Machine Learning

Volker Tresp
Professor, Ludwig Maximilian University of Munich
Distinguished Research Scientist, Siemens AG

Overview

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: Evolution

- Improvement via trial and error
- Biological evolution
- The “**Blind Watchmaker**“ (Richard Dawkins, 1986)
- Technical evolution: evolutionary improvement of technical solutions (advancement of the state of the art by trial and error in all competing companies world wide)

Advantages:

- Simple (blind); self-optimizing

Disadvantages:

- Time constant: years, decades, centuries, wasteful

Tegmark’s Life 1.0 (biological stage):

- Software/hardware is limited by slowly evolving DNA (bacteria)

2: Learning

Biological Learning:

- Lifelong optimization (improvement) of the behavior of an individual via interaction with the environment
- The **“Learning (Apprentice) Watchmaker**” that learns from a teacher and personal experience
- Basic properties of animals (“natural law”)
- Feedback of the learning success (reinforcement)
- Time constants: days

Machine Learning

- Broadest sense: attempt to mimic biological and human learning for technical purposes
- Autonomous optimization of a technical system via interaction with the environment or by analyzing acquired data; “learning instead of programming“

Tegmark’s Life 2.0 (cultural stage):

- Ability to design own software by learning but limited by slowly evolving hardware (that’s us)

3: Intelligent Design

- Almost all technical solutions are based on intelligent design
- Engineer: the “**Knowledgeable Watchmaker**“
- (We don’t mean the “Divine Watchmaker” of Fontenelle, 1686, and Paley, 1802)
- Programmer

Advantages:

- Explicit knowledge: the system is well understood and can be analyzed / improved via analytic thinking
- Time constant: years

Disadvantage:

- Need for an (expensive) designer (human)

Tegmark’s Life 3.0 (technological stage):

- Self design of software and hardware (AI future!)

Technical Progress

Advancing the state of the art of **engineering** by informed **trial and errors** by a set of **knowledgeable (learning) actors** (companies), driven by customer needs, competition and business opportunities

Occasional paradigm changes in technology and business models which might affect own business (search in service reports, digitalization) or might open up completely new businesses (Google, …)

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?)
- *Better Brain Hypothesis: My brain today is "better" than my brain yesterday*

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)

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

Overview

Non-Technical Perspectives

A	Philosophy
B	Psychology, Cognition and Cognitive Neuroscience
C	Cellular Neuroscience

A: Philosophy

Epistemology (*Erkenntnistheorie)*

- Epistemology is the theory of knowledge (and justified belief)
- What can mankind really know and how can mankind acquire knowledge (learn)?
- How can we know and study how the world functions
- What is knowledge?
- [Recall that until quite recently, in western civilization, these questions were only allowed to be addressed in a religious context]

Basic Mechanisms for Gaining Knowledge: 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 has blue eyes”) (data, measurements, Knowledge Graphs**)**
  - Can be **complex axioms** (“If something is a dog, it is also a mammal”)
- Founder: Aristotle (384 – 322 BC)
- Relationship to language theory:
  - Do we speak in axioms?
- Technical: basis for **classical Artificial Intelligence** (1954 ….)

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, often statistical, dependencies instead of assumed axioms!
- Closer to reality but learned dependencies might be difficult to explain
- Technical: basis for Machine Learning and modern **Artificial Intelligence**

Rationalism (17th Century)

- From Latin: ratio = „reason“: *let’s trust mostly our reasoning capabilities (complex axioms)*
- Priority of **rational reasoning** in knowledge acquisition (in contrast to other forms such as the senses or religious convention); search for optimal problem solution
- Representatives: Socrates (ca 470–399 BC), Plato (– 348/347 BC), René Descartes (1596–1650), Baruch Spinoza (1632–1677), Gottfried Leibniz (1646–1716)
- Since the enlightenment, rationalism is usually associated with the introduction of mathematical methods into philosophy, as in Descartes, Leibniz, and Spinoza. 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

Empiricism (17th Century)

- *Let's trust mostly our observations (facts: data, measurements, Knowledge Graphs)*
- The English term empirical derives from the Ancient Greek word empeiria, which translates to the Latin experientia; mostly a bottom-up approach
- More of a British tradition and 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 (individual/mankind)
- Representatives: Aristotle (384 – 322 BC), Francis Bacon (1562-1626), John Locke (1632-1704), David Hume (1711-1776)
- Aristotle/Locke: "One idea was thought to follow another in consciousness if it were associated by some principle" "The principal laws of association are contiguity, repetition, attention, pleasure-pain, and similarity. Both philosophers taught that the mind at birth is a blank slate and that all knowledge has to be acquired by learning. The laws they taught still make up the backbone of modern *learning theory*" (Wikipedia)
- Descriptive: how do humans act and acquire knowledge

Idealism (dominant Philosophy of the 18th/19th century)

- **Idealism:** each form of matter, including human behavior, is a reflection of ideas
- *Mind over matter: Let's trust mostly our mental self; people are driven by ideas, visions*
- 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
- It generally suggests the priority of ideals, principles, values, and goals over concrete realities
- Human ideas — especially beliefs and values — shape society
- Plato's idealism: reality is just a reflection of non-physical Ideas (ontological idealism). The material world is an illusion; also a tradition in Hinduism and Buddhism. Epistemologically, a skepticism about the possibility of knowing any mind-independent thing (also: Kant)
- George Berkeley (1685-1753)
- Immanuel Kant (1724-1804) (as an attempt of a synthesis of Rationalism and Empiricism)
- Beginning with Immanuel Kant, German idealists such as Hegel, Fichte, Schelling, and Schopenhauer dominated 19th-century philosophy

Positivism and the Scientific Method: (Beginning in the 19th Century)

- **Knowledge is derived from positive (in the sense of certain) findings**
- *Let’s trust mostly science*
- Positivism is defined as the belief that all true knowledge is scientific, and that all things are ultimately measurable (scientific revolution, Galileo Galilei)
- Data derived from sensory experience, and logical and mathematical treatments of such data, are together the exclusive source of all authentic knowledge
- <u>Even society</u> might operate according to laws like the physical world. Introspective and intuitional attempts to gain knowledge are rejected. Philosopher and founding sociologist, Auguste Comte: society operates according to its own laws, much as the physical. Also: Ernst Mach, Émile Durkheim
- Stephen Hawking: In “The Universe in a Nutshell” he writes: 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
- "The Unreasonable Effectiveness of Mathematics in the Natural Sciences" (Eugene Wigner, 1960)
- Materialism (Ludwig Feuerbach, Karl Marx): history is not driven by ideas but by laws (historic-dialectic materialism) (by most not considered to belonging to Positivism)

Neopositivism or Logical Positivism (20th Century)

- *Mathematical formalization of knowledge*
- Motivated by attempts to formalize all of mathematics by Gottlob Frege (1848-1925), and then, in the *Principia Mathematica*, by Bertrand Russell (1872-1970) and Alfred North Whitehead (1861-1947) [However, in 1931, Gödel's incompleteness theorem proved definitively that *Principia Mathematica*, and in fact any other attempt, could never succeed: For any set of axioms and inference rules proposed to encapsulate mathematics, either the system must be inconsistent, or there must in fact be some truths of mathematics which could not be deduced from them.]
- Logical positivists (or 'neopositivists') attempts to reduce statements and propositions to pure logic; Wiener Kreis, Rudolf Carnap (1891-1970); great influence in US
- Ludwig Wittgenstein (1889-1951): critical to the Wiener Kreis, but related, focusing on language

Reactions to Positivism

Anti-Positivism, Critical Theory: *Criticism as science as an ideology in sociology*

- Max Weber (1864-1920): sociology as a “science” as it is able to identify causal relationships. But: one seeks relationships not as "ahistorical, invariant, or generalizable“ as those pursued by natural scientists
- Rejections of 'scientism'; or science as ideology (Frankfurter Schule: Herbert Marcuse, Theodor Adorno, Max Horkheimer, Walter Benjamin, Erich Fromm, Jürgen Habermas)

Postpositivism: *Science is not independent of the scientist and of culture*

- Whereas for positivists: researcher and the researched person are independent of each other
- <u>Postpositivists: </u>theories/background/knowledge/values of the researcher can influence what is observed (Kuhn; critical for of the Wiener Kreis)

Critical Rationalism

- *Critical Rationalism* (Karl Popper (1902-1994): *Only falsifiable theories should be pursued* (If it's not falsifiable, it is not scientific)
- Is induction sound? Hume (1711-1776) advocated a practical skepticism based on common sense, where the inevitability of induction is accepted ("Will the sun rise tomorrow?"")
- 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 falsified (shown to be false) but never be proven (shown to be correct)!
- A theory has falsifiability or refutability if there is the possibility of showing it to be false: only falsifiable theories should be pursued
- The Logic of Scientific Discovery (Logik der Forschung, 1935)

Cultural Influence and Sociology of Science

Structure of Scientific Revolutions (Paradigm Shifts)

- The Structure of Scientific Revolutions (Thomas Kuhn, U.S. Historian of Science, 1922 –1996): 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; there can be no demarcation in terms of method between science and any other form of investigation
- Relativism has been criticized heavily!

