MASTER MEIM 2021-2022

Introduction to Artificial Intelligence

Lesson 1

Lesson given by prof. Francesco Camastra

Prof. Machine Learning at the M.Sc. Applied Computer Science of University Parthenope of Naples

Overview

- Presentation of Digital Ai Course
- Ancestors of Artificial Intelligence in History, Literature, and Science
- Birth of Artificial Intelligence (AI)
- Economic Perspectives of AI
- What is AI ?
- Machine Learning

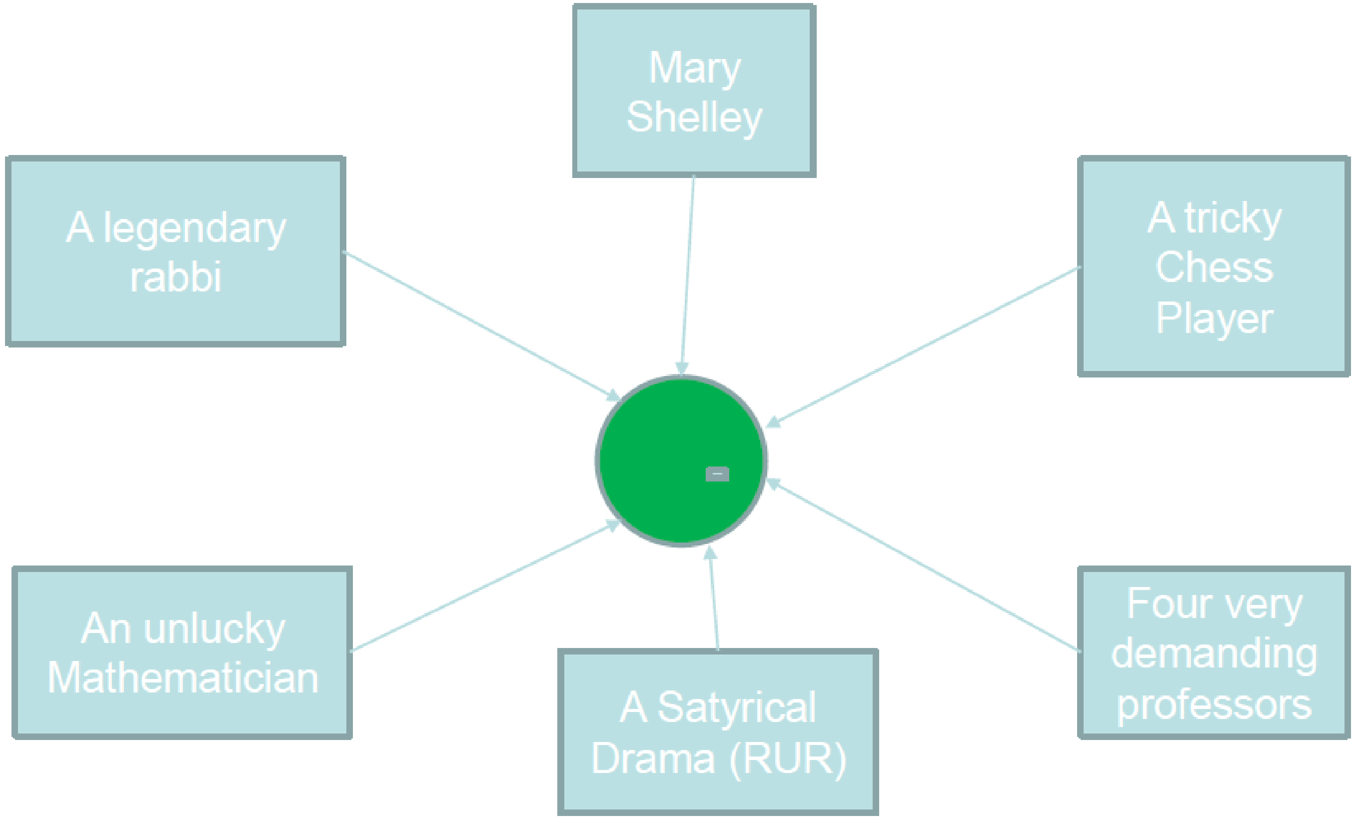
Digital AI Course

- June 28 9.00-13.00 Introduction to AI (Lesson 1) by Prof. Francesco Camastra
- June 29 14.00-18.00 Introduction to AI (Lesson 2) by Prof. Francesco Camastra
- July 6th 9.00-13.00 and 14.00-18.00, Python for AI, by Prof. Giuseppe Salvi
- July 14th 9.00-13.00 Supervised Learning (Lesson 1), by prof. Angelo Ciaramella; 14.00-18.00 Laboratory by Prof. Paola Barra
- July 15th 9.00-13.00 Supervised Learning (Lesson 2), by prof. Angelo Ciaramella; 14.00-18.00 Laboratory by Prof. Paola Barra
- July 21st 9.00-13.00 Unsupervised Learning (Lesson 1) by Prof. Antonino Staiano; 14.00-18.00 Laboratory by Prof. Alessio Ferone
- July 22nd 9.00-13.00 Unsupervised Learning (Lesson 2) by Prof. Antonino Staiano; 14.00-18.00 Laboratory by Prof. Alessio Ferone

Digital AI Course (cont.)

- September 5th 9.00-13.00 Preprocessing by Prof. Antonino Staiano; 14.00-18.00 Laboratory by Prof. Alessio Ferone
- September 6th 9.00-13.00 Deep Learning by Prof. Angelo Ciaramella; 14.00-18.00 Laboratory by Prof. Paola Barra

A Gallery for AI



Judah Loew (Rabbi Loew)

- Rabbi Loew is the subject of the legend about the creation of a **Golem**, a creature made out of clay to defend the Jews of the Prague Ghetto from antisemitic attacks, particularly the blood libel.
- He is said to have used mystical powers based on the esoteric knowledge of how God created.

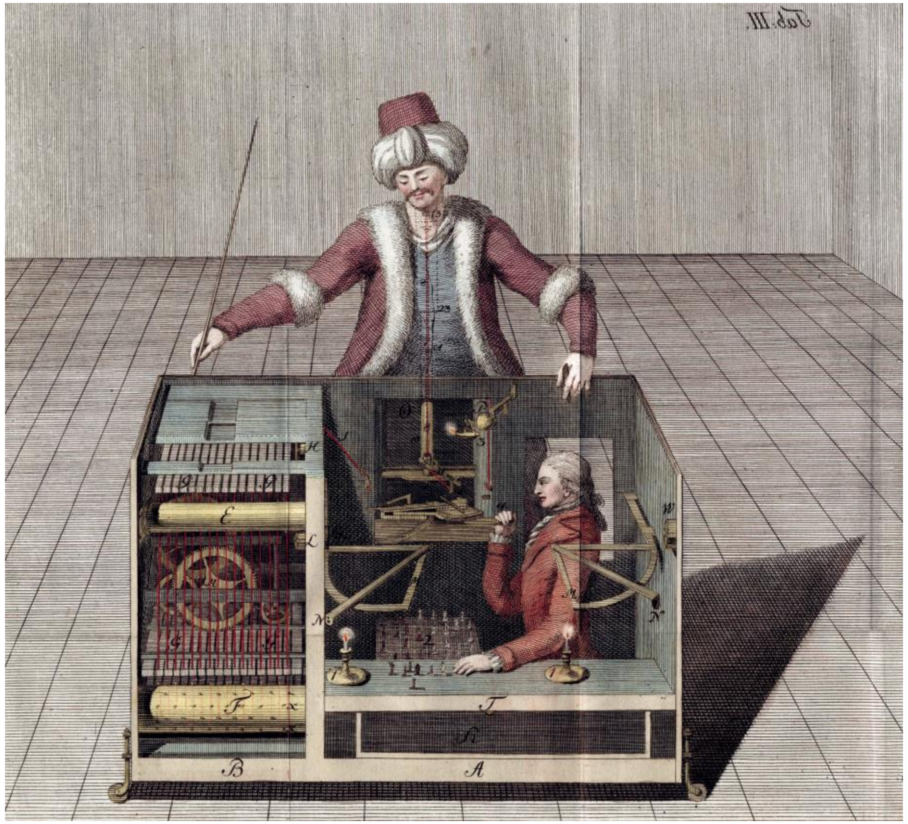
Golem

Wolfgang Von Kempelen (1734-1806)

- Wolfgang Von Kempelen was famous for his construction of **The Turk**, a chess playing automaton presented to Maria Theresa of Austria in 1769.
- The Turk consisted of a life-sized model of a human head and torso, dressed in Turkish robes and a turban, seated behind a large cabinet on top of which a chessboard was placed.

