn der wissenschaftlichen Psychologie:**
  - Herrmann von Helmholtz (1821-1894) , Wilhelm Wundt (1832-1920)
  - Wissenschaftliche Methoden auf psychologische Fragestellungen angewandt (Vision)

# Verhaltenswissenschaft

- **Behaviorismus (1920-1960)** (Iwan Pawlow (1849-1936), John Watson(1878-1958)):
  - *„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**"),
  - Die Funktionsweise kann nur aus dem Input (Reize) und dem Output (Reaktion) erschlossen werden

# *Kognition und Kognitivisten*

- Versuchen, die inneren Vorgänge in der Black Box zu verstehen
- Menschliches Verhalten ist **mehr als eine Reiz-Reaktions-Kette** (Passivität), Reaktionen sind nur z. T. von außen steuerbar/manipulierbar
- 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), Helmholtz, Frederik Bartlett (1886-1969), George Miller (The magic number seven, 1956), Noam Chomsky (Three Models of Language, 1956)



## 3: Neurobiologie

- Neuronale Plastizität
- 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.
- Aplysia (Seeschnecke):
  - Kandel (Nobelpreis 2000)
  - Gill (Kiemen) withdrawal reflex: 24 Sensor-N, 6 Motor-N.
  - Habituation (Gewöhnung): (Verringerung der Neurotransmitter auf wiederholtem Reiz)
  - Sensitization (Sensibilisierung): (Erhöhung der Neurotransmitter auf schädlichem Reiz)
  - Association: Licht/Elektrischer Schock

# Neuron

- Ruhemembranpotential: -70 mV.
- Depolarisation:  $> -50\text{mV}$  -> Öffnung der Natrium-Kanäle, Auslösung eines Aktionspotentials
- Refraktärphase
- Systemtheorie: *Leaky integrator*

# *Synapse*

Postsynaptisch:

Öffnung von Ionenkanälen und damit  
Änderung des Membranpotentials der  
postsynaptischen Nervenzelle

The figure shows a synapse that is affecting another synapse. **Short term memory** (minutes) can be produced when a weak stimulus (thin arrows in the left lower part of the figure) is causing a protein phosphorylation of ion channels, which leads to a release of an **increased amount of transmitter**. For a **long term memory** (weeks) to be created, a stronger and more long-lasting stimulus is required (bold arrows in the figure). This causes an increased level of the **messenger molecule cAMP**, which causes further activation of protein kinases. They will phosphorylate different proteins and affect the **cell nucleus**, which in turn will issue orders regarding the synthesis of new proteins. This may lead to changes in the **form and function of the synapse**. The efficacy of the synapse can then be increased and more transmitter released.

# *Hebb'sches Lernen in Psychologie und Neurophysiologie*

- **Hebb'sches Lernen:** Wiederholte Aktivierung (Feuern) eines Neurons durch ein anderes Neuron über eine bestimmte Synapse erhöht deren Effizienz
- Hebb versuchte damit das **klassische Konditionieren** neurologisch zu erklären.
- Long Term Potentiation (LTP) / Depression (LTD)
- Hebb'sches Lernen nachgewiesen: Neuronen im Hippocampus von Säugetieren
- Offene Frage: wie viel biologisches Detail benötigt man, um die Essenz des Lernens nicht zu verlieren?

# *Technisches Lernen (Maschinelles Lernen)*

*Vision: ein System, welches  
sich selbst durch Interaktion mit  
der Umwelt verbessert*

# Technisches Lernen

## ***Vor dem Computerzeitalter:***

- Ableiten von Wissen aus Beobachtungen: Statistik

## ***Mit dem Computerzeitalter:***

1. Statistik, Mustererkennung (OCR), Spracherkennung
2. Neuroinformatik; Adaption, Kybernetik
3. Symbolisches Maschinelles Lernen in der Künstlichen Intelligenz
4. Statistisches Maschinelles Lernen
5. 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)
- $P(D|H=0)$ : Wahrscheinlichkeit der Beobachtung (Data)  $D$ , falls Hypothese  $H$  unwahr ist