Philosophy as a Basis for Classical Artificial Intelligence and Machine Learning

- Poppers *falsificationism* has great influence in science today, in general
- Logical positivism (Ludwig Wittgenstein) as a motivation for early AI research (dominance of logic-based top-down approaches)
- Empiricism as a basis for Machine Learning

- Note: almost all theories would agree on simple facts, axioms, measurements; the dispute is on the possibility to generalize from simple facts, and statements that are not simple facts

B: Psychology, Cognition and Cognitive Neuroscience

- Psychology is the study of the faculty of the mind, in particular of learning and behavior
- Whereas philosophy is concerned with the question, what mankind can learn and know, here the question is how an individual learns
  - **Psychology:** humans are mostly the system that produces data to be studies (e.g., happiness as a function of income; child development)
  - **Cognition:** also models the inner working of the system, i.e., the brain (Bayesian brain)
  - **Cognitive Neuroscience:** connecting psychology and cognition with neuroscience (e.g., the role of the hippocampus in episodic memory)

Psychology as Empirical Science

Begin of empirical (experimental) psychology:

- Herrmann von Helmholtz (1821-1894, Berlin)
- Wilhelm Wundt (1832-1920, Leipzig) (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”
  - First experimental psychological lab worldwide
- Gustav Theodor Fechner (1801–1887, Leipzig): Founder of Psychophysics
  - “The Scientific Study of the Relation between Stimulus and Sensation“
- Hermann Ebbinghaus (1850-1909): first rigorous experimental studies on human memory

Psychoanalysis and Psychiatry

- Psychoanalysis was founded by Sigmund Freud (1856-1939)
- Hypothesis: people can be cured by making conscious their unconscious thoughts and motivations, thus gaining “insight”

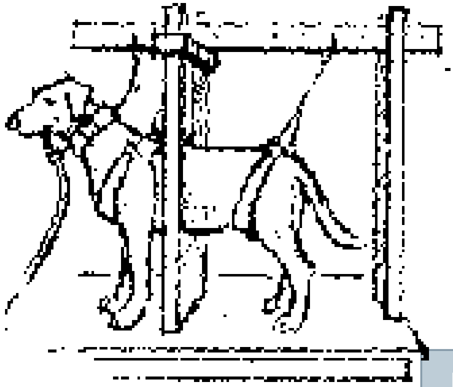
- Psychoanalysis is not part of psychology, and by some critics regarded as a pseudoscience (difficulty with falsification)

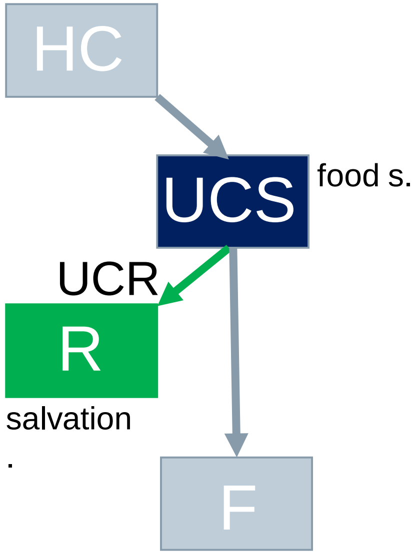
- Psychoanalysis maintains a strong influence on **psychiatry** (a branch of medicine: diagnosis, prevention, study and treatment of mental disorders)

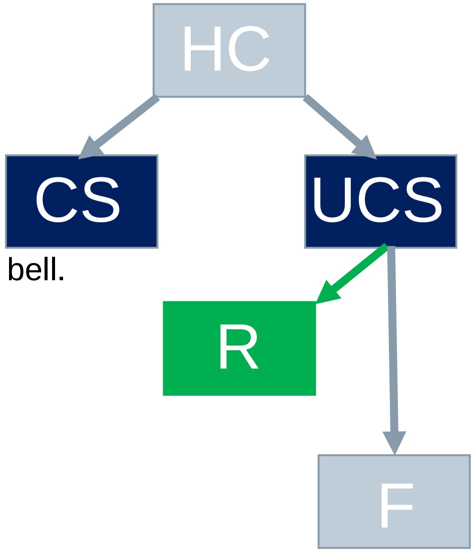
Behaviorisms (1920-1960)

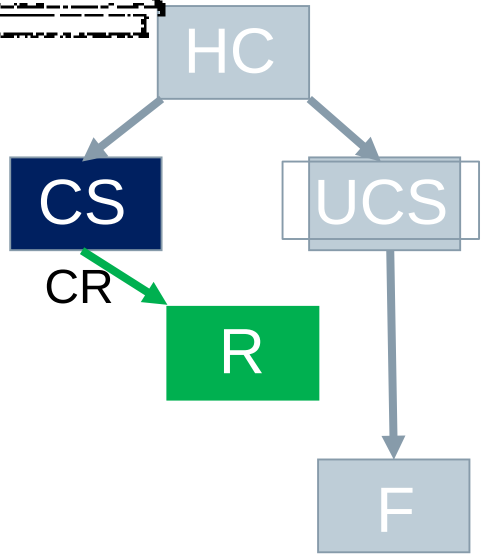
- „Belief in the existence of consciousness goes back to the ancient days of superstition and magic“
- Founded also as reaction to Sigmund Freud’s Psychoanalysis
- Rejection of theories that need to assume mental states (no interest in explainability?)
- The inner structure (of the brain) is irrelevant
- 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)









Unconditional response, UCR: The unconditional stimulus (UCS, food smell) produces the response (R, salvation), since the dog has learned that after food-smell comes food

A hidden cause (HC, experimenter) might produce both the UCS and the conditional response (CS, bell) which might come slightly earlier than the food smell

Conditional response (CR): The dog leans that after the CS, food follows and starts salvation, even without the UCS

- As a predictive model this makes absolutely sense but it might not reflect causality or be interpretable
- **Good predictive performance with difficult interpretability is a core issue in Machine Learning**

Another example. **UCR**: A stomach virus (UCS) would produce a response of nausea (R). **CR**: Chocolate (CS) which was eaten before a person was sick with a virus now produces a response of nausea

Learning in Psychology

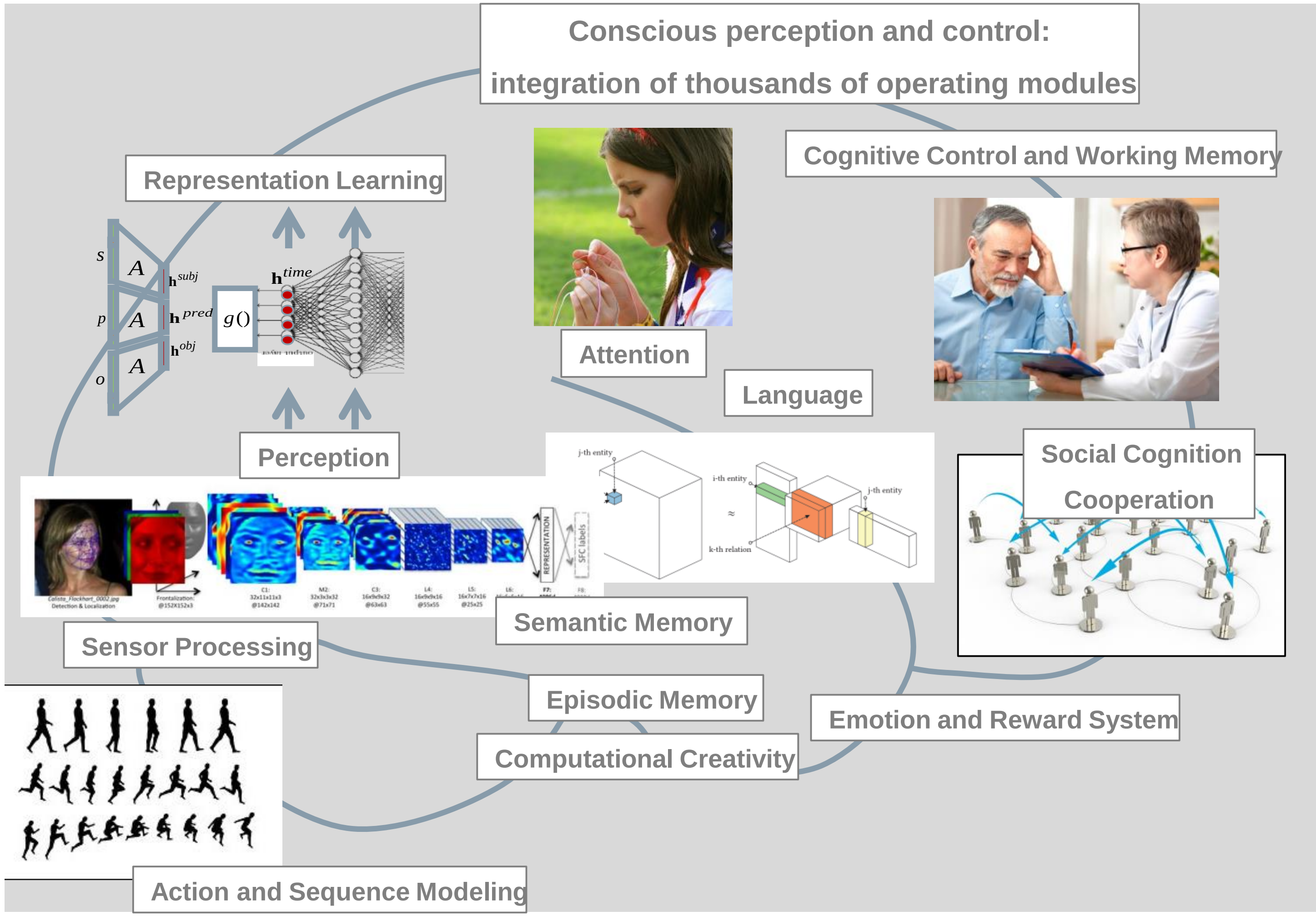
- **Habituation, Sensitization, Familiarization**
  - Learning about repeated events
- **Classical Conditioning**
  - By learning the association (Bell~Food); learning to predict
- **Operand Conditioning**
  - Learning the outcome of behavior (Learning to Act)
  - Stimulus => Response => Outcome
  -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