The Turk

- The Turk appeared to be able to play a strong game of chess against a human opponent, but was in fact merely an elaborate simulation of mechanical automation.
- A human chess master (a dwarf or a man without legs) concealed inside the cabinet puppeteered the Turk from below by a series of levers.



The Turk (cont.)

- The Turk won most of the games played during its demonstrations around Europe and the Americas for nearly 84 years, playing and defeating many challengers including statesmen such as Napoleon Bonaparte and Benjamin Franklin.
- The Turk survived at Von Kempelen's death and was used by other cheaters until 1853 when Edgar Allan Poe, the American writer of thrillers, discovered the trick.

Frankenstein

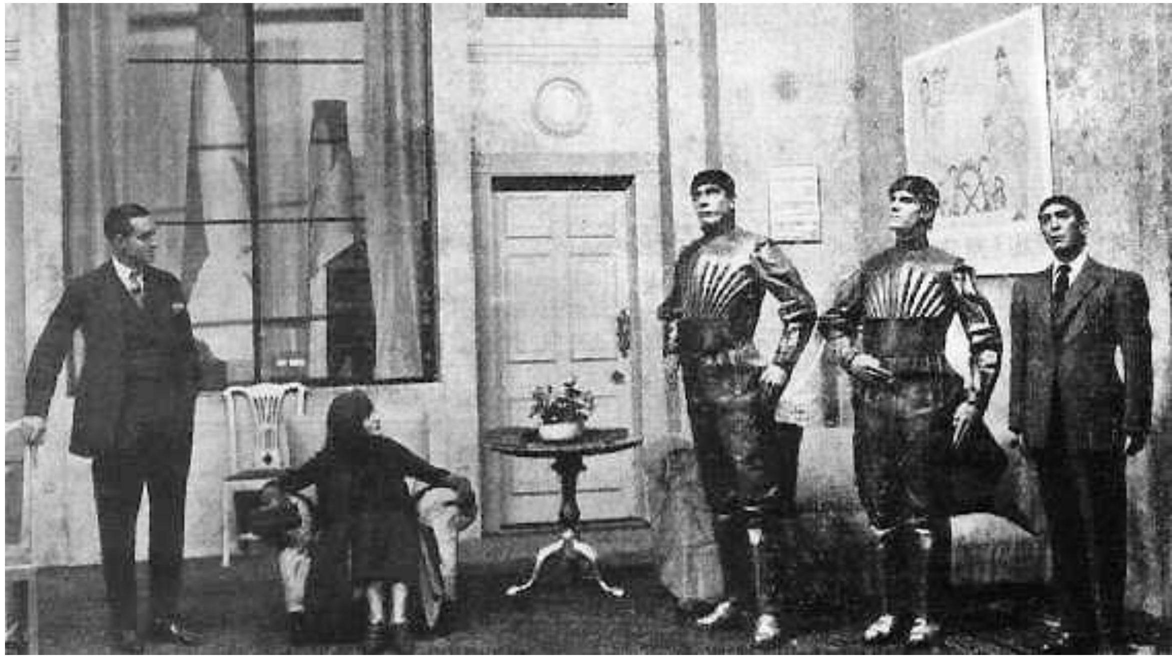
- The novel **Frankenstein or The Modern Prometheus** was written by English author Mary Shelley (1797–1851).
- The novel tells the story of Victor Frankenstein, a young scientist who creates a hideous, sapient creature in an unorthodox scientific experiment.

Frankestein (cont.)

- Since the novel's publication, the name "Frankenstein" has often been used to refer to the monster itself. This usage is considered erroneous.
- In the novel, Frankenstein's creation is identified by words such as "creature", "monster", "daemon", "wretch", "abortion", "fiend" and "it".

R.U.R.

- R.U.R. is a science fiction play written by the Czech writer Karel Čapek in 1920.
- R.U.R. stands for Rossumovi Univerzální Roboti (**Rossum's Universal Robots**).
- It premiered on 25 January 1921 and **introduced the word "robot" to the English language**.



R.U.R : the plot

- The play begins in a factory that makes artificial people, called ***roboti*** (**robots**), from synthetic organic matter.
- They are **not exactly robots** by the current definition of the term: **they are living flesh and blood creatures rather than machinery** and are **closer to the modern idea of androids** or replicants.
- They may be mistaken for humans and can think for themselves. **They seem happy to work for humans at first, but a robot rebellion leads to the extinction of the human race**.

Isaac Asimov (1920-1992)

- Isaac Asimov was a russian (and after US) scientist and one of the the science fiction writers.
- He published in 1942 a short story Runaround" (included in the 1950 collection I, Robot), where he introduced **The Three Laws of Robotics**.

The Three Laws of Robotics (The Handbook of Robotics, 56th Edition, 2058 A.D.)

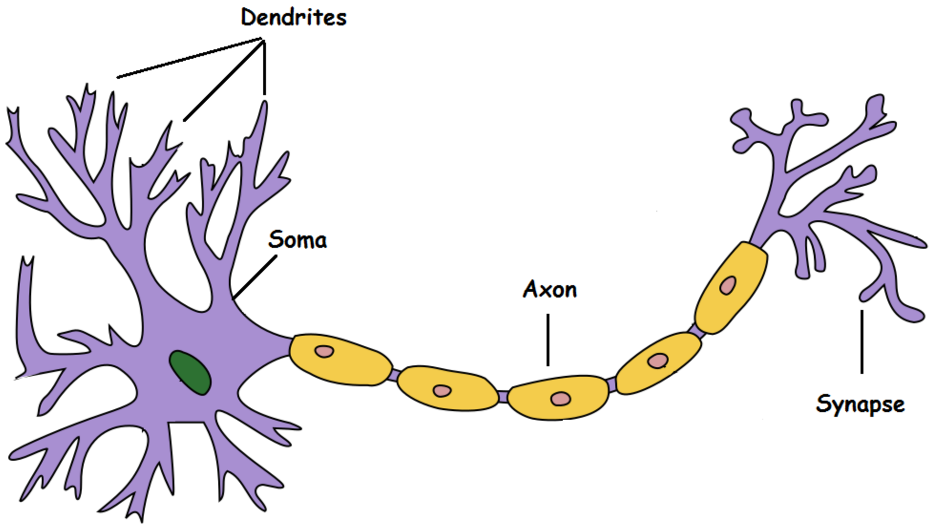
1. First Law: **A robot may not injure a human being or, through inaction, allow a human being to come to harm.**
2. Second Law: **A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.**
3. Third Law: **A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.**

Relevance of The Three Laws of Robotics

- According to most of scientists the Three laws of Robotics should rule the behavior of all AI Systems (e.g., automatic driving systems without human assistance).

Warren Mc Culloch and Walter Pitts

- Warren MuCulloch (neuroscientist) and Walter Pitts (logician) proposed the first Computational Model of the neuron in 1943. This model aimed to mimic the biological neuron of the human brain.
- The human brain contains about 100 billions of neurons.

