Machine Learning: 2017

Volker Tresp

Professor, LMU

Principal Research Scientist, Siemens AG

I. Introduction

II. Non-Technical Perspectives

- »Logic and Philosophy
- »Psychology
- »Neurobiology

III. Technical Learning

IV. Details on the Lecture

I. Introduction

Why is machine learning of interest?

Thesis: Learning is one of the three fundamental mechanisms for the design and the improvement of autonomous (intelligent) systems

1: Intelligent Design

» Almost all practical solutions are based on intelligent design

» Engineer: the knowledgeable “watchmaker“

» Programmer

Advantages:

» Explicit knowledge: the system is well understood and can be analyzed and can be improved via analytic thinking

» Time constant: years

Disadvantage:

Need for an (expensive) designer (human)

2: Evolution

- Improvement via trial and error
- Biological evolution
  - The “blind watchmaker“
- Technical evolution
  - Evolutionary improvement of technical solutions

Advantages:

- Simple (blind)
- Self-Optimizing

Disadvantages:

- Time constant: years, decades, centuries, …
- Wasteful

3: Learning

Biological Learning:

» Lifelong optimization (improvement) of the behavior of an individual via interaction with the environment

» The watchmaker that learns from a teacher to become a watchmaker and the improves h. skills over h. working life

» Basic properties of animals (“natural law”)

» Feedback of the learning success (*reinforcement*)

» Time constants: days

Human

» Change of future behavior in some sense, via processing of external information:

- » By acting in the world, from sensory inputs, but also by reading, by listening to a teacher, …
- » Reflected in any change of future behavior that was influenced by the outside world
- » Often associated with skill learning, memory, social skills, learning to read/write, reasoning, cognitive control, …

3: Learning (cont‘d)

Machine Learning

» Broadest sense: attempt to mimic biological learning for technical purposes

» Autonomous optimization of a technical system via interaction with reality or by analyzing acquired data

» “Learning instead of programming“

Characterization of Learning

- **Learning is an exclusive property of living beings** (and computers?)
  - Even quite primitive animals can learn (adaption)

- Biological Learning:
  - (beneficial? permanent?) Modifications in the (central?) nervous system (based on interactions with the environment?)

- Machine Learning:
  - Beneficial changes in a technical system based on the analysis of data or based on the interaction with the environment, by employing learning algorithms

Etymological Origin

- Etymologically:
  - Old English *leornian*
  - from Proto-Germanic *liznojan* (with a base sense of **"to follow or find the track)**
  - from Proto-Indo-European *leis* (track)
  - Related to German *Gleis* (track)

- Even etymologically, “learning” has something to do with the idea of “following traces / leaving traces”

II. Non-technical Perspectives

1. *Philosophy*
2. *Psychology*
3. *Neurobiology*

II. Non-technical Perspectives

1. *Philosophy*
2. *Psychology*
3. *Neurobiology*

Philosophy

- For most of human history, the study of learning and memory was a branch of philosophy

- Two aspects (often not clearly separated)
  - What can one know about the world?
    - Philosophy: Epistemology (“Erkennnistheorie”)
  - How does an individual (child) acquire knowledge
    - Genetics (inheritance)
    - Learning
    - Memory
    - How can we get to be so smart?
      - Logical Reasoning
      - Planning
      - Language
      - Science, …

Epistemology

Philosophy is not as much concerned in how we learn a skill, but how we can know and learn something about the world

» Epistemology is the branch of philosophy concerned with the nature and scope of knowledge and is also referred to as "theory of knowledge“(“Erkenntnistheorie”)

» Put concisely, it is the study of knowledge and justified belief

» It questions what knowledge is and how it can be acquired, and the extent to which knowledge pertinent to any given subject or entity can be acquired

Basic Concepts: Deduction and Induction

Deduction:

» From the general to the specific (*top-down*)

» Axioms are given and theorems are derived via the machinery of deductive reasoning

» Axioms:

  » Can be simple facts (“Jack’s height is 180 cm”)

  » Complex axioms (“If something is a dog, it is also a mammal”)

» Basis for the classical Artificial Intelligence

Induction:

» Generalizes observations (*bottom-up*), to generalize and to justify theories

» Inferring the validity of a hypothesis via observations and experiences

  » Simple facts (“Jack’s height is 180 cm”) as in deduction

  » **Learned dependencies instead of assumed axioms!**

» Basis for Machine Learning

Rationalism (top-down) (from Latin: ratio = „reason“)

- Priority of rational reasoning in knowledge acquisition
  - In contrast to other forms such as the senses or religious convention

- Representatives: **Socrates** (ca 470–399 BC), **Plato** (– 348/347 BC), René Descartes (1596–1650), Baruch Spinoza (1632–1677), Gottfried Leibniz (1646–1716), Immanuel Kant (1724–1804)
  - Since the Enlightenment, rationalism is usually associated with the introduction of **mathematical methods into philosophy**, as in Descartes, Leibniz, and Spinoza (Bourke 263). This is commonly called **continental rationalism**, because it was predominant in the continental schools of Europe, whereas in Britain empiricism dominated
  - Proponents of some varieties of rationalism argue that, starting with foundational basic principles, like the axioms of geometry, one could **deductively** derive the rest of all possible knowledge

- Psychology: individuals are shaped primarily by their inherited nature (nativists) (**nature** versus nurture)

Empiricism (bottom-up)

» More of a British tradition

» In contrast to Rationalism

» „There is nothing in the mind that was not first in the senses." John Locke postulated that, at birth, the mind was a blank slate or *tabula rasa*

» Representatives: **Aristotle** (384 – 322 BC), Francis Bacon (1562-1626), John Locke (1632-1704), George Berkeley (1685-1753), David Hume (1711-1776)

» Aristotle: Rules of association (near in space or time (contiguity) often (frequency); similarity)

» *Psychology:* we start with a blank slate *(tabula rasa) (Locke)* (nature versus **nurture**)

Idealism *(dominant Philosophy of the 19th century) and* ***Materialism***

» **Idealism:** each form of matter, including human behavior, is a reflection of ideas

» In philosophy, idealism is the group of philosophies which assert that reality, or reality as we can know it**, is fundamentally mental, mentally constructed, or otherwise immaterial.** Epistemologically, idealism manifests as **a skepticism about the possibility of knowing any mind-independent thing**. In a sociological sense, idealism emphasizes how human ideas — especially beliefs and values — shape society. As an ontological doctrine, idealism goes further, asserting that **all entities are composed of mind or spirit**. Idealism thus rejects physicalist and dualist theories that fail to ascribe priority to the mind

» Beginning with Immanuel Kant ("Das Ding an sich"), German idealists such as G. W. F. Hegel, Johann Gottlieb Fichte, Friedrich Wilhelm Joseph Schelling, and Arthur Schopenhauer dominated 19th-century philosophy

» **Materialism** (Ludwig Feuerbach, Karl Marx): history is not driven by ideas but by laws (historic-dialectic materialism)

Positivism (after 19th century)

- Knowledge is derived from positive findings
- Data derived from sensory experience, and logical and mathematical treatments of such data, are together the exclusive source of all authentic knowledge
  - Modern scientific thinking: Positivism is defined as the belief that all true knowledge is scientific, and that all things are ultimately measurable.