Cognitive Psychology and Cognition

- Attempt to understand the inner working of the „Black Box“
- Reaction to Behaviorism [whose scientific methodologies are the basis for much of psychology today]
- Human behavior is more than stimulus-response: Development is an active process of a subject
- Reintroduction of mental processes: explainability
- 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 Neuroscience

- Personally, one of my favorite areas
- Combines psychology and cognition with modern research in neuroscience
- Integrated understanding of the human mind
  - From behavioral studies, structural analysis of the brain (including MRI, Diffusion Tensor Imaging), EEG, fMRI, insights from brain damage, …
- Highly recommended book:
  - Gazzaniga: Cognitive Neuroscience



The Organization of Memory and Learning: Better Brain Hypothesis

Declarative Memories

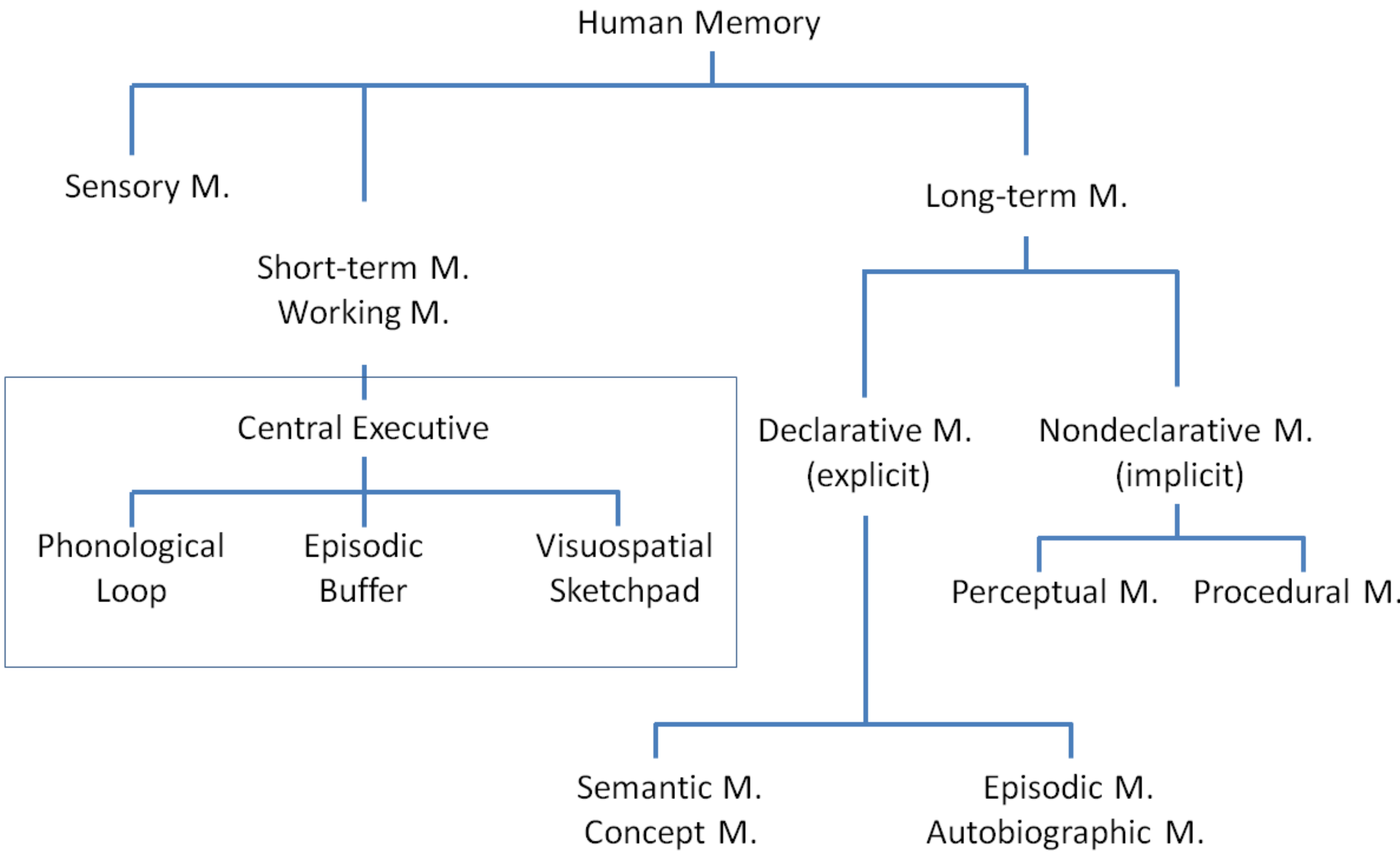
- Episodic Memory (evens we remember)
- Semantic Memories (facts we know)

Nondeclarative memories

- Skills that we have learned
- Learning to perceive and act better

Working Memory: Central Executive

More (!) Insight by introspection? Learned causality and rules? Logical and other forms of reasoning? Learning to make the right decisions. Role of language in all of his? Learned in school? Learning about social roles (mine and others) and how to improve mine. Learning how others feel. Reinforcement Learning and internal rewards (dopamine)



Influence on Machine Learning

Psychology

- The statistical approach of **psychology** greatly influenced Machine Learning
- A Neural Network as a model that predicts the response from stimuli, and over inputs

Cognition

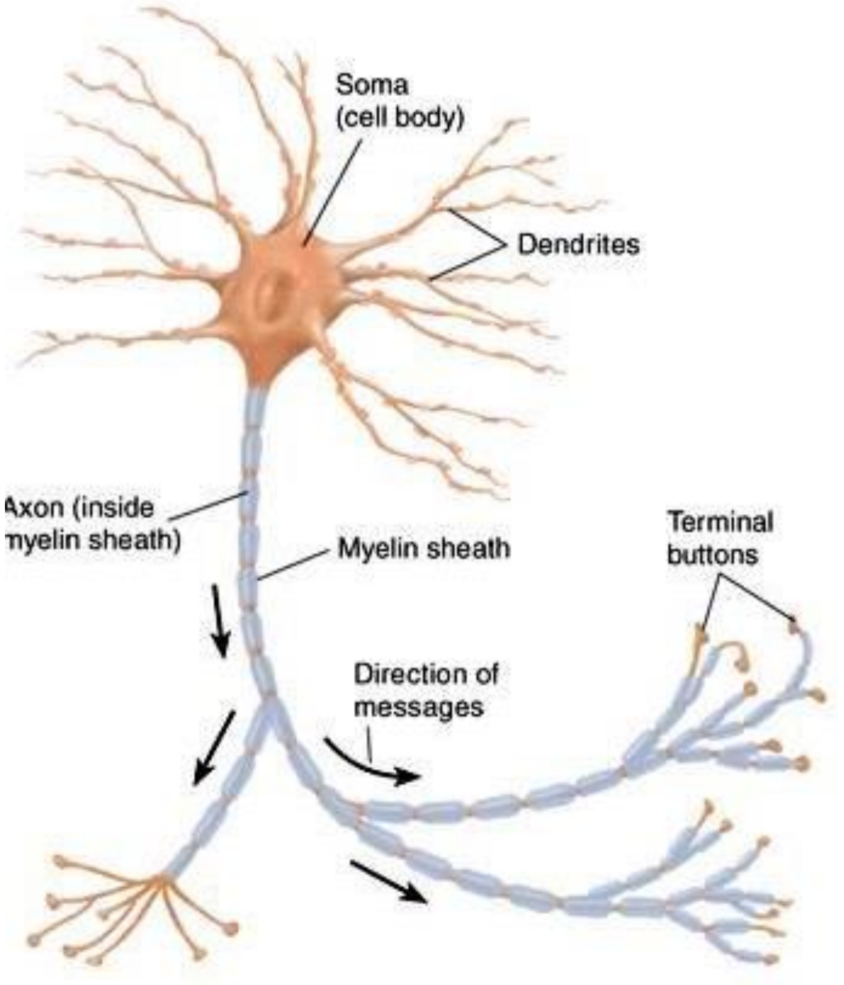
- Relating the inner working of a Neural Network to Cognition and the inner working of the brain
- Machine Learning motivates much research in Cognition
- Two of the co-inventors of the multilayer perceptron, (David Rumelhart and Geoffrey Hinton) are **cognitive psychologists**
- Geoffrey Hinton: founder of Deep Learning