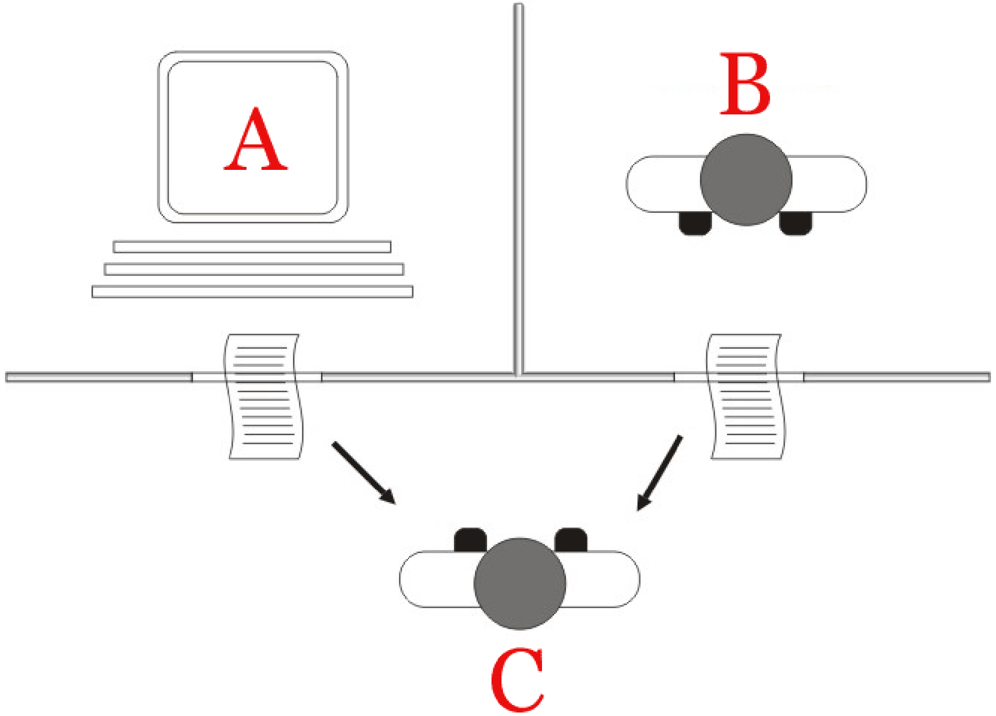


Alan Turing

- Alan Turing in 1950 developed the so-called the **Turing test**.
- The Turing test is a **test of a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human**.

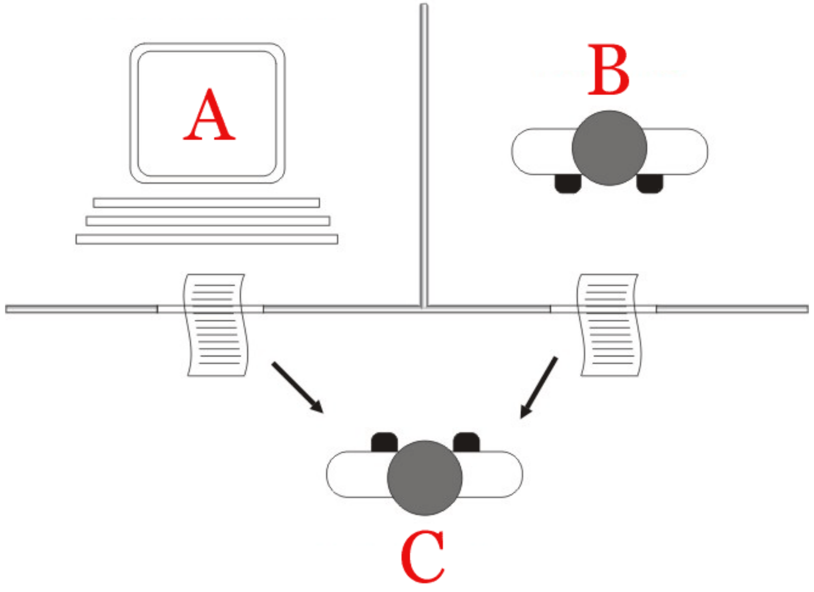
The Turing test

- Turing proposed that a **human evaluator** (**C**) **would judge natural language conversations between a human** (**B**) **and a machine** (**A**) **designed to generate human-like responses**.



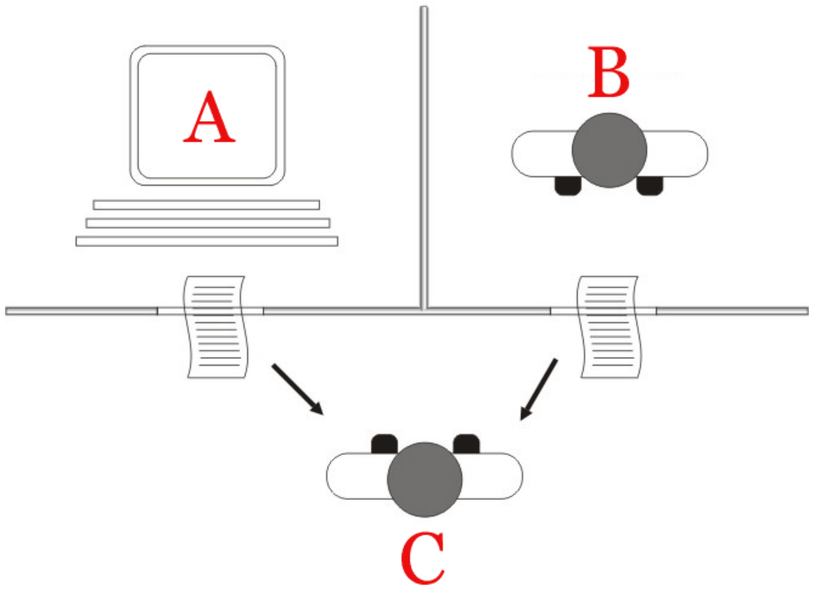
Turing Test (cont.)

- The evaluator would be aware that one of the two partners in conversation is a machine, and all participants would be separated from one another. The conversation would be limited to a text-only channel such as a computer keyboard and screen so the result would not depend on the machine's ability to render words as speech.
- If **the evaluator cannot reliably discriminate the machine from the human, the machine is said to have passed the test**.



Turing Test (cont.)

- The test results do not depend on the machine's ability to give correct answers to questions, only how closely its answers resemble those a human would give.



Have Current computers passed the Turing Test ?

- According to Tomaso Poggio, in some restricted task computer passed the Turing test.

Deep Blue

- Deep Blue was a **chess-playing expert system** run on a unique purpose-built IBM supercomputer.
- It was the **first computer to win a game, and the first to win a match, against a reigning world champion** under regular time controls.
- Its development began in 1985 at Carnegie Mellon University under the name ChipTest.

Deep Blue (cont.)

- It then moved to IBM, where it was first renamed Deep Thought, then again in 1989 to **Deep Blue**.
- It first played world champion Garry Kasparov in a six-game match in 1996, where it lost four games to two. In 1997 it was upgraded and, in a six-game re-match, it defeated Kasparov by winning three games and drawing one. **Deep Blue's victory is a milestone in the history of artificial intelligence**.

Deep Blue: History

- While a doctoral student at Carnegie Mellon University, **Feng-hsiung Hsu** began development of a chess-playing supercomputer ChipTest.
- After receiving his doctorate in 1989, Hsu and **Murray Campbell** joined IBM Research to continue their project. Then, **Arthur Joseph Hoane**, **Jerry Brody** joined the team in 1990.

Deep Blue: Hardware

- Deep Blue in 1997 was implemented on IBM RS/6000 SP Supercomputer with 30 PowerPC 604e "High 2" 200 MHz CPUs and 480 custom VLSI second-generation "chess chips“.
- Deep Blue occupied 2 cabinets and his speed was about 11.38 GigaFlops.
- A current PC with a single core of 2.5 GHz has a theoretical speed of 10 GigaFlops.

The birth of Artificial Intelligence

- The birth of Artificial Intelligence, as discipline, can be dated to **Dartmouth Summer Research Project as Artificial Intelligence**, proposed by **John McCarthy**, **Marvin L. Minsky**, **Nathaniel Rochester**, and **Claude E. Shannon** in 1956.
- Although **Herbert Simon introduced the term Artificial Intelligence for the first time**.