- This view holds that <u>society operates according to laws like the physical world</u>. Introspective and intuitional attempts to gain knowledge are rejected
- The concept was developed in the early 19th century by the philosopher and founding sociologist, **Auguste Comte**. Comte argued <u>that society operates according to its own laws, much as the physical world operates according to gravity and other laws of nature.</u> Also: Ernst Mach. **Émile Durkheim**

- Stephen Hawking is a recent high profile advocate of positivism, at least in the physical sciences. In The Universe in a Nutshell he writes:
  - Any sound scientific theory, whether of time or of any other concept, should in my opinion be based on the most workable philosophy of science: the positivist approach put forward by Karl Popper and others. According to this way of thinking<u>, a scientific theory is a mathematical model that describes and codifies the observations we make</u>. A good theory will describe a large range of phenomena on the basis of a few simple postulates and will make definite predictions that can be tested… If one takes the positivist position, as I do, one cannot say what time actually is. All one can do is describe what has been found to be a very good mathematical model for time and say what predictions it makes

Reactions to Positivism

» Anti-Positivism, Critical Theory

» **Max Weber** argued that sociology may be loosely described as a 'science' as it is able to identify causal relationships—especially among, or hypothetical simplifications of complex social phenomena. As a nonpositivist, however, one seeks relationships that are not as "ahistorical, invariant, or generalizable“ as those pursued by natural scientists.

» The antipositivist tradition continued in the establishment of critical theory, particularly the work associated with the Frankfurt School of social research. Antipositivism would be further facilitated by rejections of 'scientism'; or **science as ideology** (Frankfurter Schule: Herbert Marcuse, Theodor Adorno, Max Horkheimer, Walter Benjamin, Erich Fromm, Jürgen Habermas)

» Postpositivism

» While positivists believe that the researcher and the researched person are independent of each other, postpositivists accept that theories, background, knowledge and values of the researcher can influence what is observed.

» Logical Positivism/Neopositivism

» Logical positivists (or 'neopositivists') reject metaphysical speculation and attempted to reduce statements and propositions to pure logic

» Subgroup: Logical empiricism: A school of philosophy that combines empiricism, the idea that observational evidence is indispensable for knowledge of the world, with a version of rationalism the idea that our knowledge includes a component that is not derived from observation

Structure of Scientific Revolutions and Relativism

» Structure of Scientific Revolutions (SSR)

» The Structure of Scientific Revolutions (SSR) (Thomas Kuhn (U.S. Historian of Science, 1922 –1996). Kuhn argued that science does not progress via a linear accumulation of new knowledge, but undergoes periodic revolutions, also called **"paradigm shifts"** (although he did not coin the phrase), in which the nature of scientific inquiry within a particular field is abruptly transformed

» Kuhn did not consider himself a relativist

» Relativism

» A form of truth relativism, which is the doctrine that there are no absolute truths, i.e., that truth is **always relative to some particular frame** of reference, such as a language or a culture

» Paul Feyerabend (1924 – 1994): scientific knowledge is not cumulative or progressive and that there can be no demarcation in terms of method between science and any other form of investigation

» Relativism has strongly critics!

Karl Popper: Scientific Discovery

Karl Popper

» Is induction sound?

» Karl Popper (Sir, 1902-1994). *The Logic of Scientific Discovery*

» **If no finite set of observation can ever prove a theory, how can we ever accept a scientific theory as being true?**

» Popper accepts Empiricism as a valid means to increase knowledge, if one accepts that theories can only be tested but never be proven

» Popper is considered a postpositivist

Philosophical Connections to machine Learning and Artificial Intelligence

- **Logical positivism** was a motivation for early AI research (Dominance of Logic-based approaches)

- Machine Learning has its roots in **Empiricism** and **Logical Empiricism**

II. Non-technical Perspectives

1. *Philosophy*
2. *Psychology*
3. *Neurobiology*

Psychology

» Psychology is the study of mind and behavior

» Humans are in focus

» Special focus: human learning

Psychoanalysis

» Psychoanalysis was founded by Sigmund Freud (1856-1939)

» Hypothesis: people can be cured by making conscious their unconscious thoughts and motivations, thus gaining “insight”

» Psychoanalysis is regarded by some critics as a pseudoscience

» Maintains a strong influence on psychiatry (a branch of **medicine**: diagnosis, prevention, study and treatment of mental disorders)

Psychology as Empirical Science

Begin of empirical (experimental) psychology:

- » Herrmann von Helmholtz (1821-1894)
- » Wilhelm Wundt (1832-1920) (Assistant to Helmholtz)
  - » Wundt is considered to be the founder of psychology as a separate scientific field
  - » From 1858 to 1863, he was assistant to Hermann von Helmholtz. “Theorie der Sinneswahrnehmungen”
- » Gustav Theodor Fechner (1801–1887): Founder of Psychophysics
  - » “The Scientific Study of the Relation between Stimulus and Sensation“
- » Hermann Ebbinghaus (1850-1909): first rigorous experimental studies on human memory

Behaviorisms (1920-1960)

- *„Belief in the existence of consciousness goes back to the ancient days of superstition and magic"*
- Also as reaction to Sigmund Freud
- Rejection of theories that need to assume mental states
  - The inner structure (of the brain) is irrelevant
- The functioning can only be deduced from input (stimulus) and output (reaction)
- "Input" can include personal history
- Humans are just another animal (Freud exclusively focused on humans)
- Humans start *tabula rasa (nature versus **nurture**)*
  - At a time when racism was popular in large parts of the world
- Representatives: Iwan Pawlow (1849-1936), John Watson (1878-1958), B. F. Skinner (1904-1990)

Classical Conditioning: Learning to predict important events *(Pawlow)*

How do I know that an association has been learned? Both produce the same response!

Unconditioned stimulus (food)		Unconditioned response (salivation)	*A stimulus-response connection that required no learning: salvation(food)*	
Unconditional stimulus (food) With conditional stimulus (bell)		Unconditional response (salivation)	*A stimulus which produces no response (i.e. neutral) is associated with the unconditioned stimulus; learning the association of stimuli bell~food*	
Conditional stimulus (bell)		Conditioned (learned) response (salivation)	*A stimulus-response connection that was learned : salvation(bell)*	

More examples:

- A perfume (UCS) could create a response of happiness or desire (UCR)
- A person (CS) who has been associated with nice perfume (UCS) is now found attractive (CR)

- A stomach virus (UCS) would produce a response of nausea (UCR)
- Chocolate (CS) which was eaten before a person was sick with a virus (UCS) is now produces a response of nausea (CR).

Cognitive Psychology and Cognition

- » Attempt to understand the inner working of the „Black Box“
- » Reaction to Behaviorism
- » Human behavior is more than stimulus-response
- » Development is an active process of a subject
- » Reintroduction of *mental processes*
- » In contrast to Freud: Computer metaphor
- » Acting is dominated not only by a stimulus but by active reasoning
- » The link between stimulus and behavior is the cognitive representation
- » Williams James (1842-1910), Herrmann von Helmholtz (1821-1894), Frederik Bartkett (1886-1969), George Miller (*The magic number seven*, 1956; information theory and memory), Noam Chomsky (*Three Models of Language*, 1956)

Cognitive Psychology

All of these faculties evolve under constant interactions with the environment, thus is a sense (and according to our definition) involve learning:

» Sensation and Perception

» Object Recognition and Representation

» Attention

» Learning and Memory

» Language

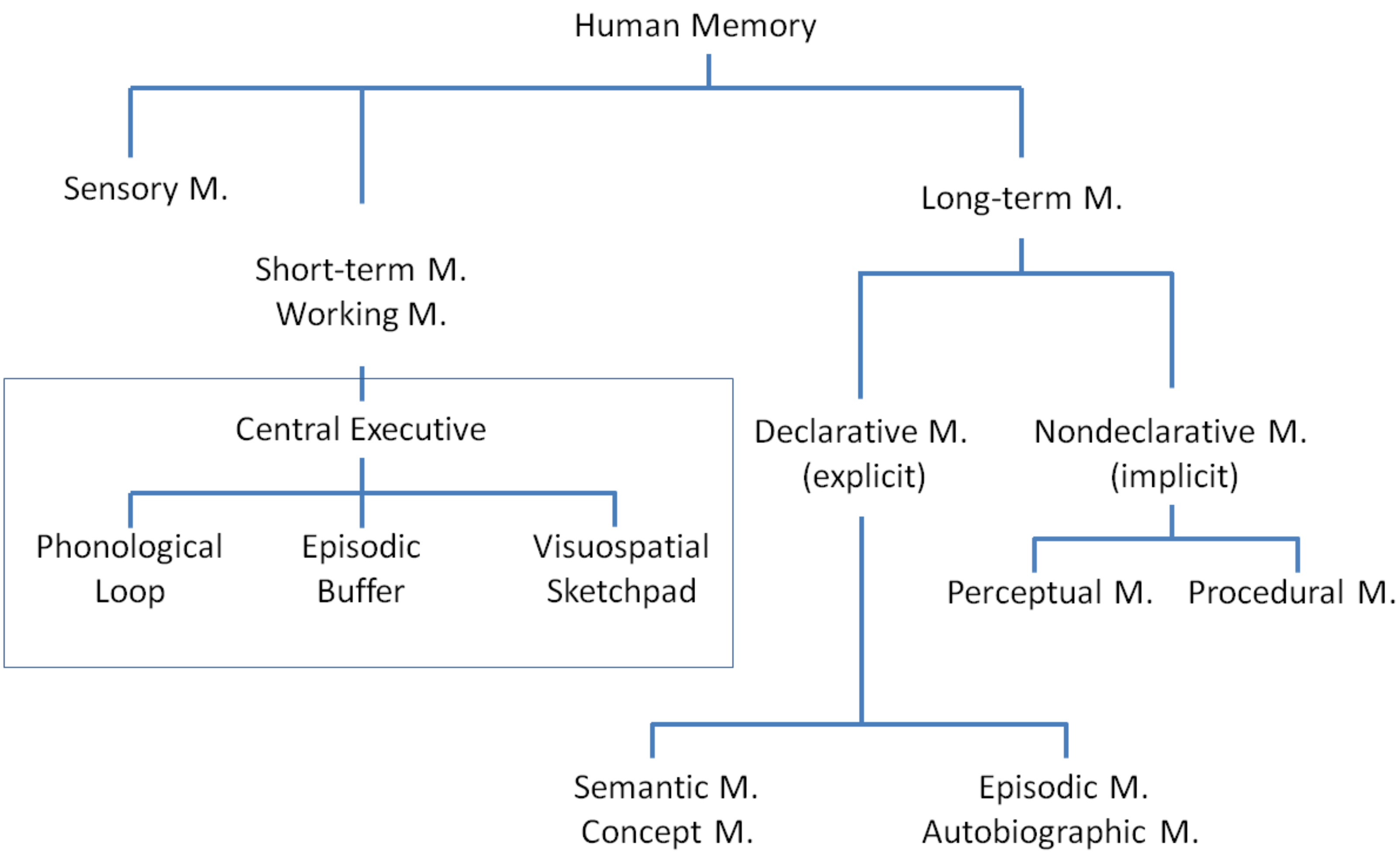
» Emotion

» Action

» Cognitive Control, Reasoning, Planning

» Consciousness

Types of Memory



Learning

- Habituation, Sensitization, Familarization
  - Learning about repeated events
- Classical Conditioning
  - By learning the association (Bell~Food) , Bell produces the same reaction as Food (learning to predict)
- Operand Conditioning
  - Learning the outcome of behavior (Learning to Act)
  - Stimulus => **Response** => Outcome
  - Reinforcement Learning
- Generalization and Discrimination Learning
- Social Learning
  - Observing, interacting and reinacting
  - Learning to copy behavior

Psychology and Machine Learning

» The statistical approach of psychology from behaviourism influenced Machine Learning

» Psychology as a guideline for technical solutions. Example: object recognition in the brain takes much less than a second and thus cannot involve much reasoning

» Machine Learning motivates much research in Cognition and vice versa

» Coinventors of the MLP (David Rumelhart and Geoffrey Hinton) are cognitive psychologists

II. Non-technical Perspectives

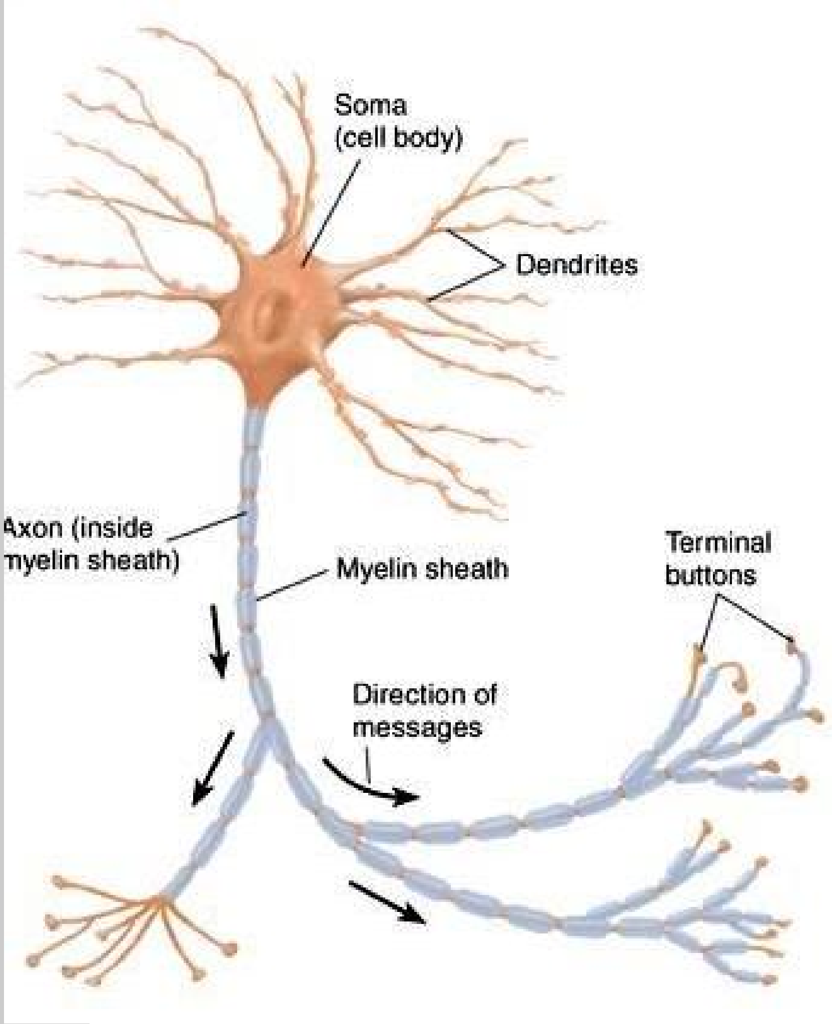
1. *Philosophy*
2. *Psychology*
3. *Neurobiology*