- *Bayes'sche Regel:*

$$P(H=1|D) = P(D|H=1) P(H=1) / P(D)$$

*(posterior Wahrscheinlichkeit der Hypothese)*

- *Wobei:*

$$P(D) = P(D|H=1) P(H=1) + P(D|H=0) P(H=0)$$



# Subjektive Wahrscheinlichkeiten

- Die Wahrscheinlichkeit, dass Partei X die nächsten Wahlen gewinnt ist 45%
- Cox (1946), wenn man seinen Überzeugungen (*Belief*) Zahlen zuordnen will, kommt man unter wenigen Konsistenzannahmen auf den Bayes'schen 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 Bayes'schen 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."

## **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) , **Jerzy Neyman** (1894-1981)

## 2: Neuroinformatik (zunächst: Ausdruckskraft)

*McCulloch and Pitts (1943)*: erster Versuch, Gehirnmechanismen über einfache Recheneinheiten zu formalisieren (Netzwerke von einfachen logischen Einheiten)

- Keine Lernfähigkeit
- Fokus: Ausdruckskraft

## *Ausdruckskraft neuronaler Strukturen (2)*

- *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)  
Hebb versuchte damit das **klass. 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).*

## *Lernfähigkeit neuronaler Strukturen (2)*

- *Taylor (1950er), Steinbuch (1961): Associative memory (Hebb'sches Lernen)*

## *Lernfähigkeit neuronaler Strukt. (3)*

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

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

Maschinelles Lernen ist eng verknüpft mit der Geschichte der Künstlichen Intelligenz (KI)

Etwas vereinfacht:

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

Die vier Ansätze

- 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: Geschichte*

## **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-)
  - (Namesgeber: AI/KI)
  - Erfinder von LISP
  - Advice Taker: a first hypothetical program to be a complete AI system
- **Marvin Minsky** (1927-) (MIT)
  - SAINT (calculus integration)
  - ANALOGY (geometric analogy)
  - STUDENT (algebra)
  - Blocks World; The Society of Mind (1985)
  - Eher Gegner Logik-basierter Ansätze
- **Claude Shannon** (1916-2001) (Bell Labs, Erfinder der Informationstheorie)
- **Arthur Samuel** (1901-1990) IBM; a system to learn to play checker better than its creator
- **Ray Solomonoff** (1926-) (MIT) Begründer der Algorithmic probability

# *Dartmouth Workshop*

- **John von Neumann** Institute for Advanced Study; Begründer der Spieltheorie
- Nathaniel Rochester (IBM), Trenchard More (Princeton), Oliver Selfridge (MIT), Cliff Shaw
- **Allen Newell** (1927-1992)(CMU), **Herbert Simon** (1916-2001) (CMU):
  - Newell und Simon entwickelten in den 1960er Jahren den General Problem Solver (GPS), ein Programm, das mit einfachen Methoden beliebige Probleme lösen können sollte (nach fast zehnjähriger Entwicklung eingestellt)
  - **Physical Symbol System Hypothesis:** a physical symbol system has the necessary and sufficient means for general intelligent action
  - Informationsverarbeitung ist ein Rechenvorgang, also Symbolmanipulation
  - Auf das Gehirn als solches kommt es nicht an: Intelligence is mind implemented by any patternable kind of matter
  - Diese Auffassung, dass Intelligenz unabhängig von der Trägersubstanz ist, wird von den Vertretern der starken KI-These geteilt, wie beispielsweise Marvin Minsky (\*1927) vom Massachusetts Institute of Technology (MIT)