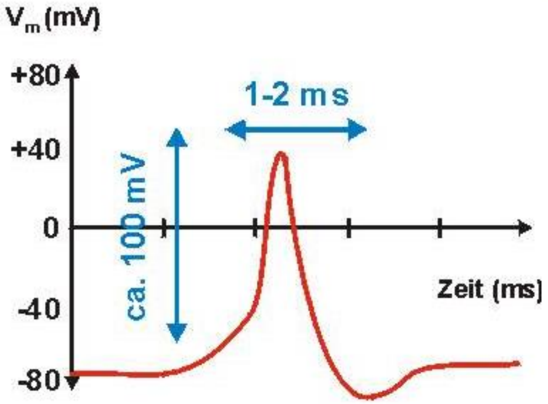
C: Cellular Neuroscience

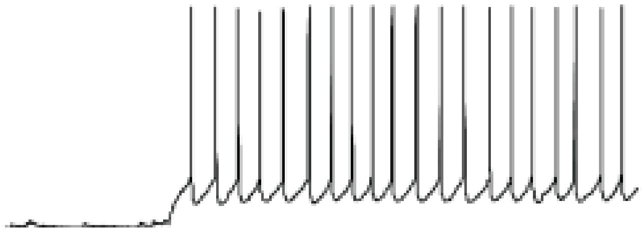
- Whereas before, we were concerned about a systemic view on brain function, here one is interested in the elementary mechanisms
- Cellular neuroscience is the study of neurons at a cellular level
- There must be a physical change if something is learned or memorized
- Central mechanism: Synapses change 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

- Integrate and fire neuron
- Resting potential: -70 mV.
- Depolarization (by inputs from dendrites):
  - > -50mV
- Leads to an opening of the sodium channels; this leads to the generation of on action potential, which is transmitted via the axon
- Refractory period: during this time no new action potential can be generated (app.. 2ms)
- Systems theory: leaky integrator
- In some models: The firing rate is interpreted as a continuous neural output

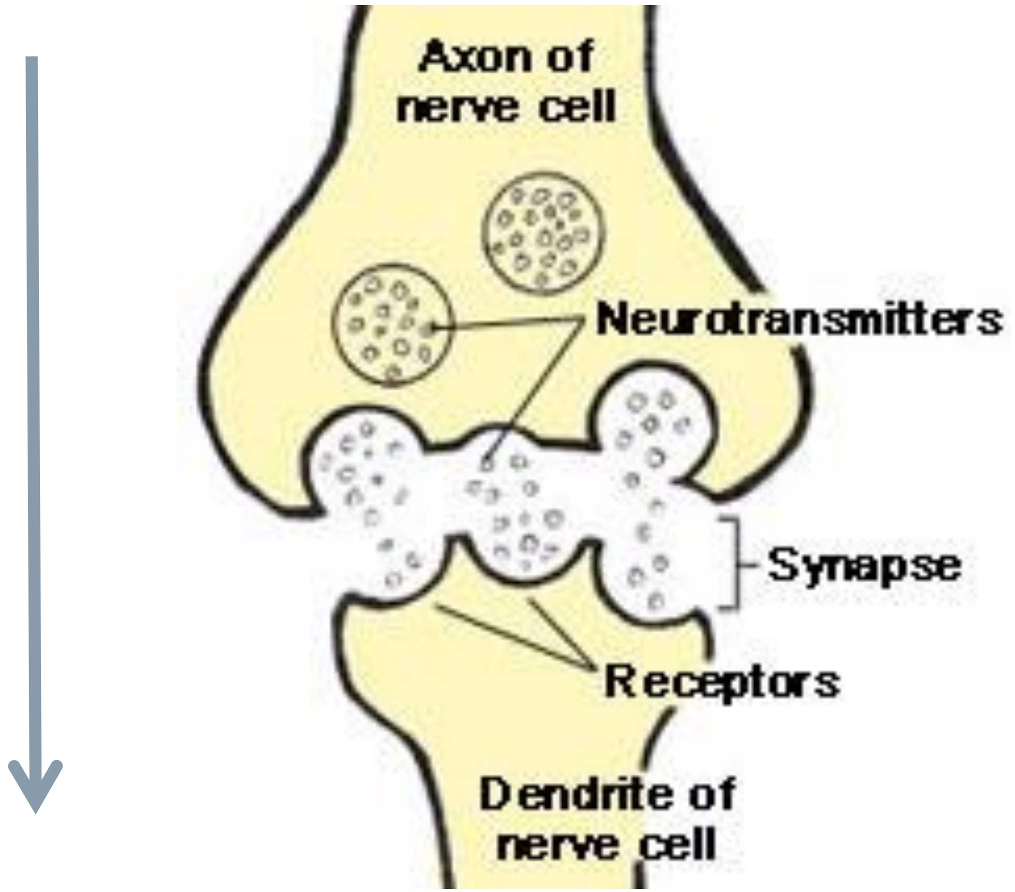






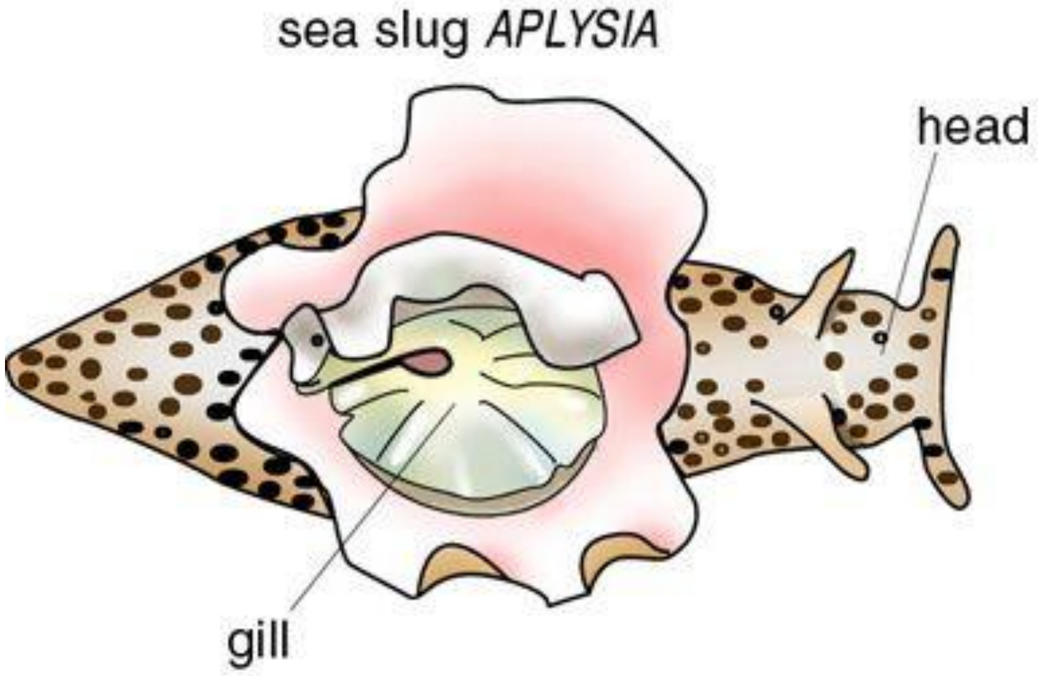
Synapse

- **Presynaptic (Axon):**
  - Presynaptic discharge of the action potential leads to the release of neurotransmitters (10 important types, but maybe 100 in total; from small to very large molecules)
- **Postsynaptic (Dendrite):**
  - Opening of ion channels and thus change of the postsynaptic membrane potential



Aplysia

- Eric Richard Kandel (1929-): US-American Neuroscientist with Austrian origin (Nobel price 2000)
- Study object: Californian see slug (Aplysia californica)
- 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



Hebb Learning in Psychology and 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."
- In short:
  - “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 confirmed biologically, i.e., in the neurons of the hippocampus
- Hebb formulates learning much more abstractly than Kandel
  - Open question in Machine Learning: how much can one ignore biological details without loosing the essence (e.g. spiking, spike timing?)

Cellular Neuroscience and Neural Circuitry

- Cellular Neuroscience and Neural Circuitry are currently not a major focus in Machine Learning but are areas of interest
- Machine Learning tries to maintain some of the inherent properties of biological learning:
    - Distributed local computing: formalized neurons as building blocks
    - Multilayered hierarchical computing
    - Convolution architectures
    - Noise and fault tolerance: graceful degradation
- The neurobiological plausibility of Machine Learning architectures and algorithms is sometimes hotly debated (is the backpropagation learning rule biologically plausible?)