Neurobiology

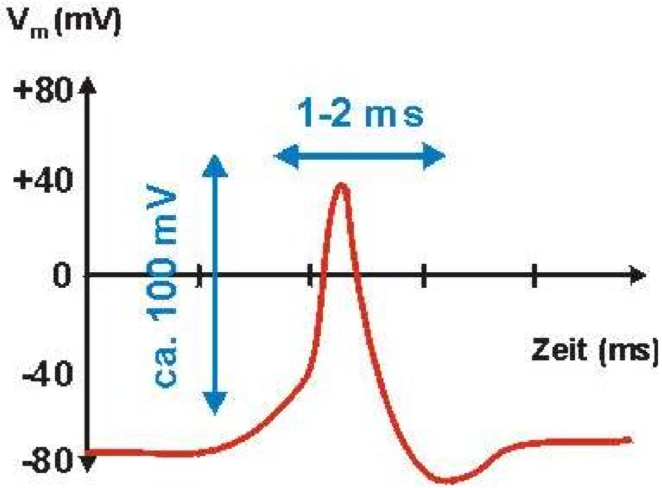
- Neurobiology or Neuroscience is the scientific study of the nervous system

- Learning: „Neuronal Plasticity“
- There must be a physical change if something is learned or memorized
- *Central mechanism: Synapses change their efficiency (synaptic plasticity)*
  - *Short-term plasticity:* the change lasts milliseconds to minutes
  - *Long-term plasticity:* the synaptic efficiency changes from hours to life-long

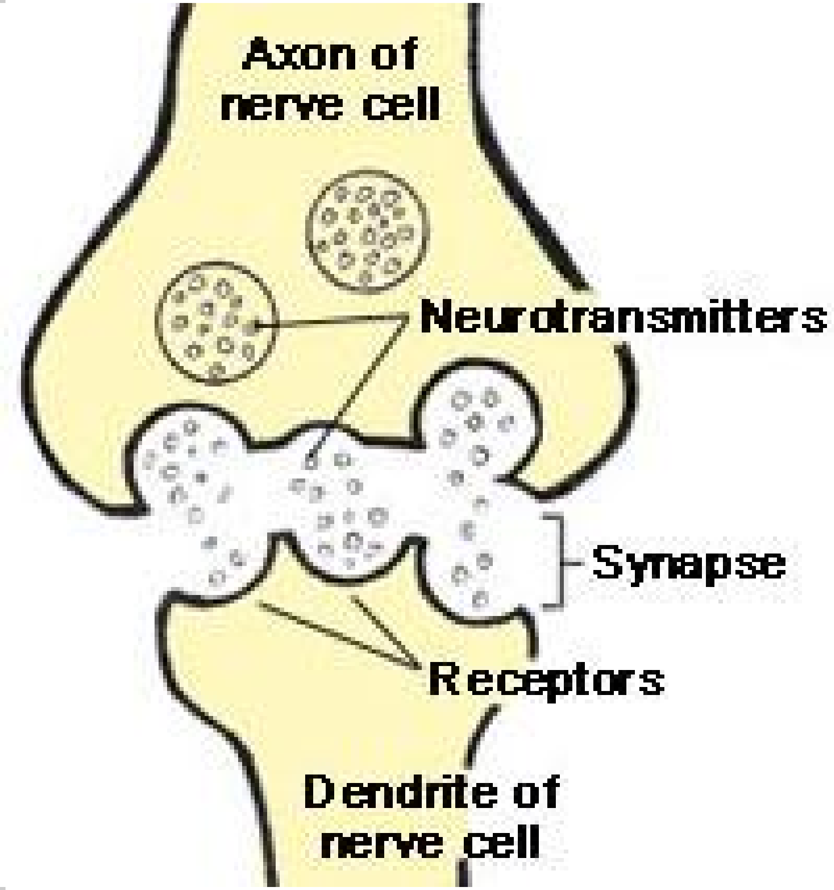
Neuron



- Resting potential: -70 mV.
- Depolarization: > -50mV
  - -> Opening of the sodium channels; action potential
- Refractory period: during this time no new action potential can be generated, independent of activation strength (app.. 2ms)
- Systems theory: *leaky integrator*



Synapse

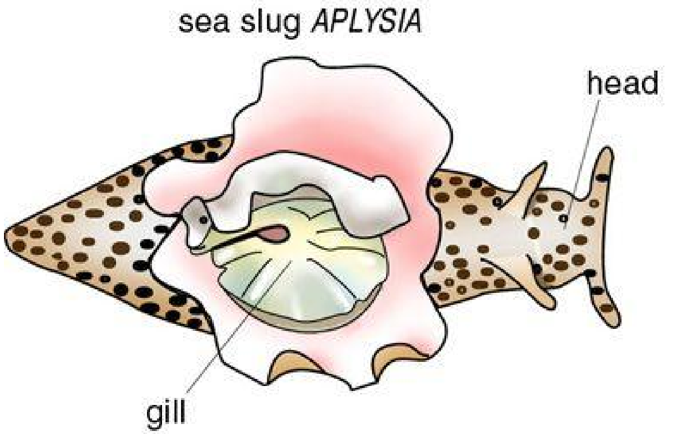


Presynaptic: Presynaptic discharge of the action potential leads to the release of neurotransmitters

Postsynaptic: opening of ion channels and thus change of the postsynaptic membrane potential

Example: Aplysia

- Eric Richard Kandel (* 7. November 1929 in Vienna): US-American Neuroscientist with Austrian origin
- Study object: Californian see slug (*Aplysia californica*)
- Nobel price 2000
- Gill-withdrawal reflex with 24 sensory-neurons and 6 motor-neurons
- Habituation
  - Reduction of neurotransmitters with repeated stimuli
- Sensitization:
  - Increase of neurotransmitters with repeated (damaging) stimuli
- Association:
  - Light/electric chock



The Synapse (left) influences a second synapse (bottom)

Short-term memory (minutes):

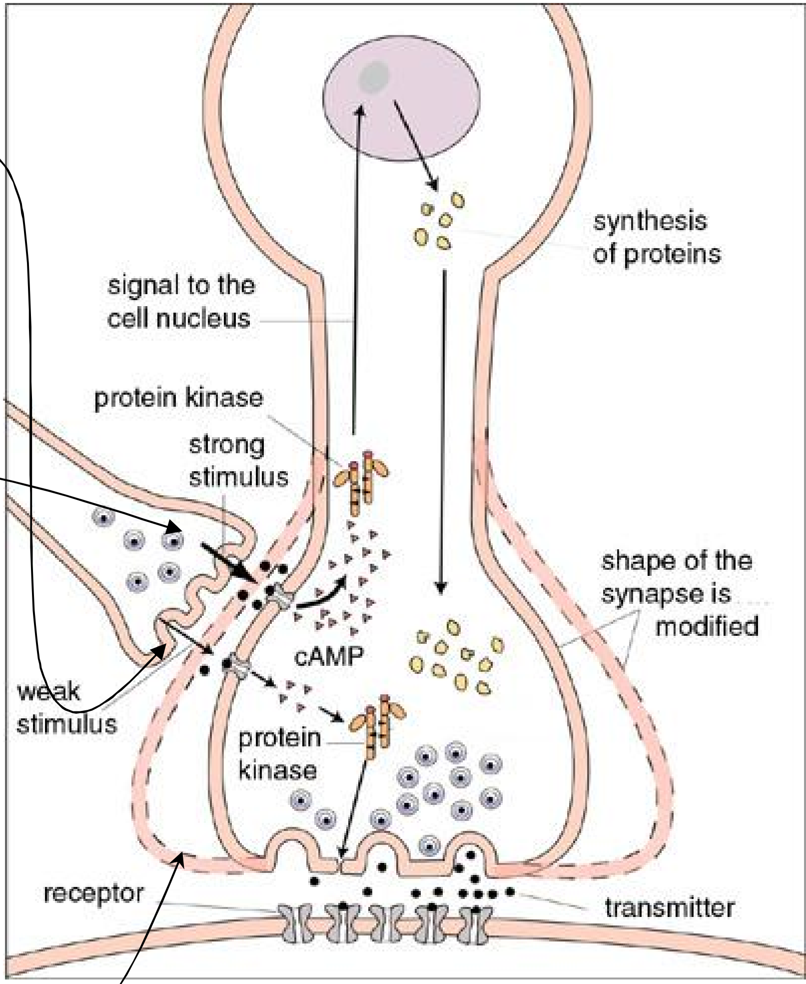
» A weak stimulus results in the phosphorylation of proteins of the ion channels (weak arrow on the left) which results in an increase of neurotransmitters