# *Kritik an der Künstliche Intelligenz*

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

### **Realitäts-Dosis (1966-1973)**

- 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).
- Der spätere Nobelpreisträger Simon prognostizierte 1957 unter anderem , dass innerhalb der nächsten 10 Jahre ein Computer Schachweltmeister werden und einen wichtigen mathematischen Satz entdecken und beweisen würde, Prognosen, die nicht zutrafen: 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.
- **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. Expertensysteme können ihre Problemlösung durch Angabe des benutzten Wissens erklären.
- Bruce Buchanan: **DENDRAL** (1969); inferring molecular structure from mass spectroscopy data; first knowledge intensive system
- Ed Feigenbaum (Stanford): **Heuristic Programming Project** (HPP) -> Expert Systems
- 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

# Symbolisches Maschinelles Lernen

In den frühen Jahren:

- Lernen in der KI eher stiefmütterlich (*only deductive inference is sound*)
- Regelbasierte Ansätze dominierten
- Kein konsistenter Umgang mit Unsicherheiten (Ursprünglich: Ablehnung von Wahrscheinlichkeitslehre und Statistik)
- Interessante Ansätze, aber wenig technisch-konkurrenzfähige Erfolge

# *Symbolisches Maschinelles Lernen (2)*

- Symbolisches ML: ML im engeren Sinne
- Erlernen von logischen Regeln: Inductive Logic Programming (ILP)
- Intuitiver Reiz: Man versucht (einfache) logische Regeln zu extrahieren (entspricht auch der Vorstellung eines Laien);
- Im Gegensatz: die im Moment dominierenden Lernansätze sind vornehmlich musterbasiert und finden oft schlecht interpretierbare Muster in hochdimensionalen Räumen
- Mächtigkeit: Man kann (first-order) Prolog Regeln erlernen (Turing-äquivalent)

# Symbolisches Maschinelles Lernen (3)

- Fallbasiertes Schließen (case-based reasoning, CBR) (1977, Schank)
- Ross Quinlan's ID3 (1979), Rivest
  - Lernen von *Identification Trees*
  - *Classification and Regression Trees* (CART, Breiman)

## *4: Statistisches Maschinelles Lernen: Revival der Neuroinformatik*

- Die Aktivitäten in der Neuroinformatik fanden erst Mitte der 80er Jahre wieder mehr Interesse
- 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 Neuroinformatik(2)*

**John Hopfield (1982, 1984):** Neural networks and physical systems with emergent collective computational abilities.

- Assoziativspeicher (Hebb'sches Lernen), kombinatorische Optimierung
- Zugang von Seiten der statistischen Physik (Spin-Gläser)
- Nichtlinear, parallel, fehlertolerant, rückgekoppelt
- Optische Rechner
- Bezug zur Funktionsweise des Gehirn
- Im Endeffekt: technisch nicht überlegen zu konkurrierenden Lösungen
- Grossberg, Kohonen

## *Revival der Neuroinformatik (3)*

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

- Diskriminantes Lernen
- Theoretisch sehr interessant aber nicht so praktikabel wie das MLP

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

- MLP: ein relativ leicht handhabbares mächtiges Werkzeug zur Modellierung hochdimensionaler nichtlinearer Abhängigkeiten
- Lösung des *exclusive-or-Problems*, Nettalk
- MLP: überlegenes Modellierungstool für hochdimensionale Probleme
- Interesse in der Statistik

Ab 1988 ein gewisser Hype („Lernen anstatt Programmieren“)

# *Statistisches Maschinelles Lernen*

- Ende der 1990er Jahre entstanden starke statistische Verfahren in der Neuroinformatik
  - Graphische Modelle; Bayes Netze
  - Support Vector Machine ...
- Die Neuroinformatik teilte sich auf in biologisches Lernen und Statistisches Maschinelles Lernen, welche gegenwärtig der dominierende Lernansatz ist
  - Bioinformatik
  - Vision
  - NLP, ...