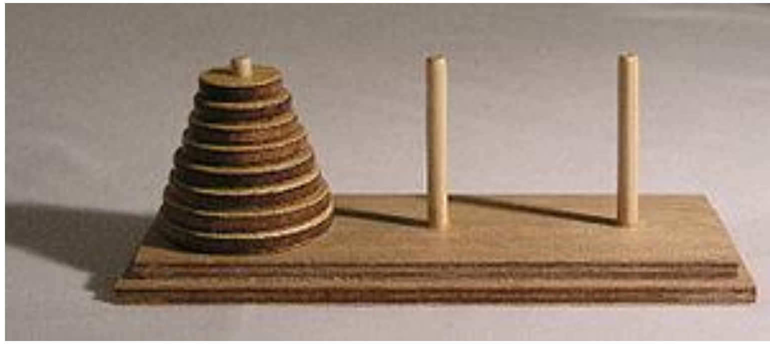
Herbert A. Simon (1916-2001)

- Herbert Alexander Simon was an American political scientist, with a Ph.D. in political science, whose work also influenced the fields of computer science, economics, and cognitive psychology. He was professor at Carnegie Mellon University from 1949 to 2001.
- He received the **Nobel Prize in Economic Sciences in 1978** and the **Turing Award in computer science in 1975**. Simon developed with Allen Newell the Logic Theory Machine (1956) and the General Problem Solver (GPS) (1957) programs. GPS may possibly be the first method developed for separating problem solving strategy from information about particular problems.

General Problem Solver (GPS) by Herbert A. Simon

- GPS was the first AI Program.
- GPS was also the first computer program which separated its knowledge of problems (rules represented as input data) from its strategy of how to solve problems (a generic solver engine).
- While GPS solved simple problems such as the **Towers of Hanoi** that could be sufficiently formalized, it could not solve any real-world problems due to computational complexity of the algorithms used in GPS.

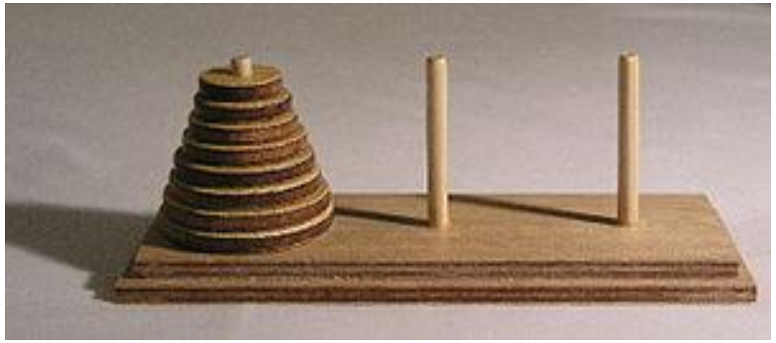
The Tower of Hanoi Problem



- The Tower of is a mathematical game or puzzle consisting of three rods and a number of disks of various diameters, which can slide onto any rod. The puzzle begins with the disks stacked on one rod in order of decreasing size, the smallest at the top, thus approximating a conical shape. The objective of the puzzle is to move the entire stack to the last rod, obeying the following rules:
- Only one disk may be moved at a time.
- Each move consists of taking the upper disk from one of the stacks and placing it on top of another stack or on an empty rod.
- No disk may be placed on top of a disk that is smaller than it.

The Tower of Hanoi Problem (cont.)

- With 3 disks, the puzzle can be solved in 7 moves. The minimal number of moves required to solve a Tower of Hanoi puzzle is $2^n - 1$, where n is the number of disks.
- The Problem was proposed by Edouard Lucas, a French mathematician



The birth of Artificial Intelligence

- The birth of Artificial Intelligence, as discipline, can be dated to **Dartmouth Summer Research Project as Artificial Intelligence**, proposed by **John McCarthy**, **Marvin L. Minsky**, **Nathaniel Rochester**, and **Claude E. Shannon** in 1956.
- Although **Herbert Simon introduced the term Artificial Intelligence for the first time**.

A proposal for the Dartmouth Summer Research Project on Artificial Intelligence

- We (John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon) propose that a **2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College** in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.

A proposal for the Dartmouth Summer Research Project on Artificial Intelligence (cont.)

- An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.
- The following are some aspects of the artificial intelligence problem:

A proposal for the Dartmouth Summer Research Project on Artificial Intelligence (cont.)

- **Automatic Computers**
  - If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

A proposal for the Dartmouth Summer Research Project on Artificial Intelligence (cont.)

- **How Can a Computer be Programmed to Use a Language**
  - It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture. From this point of view, forming a generalization consists of admitting a new word and some rules whereby sentences containing it imply and are implied by others. This idea has never been very precisely formulated nor have examples been worked out.
- **Neuron Nets**
  - How can a set of (hypothetical) neurons be arranged so as to form concepts. Considerable theoretical and experimental work has been done on this problem by Uttley, Rashevsky and his group, Farley and Clark, Pitts and McCulloch, Minsky, Rochester and Holland, and others. Partial results have been obtained but the problem needs more theoretical work.

A proposal for the Dartmouth Summer Research Project on Artificial Intelligence (cont.)

- **Theory of the Size of a Calculation**
  - If we are given a well-defined problem (one for which it is possible to test mechanically whether or not a proposed answer is a valid answer) one way of solving it is to try all possible answers in order. This method is inefficient, and to exclude it one must have some criterion for efficiency of calculation. Some consideration will show that to get a measure of the efficiency of a calculation it is necessary to have on hand a method of measuring the complexity of calculating devices which in turn can be done if one has a theory of the complexity of functions. Some partial results on this problem have been obtained by Shannon, and also by McCarthy.

A proposal for the Dartmouth Summer Research Project on Artificial Intelligence (cont.)

- **Self-Improvement**
  - Probably a truly intelligent machine will carry out activities which may best be described as self-improvement. Some schemes for doing this have been proposed and are worth further study. It seems likely that this question can be studied abstractly as well.
- **Abstractions**
  - A number of types of "abstraction" can be distinctly defined and several others less distinctly. A direct attempt to classify these and to describe machine methods of forming abstractions from sensory and other data would seem worthwhile.

A proposal for the Dartmouth Summer Research Project on Artificial Intelligence (cont.)

- **Randomness and Creativity**
  - A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of a some randomness. The randomness must be guided by intuition to be efficient. In other words, the educated guess or the hunch include controlled randomness in otherwise orderly thinking.

Partecipants

- Dr. Marvin Minsky
- Dr. Julian Bigelow
- Prof. D.M. Mackay
- Mr. Ray Solomonoff
- Mr. John Holland
- Mr. John McCarthy
- (Trenchard More)
- (W. Ross Ashby)
- (Abraham Robinson)
- (Tom Etter)
- Dr. Claude Shannon (4weeks)
- Mr. Nathanial Rochester (4weeks)
- Mr. Oliver Selfridge (4weeks)
- Mr. Allen Newell (2weeks)
- Prof. Herbert Simon (2weeks)
- (John Nash)
- (W.S. McCulloch)
- (Arthur Samuel)
- (Kenneth R. Shoulders)
- (Alex Bernstein)
- (David Sayre)

Artificial Intelligence Today

- But what is the economic impact of Artificial Intelligence today ?
- Considering the recent analysis of FortuneBusinessInsight https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114

Artificial Intelligence Market size (FortuneBusinessinsight)

- The global Artificial Intelligence market is projected to grow from $387.45 billion in 2022 to $1,394.30 billion by 2029, at a **CAGR of 20.1% in forecast period**.
- The global COVID-19 pandemic has been unprecedent and staggering with this technology experiencing higher-than anticipated demand across all regions compared to pre-pandemic level.
- Based on FortuneBusinessinsight analysis **the global market of AI had exhibited a rise of 120% in 2020 compared to 2019**.