Long-time memory (weeks):

» A strong stimulus (thick arrow, left) results in a increased level of the messenger cAMP (Cyclic adenosine monophosphate), which results in an amplification of the protein kinase

» This influences the cell's DNA and new proteins are generated

» This results in an increases efficiency of the synapse: its efficiency is increased and mode neurotransmitters are generated



Machine Learning 2017

Hebb Learning in Psychology und Neurophysiology

- Kandel's results supplied new evidence for the Hebb's law
  - "When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."
  - "Neurons that fire together wire together" (long-term potentiation (LTP))
  - "Neurons out off sync, loose their link" (long-term depression (LTD))
- Hebb learning has been conformed biologically, i.e., in the neurons of the hippocampus
- Hebb formulates learning much more abstractly than Kandel
  - Open question for Machine Learning: how much can one ignore biological details without loosing the essence (e.g. spiking, spike timing?)

Neurobiology and Machine Learning

- Neuroscience is a focus in major machine Learning conferences (like the NIPS conference)
- Machine Learning tries to maintain some of the inherent properties of biological learning:
  - Distributed computing
  - Local computing
  - Noise tolerance
  - Fault tolerance
  - Graceful degradation
- Neurobiological relevance of Machine Learning architectures and algorithms is sometimes hotly debated (is backprop biological plausible?)

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neural Computation*
3. *AI and Machine Learning*
4. *Neural Computation: Revival*
5. *Modern Machine Learning*
6. *Data mining; Big Data*

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neural Computation (I)*
3. *AI and Machine Learning*
4. *Neural Computation (II)*
5. *Modern Machine Learning*
6. *Data mining; Big Data*
7. *Neural Computation (III): Deep Learn.*

Thomas Bayes (Rev., 1701 -1761)

» Updating the degree of belief in hypothesis based on observations

» *P(H=1):* degree of belief in the truthfulness of a hypothesis *H* (a priori assumption); *top-down*

» *P(D|H=1):* Plausibility of the data *D*, if the hypothesis *H* is true (Likelihood): *bottom-up*

» *P(D|H=0):* Plausibility of the data *D*, if the hypothesis *H* is false

» Bayes' theorem:

» $P(H=1|D) = P(D|H=1)\, P(H=1) / P(D)$

(a posteriori probability of the hypothesis); the updated *top-down*

» Pierre-Simon Laplace (1749–1827) then further developed the theory

Applications of Bayes' Theorem

Probabilistic Reasoning with known probabilities for prior and likelihood

» *P(PatientHasBronchitis)*

x P(PositiveX-Ray | PatientHasBronchitis)

Probabilistic Reasoning with unknown probabilities for prior and likelihood (Hierarchical Bayesian Reasoning; *Bayesian Statistics; Bayesian Machine Learning*) involves unknown parameters. For example, in a supervised setting,

- *P(PatientHasBronchitis | PositiveX-Ray, Parameter)*

x P(Parameter)

Bayesian Statistics is based on the concept of Subjective Probability

» Subjective probability:

  » Before I throw a coin, what is the probability that it is a fair coin

  » I believe that the probability that party X wins the election is 45%

» Cox (1946): Cox's theorem implies that any plausibility model that meets the postulates is equivalent to the subjective probability model, i.e., can be converted to the probability model by rescaling

  » If a *one* corresponds to the belief that an event happens with certainty and if a *zero* corresponds to the belief that an event does not happen, and numbers in between corresponds to degrees of certainty, then these numbers exactly behave as probabilities

Critique on Bayesian Statistics

Karl Pearson (1857 – 1936) now considered the <u>founder of modern statistics (nowadays called classical or frequentist statistics)</u>

» *"I felt like a buccaneer of Drake's days -... I interpreted that sentence of **Francis Galton (1822-1911) [his advisor]** 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)

» Criticism on the role of subjective probabilities: frequentists only make statements about repeatable experiments

» One evaluates if the data contradict a hypothesis but one does not make statements about the probability of a hypothesis

Egon Pearson (1895-1980) Son of Karl P., **Jerzy Neyman** (1894-1981)

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neural Computation (I)*
3. *AI and Machine Learning*
4. *Neural Computation (II)*
5. *Modern Machine Learning*
6. *Data mining; Big Data*
7. *Neural Computation (III): Deep Learn.*

Remark

The two poles we found in Philosophy, Psychology and Biology are also present in the history of intelligent systems

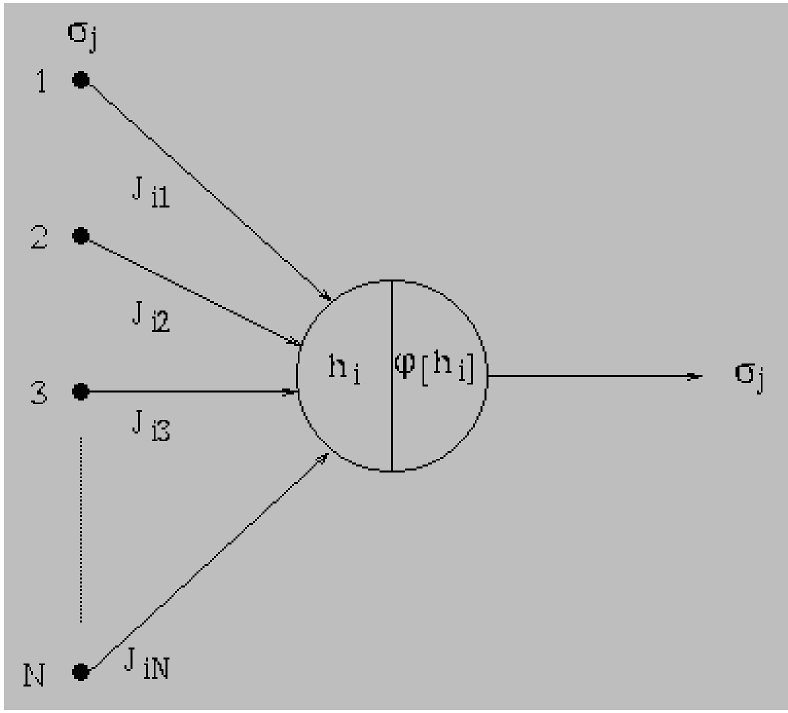
Dominance of

- Internal mechanisms; top-down view
  - Classical Artificial Intelligence (AI)