## 5: *Datamining*

- Datamining: Extraktion oder „Mining“ von Wissen aus großen Datenmengen
- Statistik: Daten sind eher Mangelware; Small-Sample Statistics
  - Typisches Betätigungsfeld: Biostatistik: verbessert mein neues Medikament die Heilungsaussichten signifikant?
- Fokus im Datamining:
  - Große Datensätze
  - Große Anzahl von Variablen
  - Das (oft kommerzielle) Ziel (Geschäftsmodell) steht im Fordergrund: Direkt-Marketing

## *Datamining (2)*

- Datamining als Teil des KDD-Prozesses (Knowledge Discovery in Databases (KDD))
- Erste Arbeiten um 1990
  - 1989: IJCAI Workshop zu KDD
  - 1995 KDD Konferenz
  - 1998 SIG KDD Konferenz

## *Datamining (3)*

- Nützliche Methoden anderer Gebiete werden übernommen
  - Statistik
  - Maschinelles Lernen
  - Neuroinformatik
  
- Neue Methoden werden entwickelt
  - Association Rules (Agrawal)
  - DBSCAN (Ester, Kriegel, Sander, XU) ...

# *Die Gegenwart*

Stand der Technik: Technische Systeme, die intelligente Dinge vollbringen, ohne wirklich intelligent zu sein

# *Was kann Maschinelles Lernen?*

- **Zustandsabhängig eine Entscheidung treffen oder eine Vorhersage generieren (Input-Output Systeme)**
  - Agent
  - OCR-Klassifikation
  - Zeitreihenprognose
- **Zusammenhänge in Daten erkennen**
  - Datamining
  - Statistische Analysen
- **Lernen in Wissensrepräsentationen (Domänen- und Weltmodellen)**
  - Lernen WWW und im Semantik Web (Gedächtnis)
  - Lernen in komplexen Strukturen



# Aktuelle Aktivitäten

- Die großen Spieler im Informationszeitalter wie Microsoft, Google, Yahoo stellen im großen Maße ML-Personen 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, Speech, Information Retrieval, ...
- THESEUS: Lernen in Semantischen Web

# *Die Vorlesung*

- Technischen Werkzeuge, die heute im Fokus stehen mit entsprechenden Anwendungen
- Mathematische Voraussetzungen (werden eingeführt):
  - Lineare Algebra
  - Wahrscheinlichkeitslehre
  - Statistik

# Übersicht (vorläufig)

- Einführung
- Das Paradigma der Mustererkennung, das **Perceptron** (Neuron)
- So was Ähnliches habe ich schon mal gesehen: **Memory-basierte Ansätze**
- Über Korrelationen zur Vorhersage: **Lineare Regression**
- „Gehirnmodelle“ für nichtlineare Probleme: **Neuronale Netze**
- Die Guten ins Töpfchen, die Schlechten in Kröpfchen: **Lineare Klassifikatoren**
- Lineare Theorie für unendlich-dimensionale nichtlineare Systeme; **Kernelbasierte Systeme, Support Vector Machine**
- Ein unnötiger Disput: **Frequentisten, Bayesianer** und was immer
- Welches Modell ist das beste? **Modellselektion**
- Probabilistische Expertensysteme: **Bayes'sche Netze**
- Erlernen einfacher Regeln: **Entscheidungsbäume**
- Die Verbindung zur Prädikatenlogik und zu relationalen Datenbanken: **Relationales Lernen**
- Optimieren von lernenden Agenten: **Reinforcement Lernen**
- Versteckte Faktoren: **Principle Component Analysis**