COVID-19 impact on Artificial Intelligence (FortuneBusinessinsight)

- To cope with the problems of the COVID-19 pandemy, the demand of AI in healthcare increased notably.
- AI provides several tools and models that enhance traditional analytics and decision-making capabilities. This improves the accuracy and efficiency of diagnosis, treatments, and forecasting.

Italian Market of AI (Artificial Intelligence Observatory of Politecnico di Milano)

- Artificial Intelligence Observatory of Politecnico di Milano in his annual book, estimates that the **annual growth of AI market in 2021 was +27% achieving the amount of 380 millions of Euro**. This value is doubled w.r.t the one of two years ago.
- Most of market demand was in Italy (290 millions of Euro corresponding to 76% of the whole market), only a residual part of AI projects was commissioned by non-Italian companies/administrations.

AI projects in Italy (Artificial Intelligence Observatory of Politecnico di Milano)

- AI projects that attract more interest are:
- 35% of whole amount: **INTELLIGENT DATA PROCESSING**, i.e., algorithms for analyzing and extracting informations from data (+32% w.r.t. 2020);
- 16% of whole amount: **NATURAL LANGUAGE PROCESSING**, i.e., systems for the natural language interpretation (+20% w.r.t. 2020);
- 16% of whole amount: **RECOMMANDATION SYSTEMS**, i.e., systems for suggesting to customers contents (or products) on the basis of their preferences (+20% w.r.t. 2020);

AI projects in Italy (cont.)

- 11% of whole amount: **INTELLIGENT ROBOT PROCESS AUTOMATION**, i.e., systems where AI algorithms machine a few project activities algorithms managing, at the same time, the all stages of project (+15% w.r.t. 2020);
- 11% of whole amount: **COMPUTER VISION**, i.e., systems that analyze images for surveillance of places or for monitoring production lines (+41% w.r.t. 2020);
- 10.5% of whole amount: **CHATBOT AND VIRTUAL ASSISTANT** (+34% w.r.t. 2020);

Italian Market of AI

- The domains in Italy that require most investments in AI are: finance (banks), utilities (energy), telecommunication/multimedia, insurances.
- AI project regard mainly big companies, only 6% Small and Medium Companies (PMI) started an AI project in 2021.
- Italian Observatory of AI identified 94 cases where the AI usage implied ethical problems. Main problems are: Freedom violation (19%), Potential Creation of Trusting by Big Tech (17%), privacy (11%), potential bias in the AI system design (23%).

But …. What is Artificial Intelligence ?

- It does **not exist a unambiguous definition of AI** .
- Therefore all previous estimation about the AI market depends on what it is considered AI.

High-level expert group on Artificial Intelligence of European Commission

- In June 2018 European Commission (EC) appointed a High-level Expert Group for suggesting EC advices on its Artificial Intelligence strategy.
- High-Level Expert group (**HLEG on AI)** is composed of 52 experts of 28 different EU countries, coming from academia, industry, and civil society.
- Member of HLEG on AI is Francesca Rossi (Distinguished Staff Member of IBM Research)

AI HLEG deliverables

- AI HLEG produced the following deliverables:
  - **Ethics Guidelines for Trustworthy AI**
  - **Policy and Investment Recommendations for Trustworthy AI**
  - **The final Assessment List for Trustworthy AI**
  - **Sectorial Considerations on the Policy and Investment Recommendations**

Ethics Guidelines for Trustworthy AI deliverables

- On 8 April 2019, the High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence. According to the Guidelines, **trustworthy AI** should be:
  - **lawful** - respecting all applicable laws and regulations
  - **ethical** - respecting ethical principles and values
  - **robust** - both from a technical perspective while taking into account its social environment

Trustworthy AI requirements

- AI HLG put forward a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy. A specific assessment list aims to help verify the application of each of the key requirements:
  - **Human agency and oversight**: AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms need to be ensured, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches

Trustworthy AI requirements (cont.)

- **Technical Robustness and safety**: AI systems need to be resilient and secure. They need to be safe, ensuring a fall back plan in case something goes wrong, as well as being accurate, reliable and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.
- **Privacy and data governance**: besides ensuring full respect for privacy and data protection, adequate data governance mechanisms must also be ensured, taking into account the quality and integrity of the data, and ensuring legitimised access to data.

Trustworthy AI requirements (cont.)

- **Transparency**: the data, system and AI business models should be transparent. Traceability mechanisms can help achieving this. Moreover, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. Humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations.

Trustworthy AI requirements (cont.)

- **Diversity, non-discrimination and fairness**: Unfair bias must be avoided, as it could could have multiple negative implications, from the marginalization of vulnerable groups, to the exacerbation of prejudice and discrimination. Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.

Trustworthy AI requirements (cont.)

- **Societal and environmental well-being**: AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should take into account the environment, including other living beings, and their social and societal impact should be carefully considered.
  Accountability: Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. Moreover, adequate an accessible redress should be ensured.

Artificial Intelligence definition according to HLEG on AI

- According to HLEG on AI, Artificial Intelligence techniques can be divided in three groups:
  - **Reasoning and Decision Making**
  - **Learning**
  - **Robotics**

Reasoning and Decision Making

- This group of techniques includes knowledge representation and reasoning, planning, scheduling, search, and optimization. These techniques allow to perform the reasoning on the da ta coming from the sensors. To be able to do this, one needs to transform data to knowledge, so one area of AI has to do with how best to model such knowledge (**knowledge representation**). Once knowledge has been modelled, the next step is to reason with it (**knowledge reasoning**), which includes making inferences through symbolic rules, planning and scheduling activities, searching through a large solution set, and optimizing among all possible solutions to a problem. The final step is to decide what action to take. The reasoning/decision making part of an AI system is usually very complex and requires a combination of several of the above mentioned techniques.

Reasoning and Decision Making: Expert Systems

- In this category there are the so-called **Expert Systems** .
- In the Expert System the knowledge of an Expert (a chess-player, a doctor, …) on a given domain is formalized in terms of a set of rules, typically IF-THEN rules.
- The set of rules is called **KNOWLEDGE BASE** of the system.

An example of Expert System: Deep Blue

- Deep Blue was a chess-player Expert System.
- The rules were set up using the knowledge of several chess grandmasters (Miguel Illescas, John Federowicz, and Nick De Firmian). Then, each single rule was fine-tuned by the chess grandmaster Joel Benjamin.

The first example of Expert System: Mycin

- MYCIN was the first example of Expert Systems in AI.
- It was developed by Edward Shortliffe, as PhD thesis under the supervision of Bruce Buchanan, over five or six years in the early 1970s at Stanford University.
- MYCIN was able **to identify bacteria causing severe infections, such as bacteremia and meningitis**, and to recommend antibiotics, with the dosage adjusted for patient's body weight.
- The Knowledge Base of MYCIN was composed of about 600 rules

An example of TERA rules

- TERA is an expert system, developed by a joint group of University Parthenope and ISPRA researchers in 2015, able to assess the environmental risk associated to the cultivation of genetically modified plants.
- The Knowledge Base was composed of 6215 IF-THEN rules and expressed the knowledge of ISPRA experts on the domain (i.e., environmental risk assessment).

F. Camastra, **A. Ciaramella**, V. Giovannelli, M. Lener, V. Rastelli, **A. Staiano**, G. Staiano, “A fuzzy decision system for genetically modified plant environmental risk assessment using Mamdani inference”, Expert Systems with Application, 42(3), 1710-1716

Example of Expert System: TERA of University Parthenope of Naples

IF number of insert copies is *Low* AND
number of introduced sequences is *High* THEN
potential risk of the insert is *High*
ELSE
IF number of insert copies is *High* AND
number of introduced sequences is *Medium* THEN
potential risk of the insert is *High*
ELSE
IF number of insert copies is *Low* AND
number of introduced sequences is *Low* THEN
potential risk of the insert is Low

Expert Systems: today

- Expert Systems are still used in the applications where the users want to know how the system achieves a given decision.
- Therefore expert systems are still popular in diagnosis applications, in the risk assessment (e.g., to assess the bank risk re to give a loan to a customer), in the healthcare.