- External influences dominate; bottom-up view
  - Statistics
  - Neural Networks
  - Machine Learning

Neural Computation

- First: focus on expressiveness of Neural Networks (and not learnability)
- *McCulloch and Pitts (1943):* first attempt to formalize brain functions via simple computational nodes (network of simple logical units)



Expressiveness of Neural Structures

- *John v. Neumann (1956):* investigated the error tolerance of Neural Networks ("*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*

Learning in Neuronal Structures

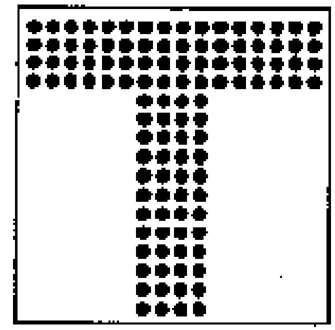
» *Hebb (1949):* Repeated activation of one neuron by another, across a particular synapse, increases its conductance (Hebb‘s theorem); "*Neurons that fire together wire together*"

» Hebb tried to explain classical conditioning via neural mechanisms

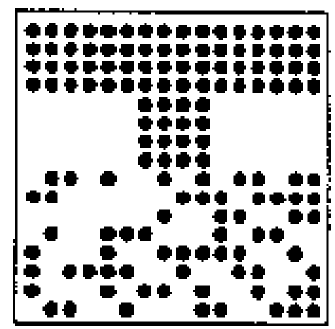
- *Wiener (1949):* Cybernetics, or control and communications in the animal and the machine *The whole world -- even the universe -- could be seen as* ***one big feedback system subject to the relentless advance of entropy,*** *which subverts the exchange of messages that is essential to continued existence (Wiener, 1954).* Book: *Cybernetics or Control and Communication in the Animal and the Machine* (1948)

Learning in Neuronal Structures: Associative Memory

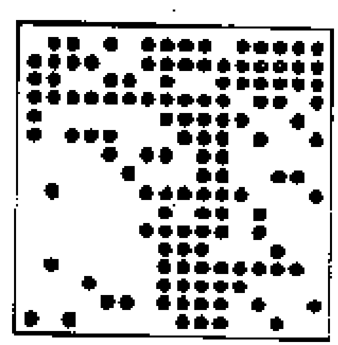
- W. K. Taylor (1950er), Karl Steinbuch (1961)
- Associative memory, “Lernmatrix”
- Relationship to Hebb Learning



Original ‘T’



half of image corrupted by noise



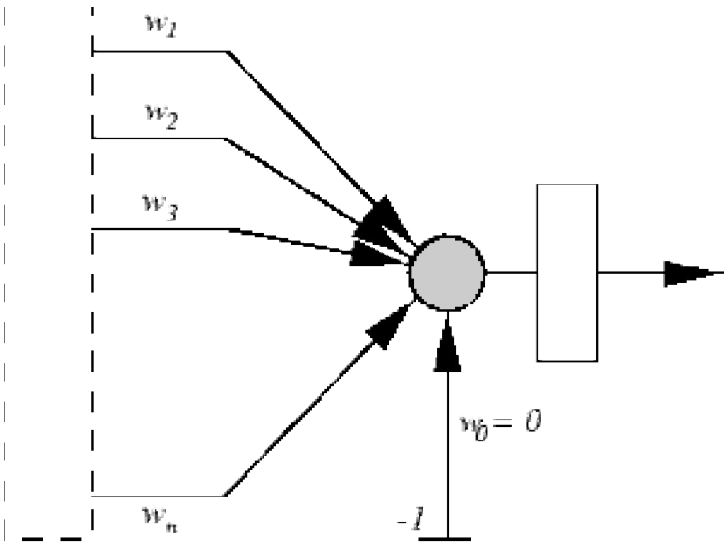
20% corrupted by noise (whole image)

Perceptron and ADALINE

» Minsky developed 1954 in his dissertation a neural computer he called SNARC (Stochastic Neural Analog Reinforcement Calculator)

» Rosenblatt developed 1958 the Perceptron learning rule and formulated a convergence proof; Mark I Perceptron

» Widrow and Hoff developed 1960 the ADALINE (ADaptive LINear Element) (used in modems)



» Minsky and Papert published 1969 the book „Perceptrons“ and demonstrated the limitations of the Perceptrons and of the ADALINE (Exclusive-Or Problem)

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neural Computation (I)*
3. *AI and Machine Learning*
4. *Neural Computation (II)*
5. *Modern Machine Learning*
6. *Data mining; Big Data*
7. *Neural Computation (III): Deep Learn.*

The Classical Area of Artificial Intelligence (AI): Neural Computation Winter

- After the book of Minsky and Papert funding almost exclusively went into the emerging field of AI
- No more funding for the study of learning systems
- A brief history of classical AI (1960s to 1980s)

AI and Induction

- Machine Learning is „Intelligence via Learning“
  - Induction
  - Empiricism
- (Classical) AI: Axioms permit the derivation of theorems form axioms via deduction (motivated form logical positivism)
  - Complex axiom: Dogs are mammals
  - Fact (simple axiom): Buster is a dog
  - Theorem: -> Buster is a mammal

Four AI Goals

1. Understanding human thinking: Cognition

2. Indistinguishably from human acting (Turing Test)
   - Language, knowledge representation, reasoning, learning, vision, robotics

3. Rational (optimal) (not necessarily human) Reasoning: Logic

4. Rational (optimal) (not necessarily human) Acting: Agents

- **Normative or prescriptive decision** theory is concerned with identifying the best decision to make, modeling an ideal decision maker who is able to compute with perfect accuracy and is fully rational. The practical application of this prescriptive approach (how people ought to make decisions) is called decision analysis, and is aimed at finding tools, methodologies and software (decision support systems) to help people make better decisions.
- In contrast, **positive or descriptive decision theory** is concerned with describing observed behaviors under the assumption that the decision-making agents are behaving under some consistent rules.

AI: Roots

Philosophical Basis (logical positivism):

» Vienna School (Rudolf Carnap (1891-1970))

» ***Logical* Positivism:** All knowledge can be characterized by logical theories …

» Ludwig Wittgenstein (1889-1951), Bertrand Russel (1872-1970)

Birth of AI: Dartmouth Workshop (1956)

John McCarthy (Dartmouth, later Stanford) (1927-2011)

» Naming: AI (to distinguish it from Cybernetics); Inventor of LISP

Marvin Minsky (1927-) (MIT)

» SAINT (calculus integration); ANALOGY (geometric analogy); STUDENT (algebra); Blocks World; The Society of Mind (1985); Critique on the dominating roles of Logic in AI and Statistics in Machine Learning

Claude Shannon (1916-2001) (Bell Labs) Inventor of Information Theory

Arthur Samuel (1901-1990) IBM; checkers program

Ray Solomonoff (1926-2009) (MIT) Founder of Algorithmic Probability

John von Neumann Institute for Advanced Study; Founder of Game Theory

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

» General Problem Solver (GPS): a program to solve general problems (terminated after 10 years)

» Representative of strong AI: Intelligence is independent of substrate

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

Further Development

Early Enthusiasm (1952-1969)

» In the first AI phase there was an unlimited expectation with respect to the capabilities of computers to „solve tasks for which intelligence is required, if they would be executed by humans “ (Minsky).