**Planung**

| Datum                               | Vorlesung  | Datum                                  | Übung                         |
|-------------------------------------|--|--|-------------------------------|
| 21.04.2009                          | <a href="#">Einführung</a>   | 23/24.04.2009                          | <i>noch keine Übu</i>         |
| 28.04.2009                          | Das Perceptron / Memory-basiertes Lernen   | 30.04/01-05.2009                       | <i>noch keine Übu</i>         |
| 05.05.2009                          | Memory-basiertes Lernen 2 / Wiederholung: Lineare Algebra / Lineare Regression   | 07/08.05.2009                          | <a href="#">Übungsblatt 1</a> |
| 12.05.2009                          | Basisfunktionen und Neuronale Netze  | 14/15.05.2009                          | Übungsblatt 2                 |
| <del>19.05.2009</del><br>22.05.2009 | Basisfunktionen und Neuronale Netze 2 / Kernels: Teil I                          | <del>21/22.05.2009</del><br>19.05.2009 | Übungsblatt 3                 |
| 26.05.2009                          | Wiederholung zur Wahrscheinlichkeitstheorie / Frequentisten und Bayesianer       | 28/29.05.2009                          | Übungsblatt 4                 |
| 02.06.2009                          | <i>vorlesungsfrei</i>  | 04/05.06.2009                          | <i>keine Übung</i>            |
| <del>09.06.2009</del><br>12.06.2009 | Frequentisten und Bayesianer 2 / Lineare Klassifikatoren                         | <del>11/12.06.2009</del><br>09.06.09   | Übungsblatt 5                 |
| 16.06.2009                          | Textmining / Reinforcement Lernen  | 18/19.06.2009                          | Übungsblatt 6                 |
| <del>23.06.2009</del><br>26.06.2009 | Lineare Klassifikatoren 2 / Kernels: Teil II / Modellvergleich und Modellauswahl | <del>25/26.06.2009</del><br>23.06.2009 | Übungsblatt 7                 |
| 30.06.2009                          | PCA  | 02/03.07.2009                          | Übungsblatt 8                 |
| 07.07.2009                          | Bayes'sche Netze   | 09/10.07.2009                          | Übungsblatt 9                 |
| 14.07.2009                          | Relationales Lernen  | 16/17.07.2009                          | Übungsblatt 10                |
| 21.07.2009                          | 3 mögliche Klausurtermine: 21., 23. und 24.07.09                                 | 23.07.2009                             |                               |

# Generelles / Übungen

- Umfang: 2 + 2 Semesterwochenstunden
- Voraussetzung: Grundstudiumsvorlesungen der Informatik
- Dozent: Dr. Volker Tresp
- Übungsleiter: Franz Graf und Marisa Thoma
- Der Vorlesungsstoff kann in der Hauptdiplomprüfung bei Prof. Dr. Hans-Peter Kriegel geprüft werden. Absprachen mit anderen Prüfern müssen individuell getroffen werden.
- Vorlesungsbegleitend werden Übungsstunden angeboten:
  - Wöchentlich ist ein Übungsblatt abzugeben (allein oder in Zweiergruppen)
  - Drei davon werden bewertet: pro Kursteilnehmer müssen mindestens zwei Blätter mit „OK“ bewertet werden um zur Klausur (und damit dem möglichen Scheinerwerb) zugelassen zu werden.
- Skript (Folien): <http://www.dbs.informatik.uni-muenchen.de/Lehre/MaschLernen/>

| <b>Veranstaltung</b> | <b>Zeit</b>              | <b>Ort</b>                         | <b>Beginn</b> |
|----------------------|--------------------------|------------------------------------|---------------|
| Vorlesung            | Di, 16.00 -<br>18.00 Uhr | Raum 0.33<br>(Oettingenstr.<br>67) | 21.04.2009    |
| Übung                | Do, 12.00 -<br>14.00 Uhr | Raum 15<br>(Oettingenstr.<br>67)   | 07.05.2009    |
| Übung                | Fr, 14.00 -<br>16.00 Uhr | Raum 15<br>(Oettingenstr.<br>67)   | 08.05.2009    |

<http://www.dbs.informatik.uni-muenchen.de/Lehre/MaschLernen/>

# Literatur

## Vorlesung

- Machine Learning. Tom Mitchel: McGraw-Hill
- The Elements of Statistical Learning: Data mining, Inference and Prediction. Hastie, Tibshirani, Friedman: Springer (2nd Ed.)
- Pattern Classification. Duda, Hart, Storck: Wiley
- Neural Networks for Pattern Recognition. Bishop: Clarendon Press
- Pattern Recognition and Machine Learning. Bishop: Springer, 2006
- 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 und Wahrscheinlichkeitslehre

## 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 Proc.. Papoulis, McGraw, Hill