Overview

Machine Learning

A	Before the Computer Age to Today: Statistics
B	Neural Computation I (Perceptron)
C	Classical Artificial Intelligence
D	Neural Computation II (Multilayer Perceptron)
E	Mathematically Well-Founded Models
F	Neural Computation III (Deep Learning); AI II

A: Before the Computer Age to Today: Statistics

- Thomas Bayes (1701 -1761): Updating the degree of belief in hypothesis based on observations (his work was published after his death by Richard Price)
- **A priori assumption (top-down)**: *P(H=1), with P(H=0) = 1- P(H=1)*
  - Degree of belief in the truthfulness of a hypothesis *H* (top-down)
- **Likelihood (bottom up):** *P(D|H=1) and P(D|H=0)*
  - *P(D|H=1):* Plausibility of the data *D* (observations), if the hypothesis *H* is true
  - *P(D|H=0):* Plausibility of the data *D* (observations), if the hypothesis *H* is false
- Then, the **a posteriori probability** (Bayes called it "inverse probability") of the hypothesis being true is given by Bayes' theorem:
  - *P(H=1|D) = P(D|H=1) P(H=1) / P(D)*
- With: *P(D) = P(D|H=1) P(H=1) + P(D|H=0) P(H=0)*
- Learning: updating ones prior belief based on data

Footnote: In some cases one might be able to formulate *P(H|D)* directly, but in most cases, is easier to specify *P(D|H)* and then apply Bayes' formula; in particular if *D* contains many observations

Bayesian Statistics

- In statistics, the Bayesian interpretation of probability was developed mainly by Pierre-Simon Laplace (1749–1827)
- Bayes rule is a theorem in the mathematical probability theory
- The application of Bayes rule to problems in the real world is disputed (in particular if hypothesis on parameter distributions are used)
- Bayesian Statistics has great influence in the sciences, in particular in Bayesian Machine Learning and Cognition (Bayesian Brain Hypothesis)

- Example:
  - $P(H = Bronchitis \mid D = PositiveXRay)$

    $= P(D = PositiveXRay \mid H = Bronchitis)\, P(H = Bronchitis) / P(D = PositiveXRay)$

Bayesian Statistics is Based on the Concept of Subjective Probability

- With subjective probabilities I can make personal statements like:
  - *Before I throw a coin, I believe that the coin is a fair coin with 99%*
  - *I believe that the probability that party X wins the election is 45%*

- Cox's theorem implies that any plausibility model that meets his postulates is equivalent to the subjective probability model, i.e., can be converted to the probability model by rescaling (Richard Threlkeld Cox, 1898–1991)

- 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

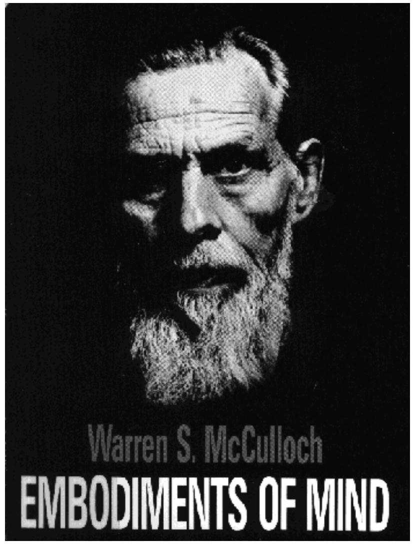
Critique on Bayesian Statistics

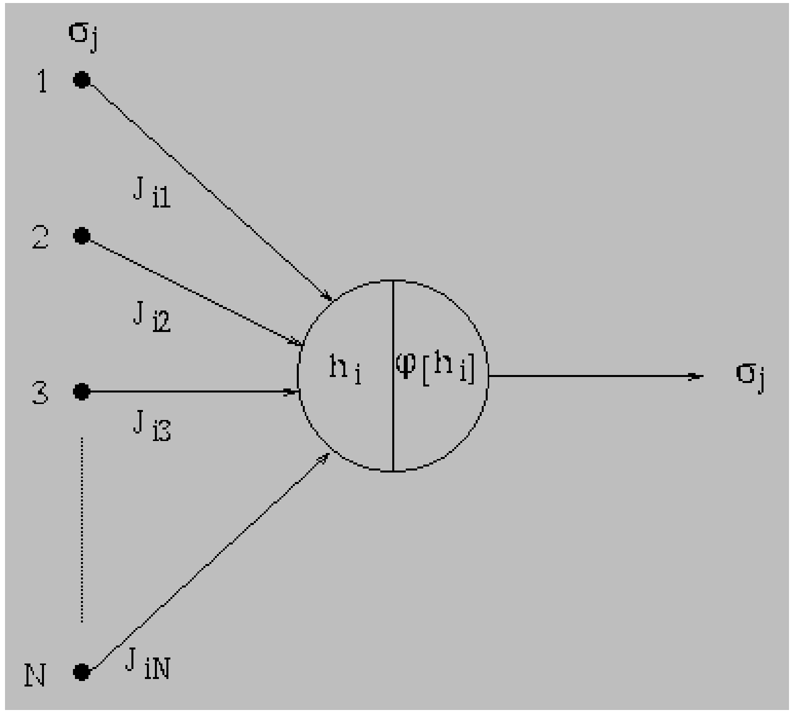
- **Karl Pearson** (1857–1936)
  - now considered the founder (“godfather”) of modern statistics (also referred to as classical or frequentist statistics)
  - "I felt like a buccaneer of Drake's days -... I interpreted that sentence of Francis Galton (1822-1911) [his advisor] to mean that there was a 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."
- **Ronald Aylmer Fisher** (1890-1962)
  - Criticism 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
- Other founding contributors: **Egon Pearson** (1895-1980), son of Karl P.; **Jerzy Neyman** (1894-1981)

Footnote: Incidentally: all of them did not only reject Bayesian statistics, but also causal analysis

B: Neural Computation I

- Started with a focus on expressiveness of Neural Networks (and not their ability to learn)
- McCulloch and Pitts (1943)
  - First attempt to formalize brain functions via networks of simple computational nodes (network of simple logical units)
  - McCulloch-Pitts Neuron





Expressiveness of Neural Networks

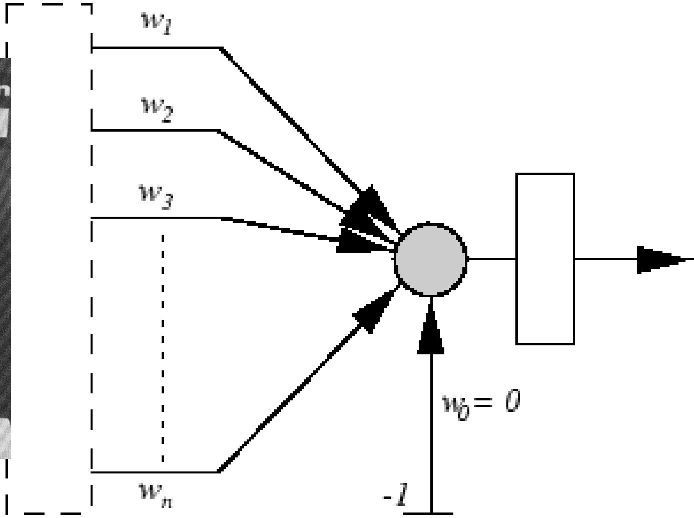
- John von Neumann (1903-1957) investigated the error tolerance of Neural Networks
  - “Computer and the Brain” (book, 1958)
- John von Neumann concluded that the brain operates in part digitally, in part analogically, but uses a peculiar statistical language unlike that employed in the operation of 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 (1954)
  - “Cybernetics or Control and Communication in the Animal and the Machine” (book, 1948)

Perceptron and ADALINE

- **Marvin Minsky** (1927-2016) developed 1954 in his dissertation a neural computer he called SNARC (Stochastic Neural Analog Reinforcement Calculator)
- **Frank Rosenblatt** (1928-1971) developed in 1958 the *Perceptron learning rule* and formulated a convergence proof; Mark I Perceptron
    - Basis for MLP and Deep Learning
- **Bernard Widrow** (1929-) and **Ted Hoff** (1937-) developed in 1960 the *ADALINE* (ADaptive LINear Element)
    - Foundation of adaptive signal processing
- Marvin Minsky and Seymour Aubrey Papert (1928-2016) published 1969 the book „Perceptrons“ and demonstrated the limitations of the Perceptrons and of the ADALINE (Exclusive-Or Problem)



C: Classical Artificial Intelligence

- 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 (1950s to 1980s)

C: Classical Artificial Intelligence

- 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)