» Herbert Simon (1957)

  » Within the next 10 years a computer will be world champion in chess and will derive an important mathematical theorem

  » 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…

» In 1958 McCarthy proposed to formalize the complete human knowledge in form of a homogeneous formal representation, first order predicate logic

First Reality-Dose (1966-1973)

» Translation of Russian into English was stopped: "the spirit is willing but the flesh is weak“ became "the vodka is good but the meat is rotten"

» Reasoning did not scale up

Knowledge-based Systems

Knowledge-based Systems(1969-1979)

- » Expert systems: In an expert system, there is a formal knowledge representation, for example as a set of rules, and these are applied to known facts to infer new facts
- » Bruce Buchanan: *DENDRAL* (1969); inferring molecular structure from mass spectroscopy data; first knowledge intensive system
- » Ed Feigenbaum (Stanford): *Heuristic Programming Project* (HPP)
- » Feigenbaum, Buchanan, Shortliffe; *MYCIN:* Diagnose blood infections; extensive interviewing of experts; uncertainty factors
- » Progress in NLP: Eugene Charniak, Roger Shank
- » PROLOG

AI becomes an Industry (1980- and a few years later)

- » McDermott: **R1** (DEC, 1982); Configuration of new computer systems; each major company has an AI group
- » Japan (1981) Fifth Generation Project; 10-year project for the realization of intelligent computers based on PROLOG

Collapse (1984) of many Silicon Valley start-ups (Beginning of the AI winter)

Machine Learning in Classical AI

- Machine Learning was not in focus in classical AI (*only deductive inference is sound*)
- The field wanted to distinguish itself from statistics and probability
- Focus on symbolic Machine Learning
- Out off this tradition
  - Case-based reasoning (*case-based reasoning*, CBR) (Schank, 1977)
  - Learning of decision trees (Ross Quinlan‘s ID3, 1979, Rivest)
  - *Inductive Logic Programming* (Stephen Muggleton, 1991)
    - Intuitively attractive: The goal is to extract simple logical rules
    - Powerful: One can learn (*first-order*) Prolog Rules (Turing-equivalent)

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neural Computation (I)*
3. *AI and Machine Learning*
4. *Neural Computation (II)*
5. *Modern Machine Learning*
6. *Data mining; Big Data*
7. *Neural Computation (III): Deep Learn.*

Revival of Neural Computation

- There was increasing interest in neural computation around the mid 80s; end of the neural winter (1969-1982)
- End of the AI hype: maybe the substrate is relevant after all
- Learning in focus; opposition to rule-based approaches
- Fascination brain: despite the biological complexity there should be a simple organizational principal, which leads to intelligence via learning. Maybe intelligence can only be reached via learning?
- Technically high-performing solutions

Revival of the Associative Memory

- **John Hopfield (1982, 1984): ** *Neural networks and physical systems with emergent collective computational abilities*
- Achievements:
  - Associative memory (Hebb learning)
  - Combinatorial optimization
- Contributions from statistical physics (Spin-glasses)
- Interesting features: nonlinear, parallel, error tolerant, feedback
- Implementation as optical computer?
- Relationship to brain functioning
- At the end: solutions were not technically competitive
- Independent and earlier: Stephen Grossberg, Teuvo Kohonen

Boltzmann Machine, Multi-layer Perceptron

Ackley, Hinton, Sejnowsky (1985): *Boltzmann Machine*
» Discriminative Learning; close connection to Statistical Physics
» Theoretically very interesting but not as practical as the MLP

Rumelhart, Hinton, Williams (1986): Multi-layer Perceptron (MLP)
» MLP: a robust powerful tool for modeling high-dimensional nonlinear dependencies
» Solution to the *exclusive-or-problems,* Nettalk
» MLP: superior modeling tool for high-dimensional problems
» *Neural Computation breakthrough*
» Interest in Statistics

» Since 1988 a certain hype („Learning instead of programming“)

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neural Computation (I)*
3. *AI and Machine Learning*
4. *Neural Computation (II)*
5. *Modern Machine Learning*
6. *Data mining; Big Data*
7. *Neural Computation (III): Deep Learn.*

Modern Machine Learning

- » In a similar way as logic dominates classical AI, there is a great influence of statistical thinking in Machine Learning from the early 1990s onward
- » In the second half of the 1990s modern machine learning increasingly became Statistical Machine Learning (SML)
- » AI became highly influenced by Machine Learning

Main Phases

- Late 1980s until around 2000
  - MLP, Neural Computation, strong influence from physics (mean field theory), model diversity (Hopfield network, Kohonen networks, Boltzmann machine, …)
- Since second half of 1990s
  - Statistical Learning Theory, Support Vector Machines, Max Margin approaches
  - Bayesian networks and Bayesian learning, Gaussian Processes
  - Committee machines, ensemble methods
- Since early '00s
  - Nonparametric Statistics (Infinite Models)
- Since second half of '00s
  - Factorization approaches; Topic Models; Web and search
- After 2010
  - Deep Learning, MLP revival (sparked by Geoffrey Hinton, Yann LeCun, Joshua Bengio); Deep Learning has significantly improved a number of benchmarks in vision and speech recognition!

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neural Computation (I)*
3. *AI and Machine Learning*
4. *Neural Computation (II)*
5. *Modern Machine Learning*
6. *Data mining; Big Data*
7. *Neural Computation (III): Deep Learn.*

Fields by Applications Foci

- **Statistics**
  - Focus on the significance of a dependency (does the medication work or not) and interpretability
  - Classically few data points, few variables
- **Machine Learning**
  - Vision of a intelligent learning machine; great challenge brings great dynamics; focus on predictive models
- **Data Mining**
  - Analysis of large data bases; discovery of patterns in data; focus on data understanding
  - Today: also WWW focus
  - Large number of variables, large data sets
- **Big Data**
  - Google, Facebook and others make huge profits in analyzing huge data sets using Map Reduce, Hadoop, Storm, Spark …
    - Is this also relevant to other industries?
  - Very Large Databases meets Machine Learning, Data Mining, Statistics (Log-linear models, Random Forests, Deep Learning)

Data Mining

- Data Mining as part of the KDD - Process (**Knowledge Discovery in Databases** (KDD))
- History:
  - 1989 IJCAI Workshop on KDD
  - 1995 KDD Conference
  - 1998 SIG KDD Conference
- Based on available statistical and machine learning approaches. But also development of novel approaches
  - Frequent Item Sets, pattern discovery, Association Rules
  - DBSCAN (Ester, Kriegel, Sander, XU) ...

Impact

» Machine Learning has increasing impact on other areas such as Vision, Speech Recognition, Information Retrieval, Information Extraction, Bioinformatics, …

» The big players in information age, such as Microsoft, Google, Yahoo, Amazon, Facebook, are hiring huge quantities of machine learners

» Big Data Hype: most large companies have activities and are looking for “data scientists” …. Now: AI experts

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neural Computation (I)*
3. *AI and Machine Learning*
4. *Neural Computation (II)*
5. *Modern Machine Learning*
6. *Data mining; Big Data*
7. *Neural Computation (III): Deep Learn.*

Neural Computation III

- » Larger neural network models (more layers), more data, faster computers, new algorithms and tricks
- » Currently attracts best talents
- » Performance oriented: Heuristics and engineering but missing principals?
- » Deep belief network (Hinton et al., 2006)
- » In 2012, ImageNet breakthrough (G. Hinton, A. Krizhevsky, I. Sutskever)
- » Around 2012: In consumer products (improving speech recognition in Smart Phones)
- » Best performance in a number of benchmarks (vision, speech recognition, computational linguistics)
- » An order of magnitude improvement on ImageNet benchmark data
- » Deep Learning has increasing impact in application and industry

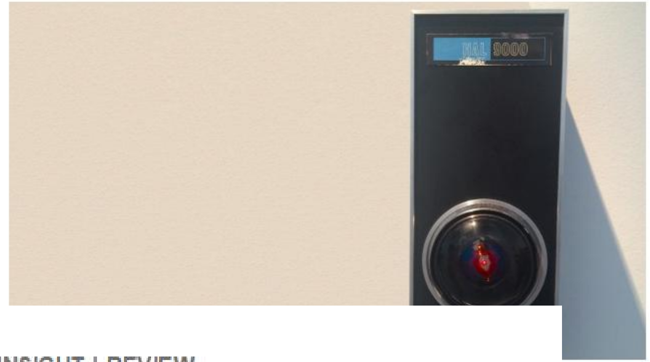
Learning 2017

Baidu's chief scientist explains why computers won't take over the world just yet

by Derrick Harris @derrickharris SEPTEMBER 23, 2015, 11:30 AM EDT

Are AI and "deep learning" the future of, well, everything?