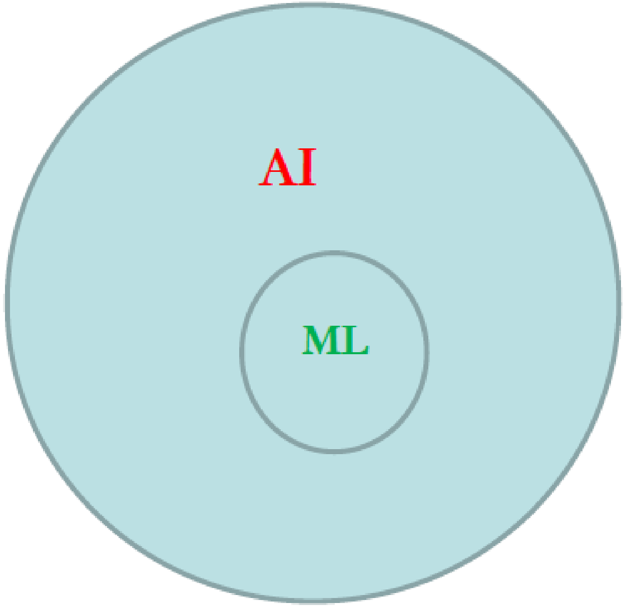
Learning

- The ability to learn is one of the **distinctive attributes of intelligent behavior**. Following Carbonell et al, we can say that "*Learning process includes the acquisition of new declarative knowledge, the development of motor and cognitive skills through instruction or practice, the organization of new knowledge into general, effective representations, and the discovery of new facts and theories through observation and experimentation*".

Machine Learning

- The **study and computer modeling of learning process in their multiple manifestations constitutes the topic of *Machine learning***.

Artificial Intelligence vs. Machine Learning



Machine Learning Market size (FortuneBusinessinsight)

- The global world Machine Learning market was 15.44 billion in 2021 and is projected to grow from $21.17 billion in 2021 to $209.91 billion by 2029, at a **CAGR of 38.8% in forecast period**.
- The global world Artificial Intelligence market is projected to grow from $387.45 billion in 2022 to $1,394.30 billion by 2029, at a **CAGR of 20.1% in forecast period**.

- https://www.globenewswire.com/news-release/2022/04/04/2415724/0/en/Machine-Learning-Market-Size-2022-2029-Worth-USD-209-91-Billion-Exhibiting-a-CAGR-of-38-8.html

Growth Drivers of Machine Learning Market

- Increasing Demand in Healthcare Sector to Boost Market
- Rising Demand for Data Analysis to Propel Market Growth

https://www.globenewswire.com/news-release/2022/04/04/2415724/0/en/Machine-Learning-Market-Size-2022-2029-Worth-USD-209-91-Billion-Exhibiting-a-CAGR-of-38-8.html

Machine Learning market in Europe

- The Germany market dominated the Europe Machine learning as a Service Market by Country in 2021, and would continue to be a dominant market till 2028; thereby, achieving a market value of $2,375.7 million by 2028. The UK market is estimated to grow at a CAGR of 30% during (2022 - 2028). Additionally, The France market would witness a CAGR of 32% during (2022 - 2028).

- There are no news about Machine Learning market in Europe since Machine Learning market is included in Artificial Intelligence' one.

But …

- If you go on Linkedin you can find 3917 job advertisements (29 new) in Italy with machine learning as a topic.
- https://www.linkedin.com/jobs/machine-learning-offerte-di-lavoro/?originalSubdomain=it

Machine Learning (by Michalski's book)

- Machine learning has been developed around the following primary research lines:
  - ***Task-Oriented Studies***, that is the development of learning systems to improve performance in a predetermined set of tasks.
  - ***Cognitive Simulation***, namely the investigation and computer simulation of human learning processes.
  - ***Theoretical Analysis***, i.e., the theoretical investigation of possible learning methods and algorithms independently of application domain.

Machine Learning Systems

- Although machine learning systems can be classified according to different view points, a common choice is to classify machine learning systems on the basis of the underlying learning strategies used.

The Teacher and The Learner

- In machine learning two entities, the teacher and the learner, play a crucial role.
- The teacher is the entity that has the required knowledge to perform a given task.
- The learner is the entity that has to learn the knowledge to perform the task.
- We can distinguish learning strategies by the amount of inference the learner performs on the information provided by the teacher.

First Case

- If a computer system (the learner) is programmed directly, its knowledge increases but it performs no inference since all cognitive efforts are developed by the programmer (the teacher).

Second Case

- On the other hand, if a system independently discovers new theories or invents new concepts, it must perform a very substantial amount of inference; it is deriving organized knowledge from experiments and observations.

Intermediate Case

- An intermediate case it could be a student determining how to solve a math problem by analogy to problem solutions contained in a textbook. This process requires inference but much less than discovering a new theorem in mathematics.
- Increasing the amount of inference that the learner is capable of performing, the burden on the teacher decreases.

A Taxonomy of Machine Learning

- This taxonomy tries to capture the notion of trade-off in the amount of effort required by both the learner and the teacher. We can identify four different learning types:
  - **Rote Learning**
  - **Learning from instruction**
  - **Learning by analogy**
  - **Learning from examples**

Rote Learning

- Rote learning consists in the direct implanting of new knowledge in the learner. No inference or other transformation of the knowledge is required on the part of the learner.

Rote Learning (cont.)

- Variants of this method include:
  - **Learning by being programmed or modified by an external identity.** It requires no effort on the part of the learner. For instance, the usual style of computer programming.
  - **Learning by memorization of given facts and data with no inferences drawn from the incoming information.** For instance, the primitive database systems.

Learning from Instruction

- Learning from instruction (or *Learning by being told*) consists in **acquiring knowledge from a teacher or other organized source**, such as a textbook, requiring that the **learner transform the knowledge from the input language to an internal representation**.

Learning from Instruction (cont.)

- The new information is integrated with prior knowledge for effective use. The learner is required to perform some inference, but a large fraction of the cognitive burden remains with the teacher, who must present and organize knowledge in a way that incrementally increases the learner's actual knowledge. **Learning from instruction mimics education methods**.
- Therefore, the machine learning task is to build a system that can accept instruction and can store and apply this learned knowledge effectively.

Learning by Analogy

- Learning by analogy consists in **acquiring new facts or skills by transforming and increasing existing knowledge that bears strong similarity to the desired new concept** or skill into a form effectively useful in the new situation.
- A Learning-by-analogy system might be applied to convert an existing computer program into one that performs a closely-related function for which it was not originally designed.

Learning by Analogy (cont.)

- Learning by analogy **requires more inference on the part of the learner that does rote learning or learning from instruction**.
- A fact or skill analogous in relevant parameters must be retrieved from memory; then the retrieved knowledge must be transformed, applied to the new situation, and stored for future use.

Learning from Examples

- Given a set of examples of a concept, **the learner induces a general concept description that describe the examples**. The amount of inference performed by the learner is much greater than in learning from instruction and in learning by analogy. **Learning from examples** has become so popular in the last years that it **is** often **called simply *learning***.
- In a similar way, **the learner and examples** are respectively referred as ***learning machine*** and ***data***.

Learning Problem

- The ***learning problem*** can be described as finding a general rule that explains data given only a sample of limited size.
- The difficulty of this task is similar to the problem of children learning to speak from the sounds emitted by the grown-up people.
- The **learning problem** can be stated as follows: **Given an example sample of limited size, *to find a concise data description***.

Learning techniques

- Learning techniques can be grouped in four big families:
  - **supervised learning**
  - **reinforcement learning**
  - **unsupervised learning**
  - **semi-supervised learning**

Supervised Learning

- In supervised learning (or ***learning with a teacher***), the data is a sample of input-output patterns. In this case, **a concise description of the data is the *function* that can yield the output, given the input**.
- This problem is called **supervised learning** because **the objects under considerations are already associated with target values**, e.g., classes and real values.