Roots of AI in Philosopy

- Motivated by attempts to formalize all of mathematics by Gottlob Frege (1848-1925), and then, in the *Principia Mathematica*, by Bertrand Russell (1872-1970) and Alfred North Whitehead (1861-1947)
- Logical positivists (or 'neopositivists') attempts to reduce statements and propositions to pure logic; Wiener Kreis, Rudolf Carnap (1891-1970); great influence in US (after 1936, worked in the US)
- Ludwig Wittgenstein (1889-1951): critical to the Wiener Kreis, but related, focusing on language

Perspectives of AI

- **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 (tradition in Rationalism)
  - Rational (optimal) Reasoning: Logic
  - Rational (optimal) Acting: Agents

- 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 (tradition in Empiricism)
  - Understanding human thinking (Cognition)
  - Indistinguishably from human acting (Turing Test)

Birth of AI: Dartmouth Workshop (1956)

- **John McCarthy** (Dartmouth, later Stanford) (1927-2011)
  - Suggested the term *Artifical Intelligence* (to distinguish it from Cybernetics); inventor of LISP
- **Marvin Minsky** (1927-2016) (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** (1903–1957) 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**

Early Classical AI Phase

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“
- 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)

- Project to translate 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

Late Classical AI Phase

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

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

- McDermott: R1 (DEC, 1982); Configuration of computer systems; each major company has an AI group
- Fifth Generation Project in Japan (1981); 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 (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)

D: Neural Computation II

- 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; in 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

Hopfield Networks

- John Hopfield (1933-): Neural networks and physical systems with emergent collective computational abilities (1982, 1984)
- Achievements:
  - Associative content-addressable memory (Hebb learning)
  - Solving combinatorial optimization problems (travelling salesman) (new interest in Hopfield nets in in quantum computing, e.g., quantum annealing)
- Important: he made the link to statistical physics (spin-glasses), which brought in many physicists
- Interesting computational features: nonlinear, parallel, error tolerant, with feedback
- Implementation as optical computer?
- Relationship to brain functioning?
  - At the end: solutions were not technically competitive
- Prior work on associative memory: W. K. Taylor (1956); Karl Steinbuch (1961); James A. Anderson (1968), D. J. Willshaw (1969), Stephen Grossberg (1967), Teuvo Kohonen (1974)
- Prior work on links to statistical physics: E. R. Caianiello (1961), W.A. Little and Gordon L. Shaw (1975)

Boltzmann Machine, Multilayer Perceptron

Ackley, Hinton, Sejnowsky (1985): Boltzmann Machine

- Discriminative Learning; close connection to Statistical Physics
- Theoretically very interesting but not as practical as the MLP
- Come-back in Deep Learning as Restricted Boltzmann Machine

Rumelhart, Hinton, Williams (1986): Multilayer Perceptron (MLP)

- MLP: a robust powerful tool for modeling high-dimensional nonlinear dependencies
- Beyond the limitations of the Perceptron: Solution to the exclusive-or-problems, Nettalk
- MLP: superior modeling tool for high-dimensional problems
- Breakthrough in Neural Computation
- Interest in Statistics and from statisticians

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

E: Mathematically Well-Founded Models

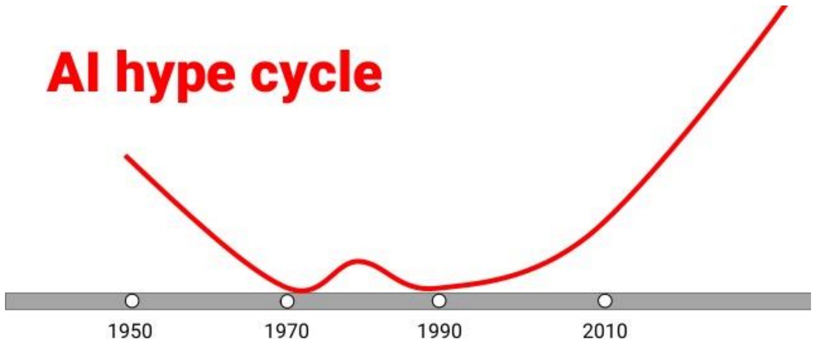
- In the middle 1990s the interest in neural networks slowly faded
  - Limited interest by other fields in this “heuristic” approach
  - Connection to Neuroscience became less strong
- Neural computation transformed into Machine Learning
  - (The term was used earlier only for symbolic machine learning)
- Keywords: Statistical Machine Learning; Bayesian networks and Graphical Models; VC-Theory; Kernel Systems; Gaussian Processes; Infinite Models
- Most famous model: Support Vector Machine
- High recognition and deep impact on other fields: Bioinformatics, Vision, Natural Language Processing, Information Retrieval, Search Engines, …
- AI increasingly became influenced by Machine Learning (other fields became NIPS-ified)

F: Neural Computation III, AI II

- By around 2010 theoreticians dominated the core of Machine Learning
- The field converged to a number of beautiful theories but in terms of practical usefulness, there was little progress (in some areas, the SVM excelled, but in some applications, a neural network from 1986 was equally good and sometimes even better than an SVM)
- This situation totally changed with deep learning
  - Sometimes after 2010 the deep learning revolution got started
  - Big Names; Geoffrey Hinton, Yoshua Bengio, Yann LeCun, Andrew Ng, Jürgen Schmidhuber, …
- Some difficult benchmarks were improved by a factor of 10 (ImageNet)!
- Dominates: Vision research, NLP, …
- Immediately in products: Speech recognition, face recognition, language translation, …
- The perceived AI revolution is due to Deep Learning

Artificial Intelligence

- *Creating machines that perform functions that require intelligence when performed by people (Kurzweil, 1990)*

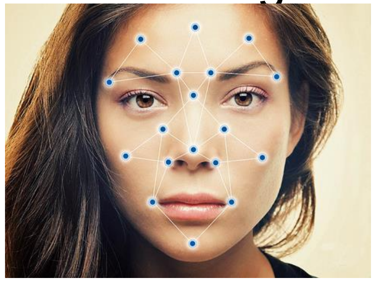


Games

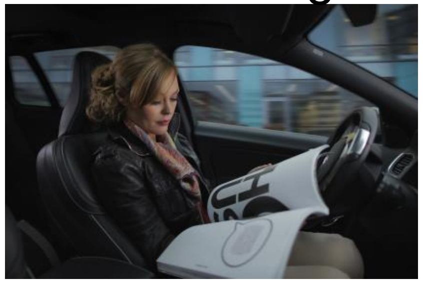
Translation

Q&A

Face Recognition



Auton. Driving



Speech Recognition

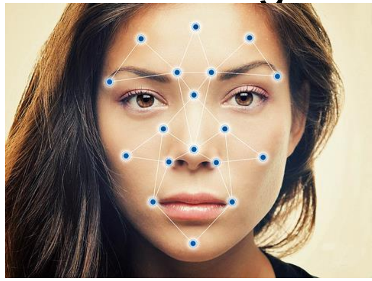
Drones, Robots

?

Deep Learning

- Deep Learning is the reason for the emerging huge interest in AI
  - Convolutional DL
  - Recurrent DL
  - Reinforcement DL

Face Recognition



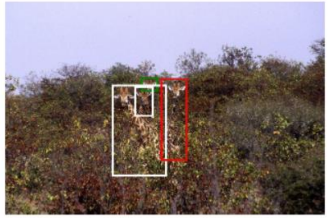
Translation

Speech Recognition

Games

Student Magic: Visual Q&A

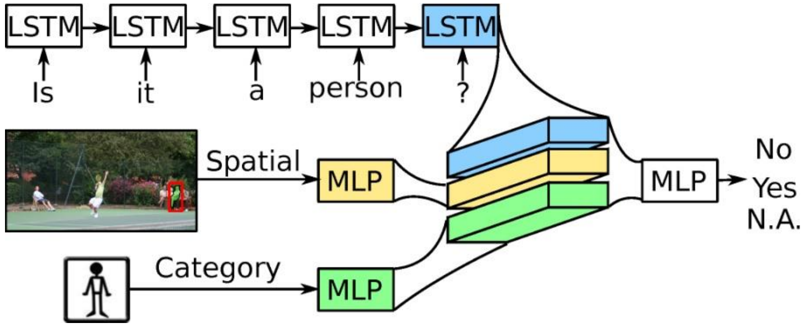
"I spy with my little eye ..."