Thanks to the advances in deep machine learning, technology companies across the globe are teaching computers to think for themselves



n your
Google
box and it

e,

Forbes / Tech

2 FREE Issue

DEC 29, 2014 @ 11:37 AM 82,136 VIEWS

Tech 2015: Deep Learning And Machine Intelligence Will Eat The World

Anthony Wing Kosner

Despite what Stephen Hawking or Elon Musk say, hostile Artificial Intelligence is not going to destroy the world anytime soon. What is certain to happen, however, is the continued ascent of the practical applications of AI, namely deep learning and machine intelligence. The word is spreading in all corners of the tech industry that the biggest part of big data, the unstructured part, possesses learnable patterns that we now have the computing power and algorithmic leverage to discern—and in short order.

For some of life's

NATURE | INSIGHT | REVIEW

Deep learning

Yann LeCun, Yoshua Bengio & Geoffrey Hinton

Affiliations | Corresponding author

Nature **521**, 436–444 (28 May 2015) | doi:10.1038/nature14539

Received 25 February 2015 | Accepted 01 May 2015 | Published

Citation Reprints Rights & permissions Article me

Abstract

Abstract • References • Author information

Deep learning allows computational models that are composed of n

IV. Details on the Lecture

The Lecture

» Technical foundation of approaches which are in focus today

» Mathematics

- » Linear Algebra (Vectors, Matrices, …)
- » Probability, Statistics
- » Optimization

» Often: Machine Learning is based on the minimization of a cost function (optimization) with respect to unknown parameters. The cost function is derived using probabilistic assumptions (probability) and model performance is analyzed by statistical methods. With quadratic loss functions, solutions can be derived with methods from linear algebra

Literature

Lecture

- *The Elements of Statistical Learning: Data mining, Inference and Prediction*. Hastie, Tibshirani, Friedman: Springer (2nd Ed.). [Modern Statistics; frequentist] Download at http://www-stat.stanford.edu/~tibs/ElemStatLearn/
- *Machine Learning: a Probabilistic Perspective. Kevin Murphy: MIT Press [very popular; Bayesian orientation]*
- *Bayesian Reasoning and Machine Learning. David Barber.* Download at *http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage [Bayesian orientation]*
- *Pattern Classification. Duda, Hart, Storck: Wiley [Pattern recognition]*
- *Pattern Recognition and Machine Learning*. Bishop: Springer *[Bayesian touch]*
- *Data Mining: Concepts and Techniques*. Han and Kamber: Morgan Kaufmann [Data mining]
- *Artificial Intelligence-a Modern Approach*. Russel and Norvig, Prentice Hall [All of AI]
- *Kernel Methods for Pattern Analysis.* John Shawe-Taylor and Nello Cristianini: Cambridge University Press [Kernel approaches]
- *Machine Learning*. Tom Mitchell: McGraw-Hill [Some excellent Chapters; some outdated]
- Andrew Ng's coursera course: https://www.coursera.org/course/ml
- *Deep Learning.* Ian Goodfellow, Yoshua Bengio, and Aaron Courville . *http://www.deeplearningbook.org/*

Literature (cont'd)

Time Series

- *Time Series Analysis*. Hamilton

Reinforcement Learning and Game Theory

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

Statistics

- *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
- *Statistik.* Fahrmeir, Kuenstler, Pigeot, Tutz: Springer
- *Introduction to Mathematical Statistics.* Hogg, Craig: Prentice Hall
- *Probability, Random Variables and Stochastic Processes.* Papoulis, McGraw, Hill

Cognition

- *Cognitive Neuroscience: The Biology of the Mind.* Gazzaniga, Ivry, Mangun, Norton
- *Learning and Memory: From Brain to Behavior. Gluck, Mercado, Myers, Worth*

Machine Learning im SS 2017

The lecture is given in English.

News

Topic

Machine Learning is a data-driven approach for the development of technical solutions. Initially motivated by the adaptive capabilities of biological systems, machine learning has increasing impact in many fields, such as vision, speech recognition, machine translation, and bioinformatics, and is a technological basis for the emerging field of Big Data.

The lecture will cover:

- Supervised learning: the goal here is to learn functional dependencies for classification and regression. We cover linear systems, basis function approaches, kernel approaches and neural networks. We will cover the recent developments in deep learning which lead to exciting applications in speech recognition and vision.
- Unsupervised Learning: the goal here is to compactly describe important structures in the data. Typical representatives are clustering and principal component analysis
- Graphical models (Bayesian networks, Markov networks), which permit a unified description of high-dimensional probabilistic dependencies
- Reinforcement Learning as the basis for the learning-based optimization of autonomous agents
- Some theoretical aspects: frequentist statistics, Bayesian statistics, statistical learning theory

The technical topics will be illustrated with a number of real-world applications.

Organisation

- **Umfang:** 3+2 Semesterwochenstunden (6 ECTS)
- **Vorlesung:** Prof. Dr. Volker Tresp
- **Vorkenntnisse:** Die Beherrschung mindestens einer Programmiersprache
- **Ansprechpartner Übungen:** Janina Bleicher
- **Anmeldung:** über UniWorX

Time and Location

Event	Time	Location	Start
Lecture	Wed 9.00 c.t. - 12.00 p.m.	Room **C 112 (Theresienstr. 41)**	26.04.2017
Exercise	Thu, 14.00 - 16.00 p.m.	Room 020 (Amalienstr. 73A)	04.05.2017
	Thu, 16.00 - 18.00 p.m.	Room M105 (Geschwister-Scholl-Platz 1)	04.05.2017

Planung

Vorlesung		Übung	
Datum	**Thema**	**Datum**	**Blatt**
26.04.17		27.04.17	entfällt
03.05.17		04.05.17	

- **Introduction:** learning from the perspectives of philosophy, logic and philosophy, psychology, neurobiology; history on technical learning
- **Basic learning machines:** Perceptron (linear classifier) and linear regression (linear predictor; linear regression); regularization; review on linear algebra
- **Basis functions:** Adding nonlinearity by a fixed transformation
- **Neural networks:** adding representational power by adapting the basis functions
- **Deep Learning:** Neural networks define the state of the art in some problems by using deep layers
- **Kernels:** adding representational power by using an infinite number of basis functions
- **Frequentist statistics and Bayesian statistics**; review on probability
- **Model comparison:** which model gives best results?
- **More linear classifiers and the support vector machine**
- A subset of: reinforcement learning, Bayesian networks, causality, factor models