Supervised Learning (cont.)

- Examples of this learning task are the recognition of handwritten letters and digits, the prediction of stock market indexes.
- In the problem of supervised learning, given a sample of input-output pairs, called the ***training sample*** (or ***training set***), the task is to find a deterministic function that maps any input to an output that can predict future input-output observations, minimizing the errors as much as possible.

Supervised Learning (cont.)

- Examples of this learning task are the recognition of handwritten letters and digits, the prediction of stock market indexes.
- In the problem of supervised learning, given a sample of input-output pairs, called the ***training sample*** (or ***training set***), the task is to find a deterministic function that maps any input to an output that can predict future input-output observations, minimizing the errors as much as possible.

Supervised Learning (cont.)

- Whenever asked for the target value of an object present in the training sample, it can return the value that appeared the highest number of times together with this object in the training sample.
- According to the type of the outputs, supervised learning can be distinguished in:
  - *classification learning*.
  - *regression learning*.

Classification

- If the output space has no structure except whether two elements of the output are equal or not, this is called the problem of ***classification learning*** (or simply ***classification***).
- Each element of the output space is called a ***class***. The learning algorithm that solves the classification problem is called the ***classifier***. In classification problems the task is to assign new inputs to one of a number of discrete classes or categories.
- This problem characterizes most pattern recognition tasks. A typical classification problem is to assign to a character bitmap the correct letter of the alphabet.

Training Set of Animals

DOG

CAT

CAT

RABBIT

TIGER

HORSE

SHEEP

GOAT

Training Set Labeling

- If the Labeling cannot do in automatic way, it must be an expensive procedure in terms of human effort, i.e., human must associate manually the correct label to the pattern (i.e., the image)
- At the same time, training labeling is a crucial phase in the supervision, since misslabeled pattern (i.e., a wrong label associated to a pattern), can affect notably the learning quality in the learner.
- **MISSLABELING**

CAT

Training Set Representativeness

- Training Set must be representative of all possible shapes that a given class may assume.
- In order to explain that, we consider an island composed of blond men and women. Suddenly, a boat arrives at the isle, with a brune woman.
- All island inhabitants think that the brune woman is not a woman since they saw in their life only women that are blond.

Regression

- If the outputs space is formed by the outputs representing the values of continuous variables, for instance the prediction of a stock exchange index at some future time, then the learning task is known as the problem of ***regression*** or ***function learning***.
- Typical examples of Regression are to predict the value of shares in the stock exchange market, to predict the electricity consumption, and to estimate the value of a physical measure (e.g., pression, temperature) in an environmental station.

Examples of Supervised Learning Algorithms

- Neural Networks, i.e., MLP
- Deep Neural Networks, i.e., Convolutional Neural Networks (CNN)
- Support Vector Machines (SVM)
- Decision Trees, i.e., Random Forests

Reinforcement Learning

- Reinforcement learning has its roots in control theory. It considers the scenario of a dynamic environment that results in state-action-reward triples as the data.
- The difference between reinforcement and supervised learning is that **in reinforcement learning no optimal action exists in a given state**, but **the learning algorithm must identify an action in order to maximize the expected reward over time**.

Reinforcement Learning (cont.)

- The concise description of data is the strategy that maximizes the reward.
- The problem of reinforcement learning is to learn what to do, i.e., how to map situations to actions, in order to maximize a given reward.
- Unlike supervised learning task, **the learning algorithm is not told which actions to take in a given situation**.

Reinforcement Learning (cont.)

- Instead, **the learner is assumed to gain information about the actions taken by some reward not necessarily arriving immediately after the action is taken**.
- An example of such a problem is learning to play chess. Each board configuration, namely the position of chess pieces on the chess board, is a given state; the actions are the possible moves in a given configuration.

Reinforcement Learning (cont.)

- The **reward for a given action** (e.g., the move of a piece), is winning the game. On the contrary, **the punishment is losing the game**.
- This reward, or this punishment, is delayed which is very typical for reinforcement learning. Since **a given state has no optimal action, one of the biggest challenges of a reinforcement learning algorithm is to find a trade-off between exploration and exploitation**.

Reinforcement Learning (cont.)

- In order to maximize reward (or minimize the punishment) a learning algorithm must choose actions which have been tried out in the past and found to be effective in producing reward, that is it must exploit its current knowledge.
- On the other hand, to discover those actions the learning algorithm has to choose actions not tried in the past and thus explore the state space.
- There is no general solution to this dilemma, but that neither of the two options can lead exclusively to an optimal strategy is clear.

Unsupervised Learning

- If the data is only a sample of objects without associated target values, the problem is known as ***unsupervised learning***.
- In **unsupervised learning there is no teacher**. Hence a concise description of the data could be a set of clusters or a probability density stating how likely it is to observe a certain object in the future.

Unsupervised Learning (cont.)

- Typical examples of unsupervised learning tasks include the problem of image and text segmentation and the task of novelty detection in process control.
- In unsupervised learning we are given a training sample of objects (e.g., images) with the aim of extracting some *structure* from them.
- For instance, identifying indoor or outdoor images or extracting face pixels in an image.

Unsupervised Learning (cont.)

- If some structure exists in training data, it can take advantage of the redundancy and find a short description of data. A general way to represent data is to specify a similarity between any pairs of objects. If two objects share much structure, it should be possible to reproduce the data from the same prototype. This idea underlies ***clustering algorithms*** that form a rich subclass of unsupervised algorithms.

Clustering algorithms

- Clustering algorithms are based on the following idea. **Given a fixed number of clusters, we aim to find a grouping of the objects such that similar objects belong to the same cluster.**
- If it is possible to find a clustering such that the similarities of the objects in one cluster are much greater than the similarities among objects from different clusters, we have extracted structure from the training sample so that the whole cluster can be represented by one representative data

Bayesian Theory of Decision (BTD)

- *Bayesian theory of decision* is a fundamental tool of analysis in Machine Learning. Several machine learning algorithms have been derived using BTD.
- The **fundamental idea in BTD** is that the **decision problem can be solved using probabilistic considerations**.

An Example

- We suppose to have a classroom in which there are students of both genders. There is an examiner, outside the classroom, that has to call the students for the examination. He has a list of the surnames of the students, but the surnames are not accompanied by the first names.

An Example (cont.)

- How can the examiner decide if to a given surname corresponds a girl or a boy?

BTD

- The aim of this lesson is to answer this question by introducing BTD. We will show that **BTD is a *formalization of the common sense***. This lesson is inspired by Duda and Hart's book, that represents a milestone in the history of pattern recognition and machine learning.

Bayesian Decision Rule

- We consider again our classroom with boys and girls and the examiner that has only a list with the surnames of the students. When the examiner calls a student (e.g., Esposito) and the student appears, in decision-theoretic terminology we say that **the student replies the *nature* in one of two possible states**, i.e., **either the student is a boy or the student is a girl**.
- We identify the ***state of nature*** (or ***class***) with C. If the student is a girl $C = C_1$, otherwise $C = C_2$.

Bayesian Decision Rule (cont.)

- A classifier with two classes is called **dichotomizer** (binary classifier)
- **Since the state of nature is unknown a natural choice is to describe C in a probabilistic way.**
- We assume that there is ***prior probability*** $p(C_1)$that the student called by the examiner is a girl and $p(C_2)$ that is a boy. The sum of the prior probability over all possible classes, i.e., C_1 and C_2 in our example, must be one.
- If our examiner must decide if the student Smith is a girl or a boy, in absence of further information he is forced to base his decision on prior probabilities.