Image	Policy Gradient		Tempered Policy Gradient	
	Is it in left?	No	Is it a person?	No
	Is it in front?	No	Is it a **vehicle**?	Yes
	Is it in right?	Yes	Is it a **truck**?	Yes
	Is it in middle?	Yes	Is it in front of photo?	No
	Is it person?	No	In the left half?	No
	Is it ball?	No	In the middle of photo?	Yes
	Is it bat?	No	Is it to the right photo?	Yes
	Is it **car**?	Yes	Is it in the middle of photo?	Yes
	Status:	Failure	Status:	Success
	Is it in **left**?	No	Is it a giraffe?	Yes
	Is it in front?	Yes	In front of photo?	Yes
	Is it in right?	No	In the **left half**?	Yes
	Is it in middle?	Yes	Is it in the middle of photo?	Yes
	Is it person?	No	Is it **to the left** of photo?	Yes
	Is it giraffe?	Yes	Is it to the right photo?	No
	Is in middle?	Yes	**In the left** in photo?	No
	Is in middle?	Yes	In the middle of photo?	Yes
	Status:	Failure	Status:	Success

Convolutional DL
+ Recurrent DL
+ Reinforcement DL

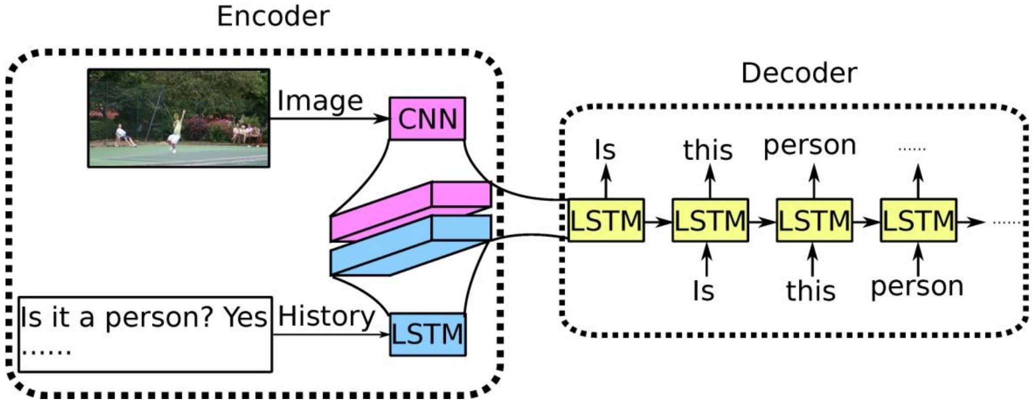
Talents, Talents Talents!

The Oracle Model

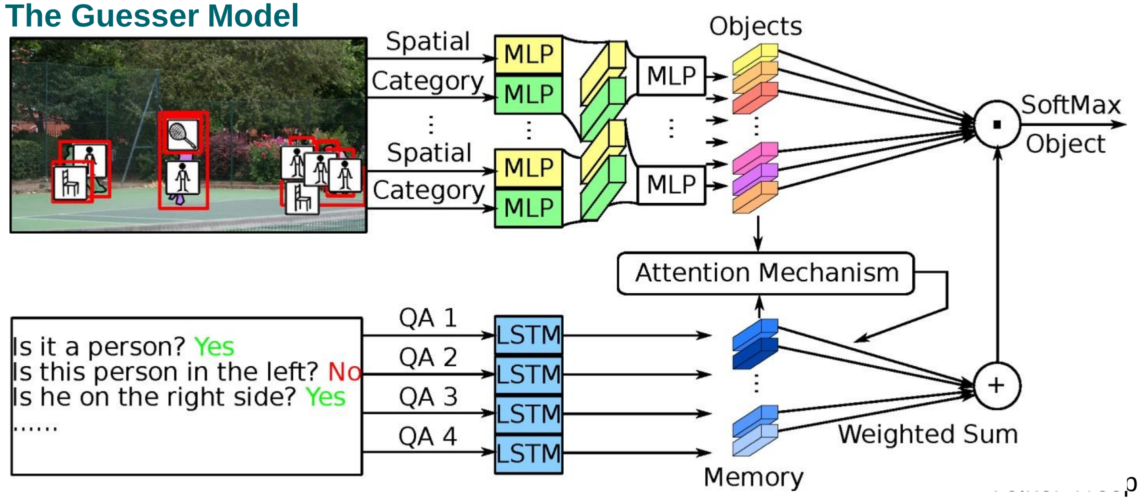


Rui Zhao, 2018

The Question-Generator Model



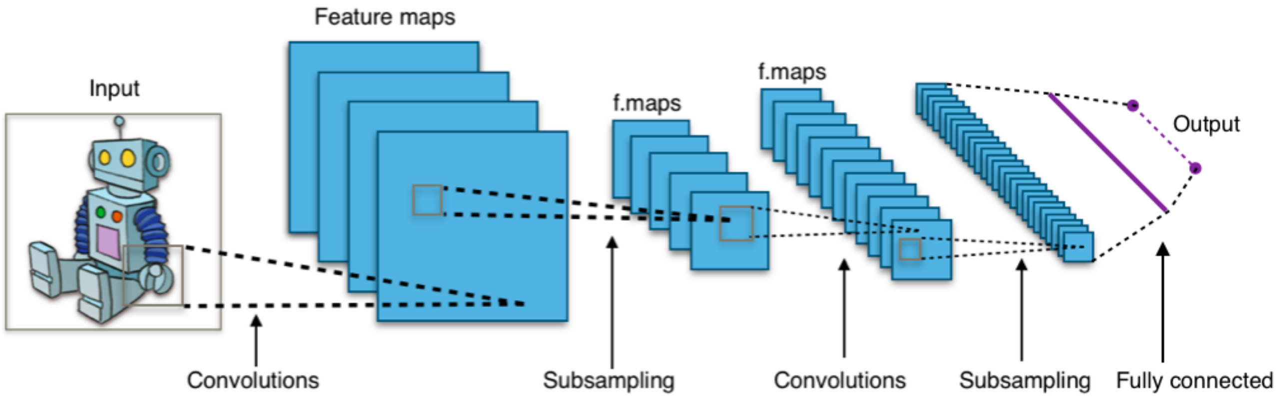
The Guesser Model

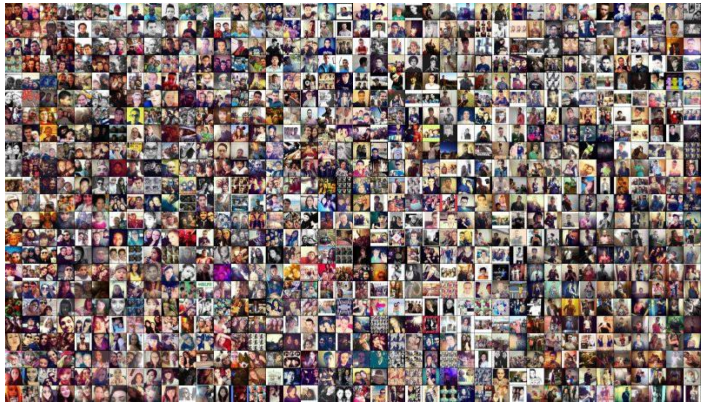


Deep X Technologies behind Artificial Intelligence

- **Deep Learning**; Machine Learning; Data Mining; Statistics
  - More (Labeled) Data
  - Deeper Models
  - New Algorithms
  - End-to-End Training; Differentiable-Computing
  - Computational Power
  - Community

- **Deep Knowledge**
  - Huge Document Repositories with Rapid IE / QA (IBM Watson)
  - Maps with GPS for Autonomous Driving
  - Ubiquitous Big Data in Industry
  - Detailed (Patient) Profiles
  - Web Content, Wikipedia for Humans
  - **Knowledge Graphs for Machines**





Current Trends

- Deep Learning: Whatever-works-best mentality (*ResNet* with hundreds of layers)
- Incredible talents that are eager to attack any problem of relevance
- Explainable AI
- Drive towards General AI; AI complete systems
- Personally I believe the most interesting and promising path to AI is via an understanding of the brain from a cognitive neuroscience perspective
- AI ethics (should we feel threatened? Jobs? As mankind? Military applications?)
- Hype problems: Big Data, Digitalization, ML, DL, AI, …
  - People in the field are quite aware of the limitations of current AI but might also benefit from the hype (both in academia and industry) (the term AI was pushed by the press)
  - Immediate mega business? There will be continuous progress in many fields but it is not clear how fast a “mega” business will develop outside of the companies with a clear digital business (Google, Facebook, Amazon, …)