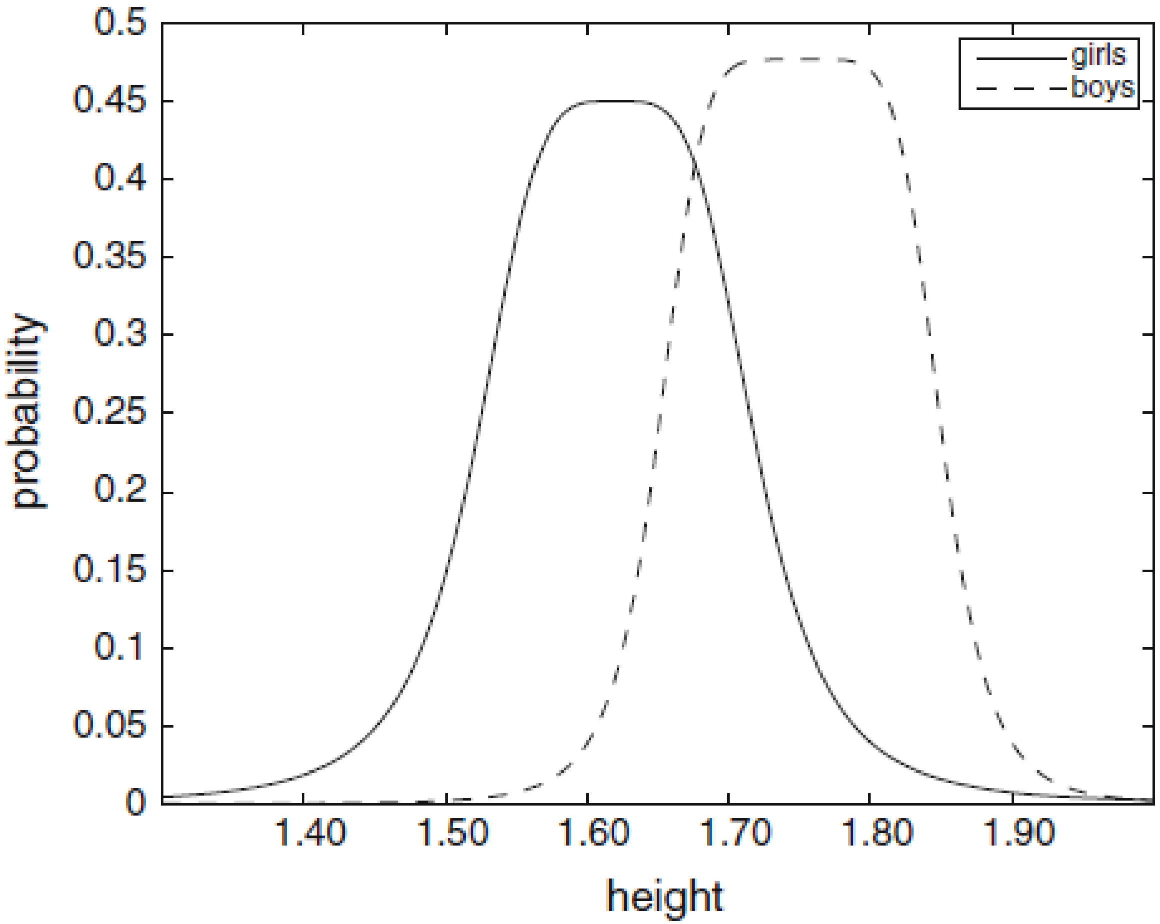
Prior Probability Decision Rule

- Hence he has to apply the following *decision rule*:
- (*Prior Probability Decision Rule*)
  Decide C_1 if $p(C_1) > p(C_2)$; decide C_2 otherwise.

Bayesian Decision Rule

- We suppose that the examiner for each student knows n numeric measurements (or ***features***) $\vec{\boldsymbol{x}} = (\boldsymbol{x_1}, \cdots, \boldsymbol{x_n})$, where, for instance, x_1 is the height, x_2 is the weight and so on. For sake of simplicity we suppose that the features are two and that are height and weight.
- For instance, if the height and the weight of the student Esposito are, respectively, 1.60m and 59 Kg, he can be represented by the *feature vector* (1.60, 59).
- Generalizing we say that each student can be represented by a feature vector (or a ***pattern***) $\vec{\boldsymbol{x}}$.

Hypothetical distribution of students' heights in the classroom



Class-conditional probability density function

- Since different features are associated to different students we can model the feature vector $\vec{\boldsymbol{x}}$ as a random variable whose distribution $p(\vec{x}|C)$ depends on the state of nature C.
- The distribution $p(\vec{x}|C)$ is the ***class-conditional probability density function***, i.e., the probability density function for $\vec{\boldsymbol{x}}$ when the state of the nature is C.

Data Set

- The set of pairs $(X, C) = \{(\overrightarrow{x_1}, C_1), (\overrightarrow{x_2}, C_2), \cdots, (\overrightarrow{x_\ell}, C_\ell)\}$, where the generic $(\overrightarrow{x_i}, C_i)$ means that C_i is the state of nature of $\overrightarrow{x_i}$, is called simply ***data*** (or a ***data set***).
- we assume that **data are *i.i.d*** that stands for ***independent and identically distributed* random variables**.

i.i.d.

- i.i.d. stands for ***independent and identically distributed* random variables**.
- Saying that **data are i.i.d. means that they are drawn independently according to the probability density distribution $p(\vec{x}|C)$** .

Bayes Theorem

- BTD assumes that all the relevant probability values are known, namely we assume that the prior probabilities $p(C_1), p(C_2)$ and the class-conditional probability densities $p(\vec{x}|C_1), p(\vec{x}|C_2)$ are known.
- The joint probability density of finding a pattern $\vec{x}$ in the class $p(C_j)$ is: $p(C_j, \vec{x}) = p(C_j|\vec{x})p(\vec{x})$
- But we have also: $p(C_j, \vec{x}) = p(\vec{x}|C_j)p(C_j)$
- Then we have: $p(C_j|\vec{x})p(\vec{x}) = p(\vec{x}|C_j)p(C_j)$

Bayes Theorem (cont.)

- Then we have the so-called **Bayes Formula**:

$$p(C_j|\vec{x}) = \frac{p(\vec{x}|C_j)p(C_j)}{p(\vec{x})}$$

A posteriori Probability

- The term $p\left(C_j|\vec{x}\right)$ is called a posteriori probability (or simply posterior) expresses the probability that the state of nature is C_j when the pattern $\vec{x}$ has been observed.

Evidence

- The term $p(\vec{x})$ is called ***evidence*** and in the case of two classes is: $p(\vec{x}) = \sum_{j=1}^{2} p(\vec{x}|C_j)p(C_j)$.
- Evidence can be viewed as a normalization factor ensuring that the sum of the probabilities is one.

A posteriori Probability

- Since a posteriori probability $p(C_j|\vec{x})$ is given, for Bayes Theorem, by $p(C_j|\vec{x}) = \frac{p(\vec{x}|C_j)p(C_j)}{p(\vec{x})}$ evidence $p(\vec{x})$ can be neglected and we can conclude that **posterior probability $p(C_j|\vec{x})$ is determined by the product $p(\vec{x}|C_j)p(C_j)$**.
- When **$p(\vec{x}|C_j)$ is large it is *likely* that the sample $\vec{x}$ belongs to the class C_j**.

Bayes decision Rule

- If we consider a pattern $\vec{x}$ for which $p(C_1|\vec{x})$ is larger than $p(C_2|\vec{x})$, it is natural to decide that the pattern $\vec{x}$ belongs to the class C_1; otherwise we assign the pattern to the class C_2.
- This is formalized by the so-called **Bayes decision rule**

Bayes decision Rule

- **Decide C_1 if $p(C_1|\vec{x}) > p(C_2|\vec{x})$; *otherwise decide* C_2.**

Likelihood of C_j with respect to $\vec{x}$

- Therefore the term $p(\vec{x}|C_j)$ **is called the *likelihood of* C_j *with respect to* $\vec{x}$.**

Optimality of Bayes Rule

- Bayes decision rule (or **Bayes classifier**) **is optimal**, i.e., **no other** rule (or **classifier**) **exists that yields a smaller error**.