The Organization of Memory and Learning: Better Brain Hypothesis

Declarative Memories

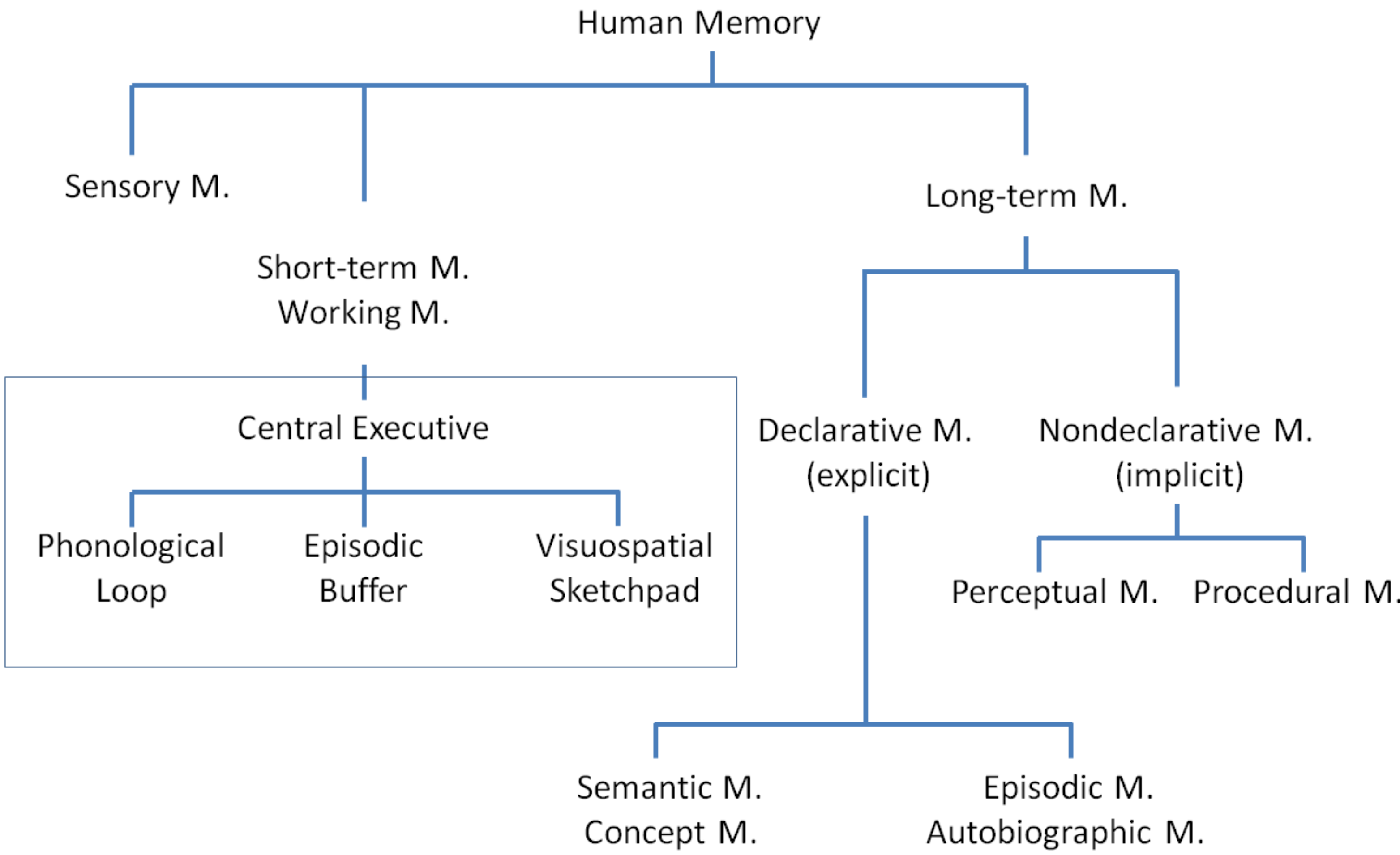
- Episodic Memory (evens we remember)
- Semantic Memories (facts we know)

Nondeclarative memories

- Skills that we have learned
- Learning to perceive and act better

Working Memory: Central Executive

More (!) Insight by introspection? Learned causality and rules? Logical and other forms of reasoning? Learning to make the right decisions. Role of language in all of his? Learned in school? Learning about social roles (mine and others) and how to improve mine. Learning how others feel. Reinforcement Learning and internal rewards (dopamine)



Overview

The Lecture

- Technical foundation of approaches which are in focus today
- From applied mathematics
  - Linear Algebra (vectors, matrices, …)
  - Calculus: how to calculate derivatives, chain rule, …
  - Optimization: necessary conditions to be in an optimum; gradient descent optimization
  - Probability, Bayesian and Frequentist Statistics
    - Lectures from different experts agree on what should be done, but might differ in in the reason why something should be done (PAC, VC-theory, Frequentist Statistics, Bayesian Statistics, …)
    - I present the path I find most useful

- Often: Machine Learning is based on the minimization of a data-dependent cost function (optimization) with respect to unknown model 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 by setting derivatives to zero (calculus)

Recommended Literature

- **Deep Learning**. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. http://www.deeplearningbook.org/
- Andrew Ng's Coursera and Stanford courses on Machine Learning and Deep Learning
- 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]
- Pattern Classification. Duda, Hart, Storck: Wiley [Pattern recognition]
- Pattern Recognition and Machine Learning. Bishop: Springer [Bayesian touch]
- Artificial Intelligence-a Modern Approach. Russel and Norvig, Prentice Hall [All of AI]
- Machine Learning. Tom Mitchell: McGraw-Hill [Some excellent Chapters; some outdated]
- Understanding Machine Learning: From Theory to Algorithms. Shai Shalev-Shwartz and Shai Ben-David. Cambridge University Press. (covers VC-Theory, Statistical Learning Theory)

Specific Topics

- **Data Mining:** Concepts and Techniques. Han and Kamber: Morgan Kaufmann [Data mining]
- **Kernel Methods** for Pattern Analysis. John Shawe-Taylor and Nello Cristianini: Cambridge UP
- **Reinforcement Learning**: an Introduction. Sutton and Barto: MIT Press
- **Bayesian Data Analysis**. Gelman, Carlin, Stern, Rubin: Chapman
- **Statistik.** Fahrmeir, Kuenstler, Pigeot, Tutz: Springer (introduction to classical statistics)
- **Probability**, Random Variables and Stochastic Processes. Papoulis, McGraw, Hill
- **Cognitive Neuroscience:** The Biology of the Mind. Gazzaniga, Ivry, Mangun, Norton

Machine Learning (SS 2019)

News

- **28.02.2019:** The registration for this course via UniWorx will be open from April 1st, 2019 onwards.

Organisation

- **Course:** 3+2 hours weekly (equals 6 ECTS)
- **Lecture:** Prof. Dr. Volker Tresp
- **Assistants:** Christian Frey, Sabrina Friedl
- **Required:** Professional skill of at least one programming language
- **Audience:** The course is directed towards master students in informatics, bioinformatics and media informatics
- **Course Language:** English

Time and Locations

All times are c.t. (cum tempore)

Component	When	Where	Starts at
Lecture	Thu, 9,00 - 12,00 h	Room M 218 (Geschwister-Scholl-Platz 1)	25.04.2019
Tutorial 1	Tue, 14,00 - 16,00 h	Room B 015 (Geschwister-Scholl-Platz 1)	30.04.2019
Tutorial 2	Tue, 16,00 - 18,00 h	Room B 015 (Geschwister-Scholl-Platz 1)	30.04.2019
Tutorial 3	Wed, 14,00 - 16,00 h	Room B 015 (Geschwister-Scholl-Platz 1)	01.05.2019
Tutorial 4	Wed, 16,00 - 18,00 h	Room B 015 (Geschwister-Scholl-Platz 1)	01.05.2019

Lecture		Tutorial	
Date	**Topic**	**Date**	**Topic**
12.04.18	Lecture 1: Introduction	17.04.18 18.04.18	Python Introduction (.ipynb) Suggested Solutions (.ipynb)
19.04.18	Lecture 2: Linear Algebra (Review), Perceptron, Linear Regression	24.04.18 25.04.18	Exercise Sheet 1 Exercise 1-3 (.ipynb) Suggested Solutions Exercise 1-3 Solutions (.ipynb)
26.04.18	Lecture 3: Basis Functions, Neural Networks	01.05.18 02.05.18	**no tutorials (May Day)**
03.05.18	Lecture 4: Deep Learning, Manifolds	08.05.18 09.05.18	Exercise Sheet 2 Tensorflow Introduction (.ipynb) Exercise 2-5 (.ipynb) Suggested Solutions Exercise 2-5 Solutions (.ipynb)
10.05.18	**no lecture (Ascension Day)**	15.05.18 16.05.18	Exercise Sheet 3 Exercise 3-3 (.ipynb) Suggested Solutions Exercise 3-3 Solutions (.ipynb)
17.05.18 **8am-10am**	Lecture 5: Kernels	22.05.18 23.05.18	**no tutorials (Whit Tuesday)**
24.05.18	Guest Lecture by Dr. Denis Krompaß: Deep Learning	29.05.18 30.05.18	Exercise Sheet 4 Exercise 4-1 (.ipynb) Suggested Solutions Exercise 4-1 Solutions (.ipynb)
31.05.18	**no lecture (Corpus Christi)**	05.06.18 06.06.18	Exercise Sheet 5 Suggested Solutions
07.06.18	Lecture 6: Probability (Review), Frequentists and Bayesians	12.06.18 13.06.18	Exercise Sheet 6 Suggested Solutions
14.06.18	Guest Lecture by Dr. Florian Buettner: PCA	19.06.18 20.06.18	Exercise Sheet 7 body_sizes.txt Suggested Solutions
21.06.18	Lecture 7: Linear Classifiers, Support Vector Machine	26.06.18 27.06.18	Exercise Sheet 8 Suggested Solutions
28.06.18	Lecture 8: Model Comparison	03.07.18 04.07.18	Exercise Sheet 9 Suggested Solutions
05.07.18	Lecture 9: Bayes Nets	10.07.18 11.07.18	Exercise Sheet 10 Suggested Solutions