# MACHINE LEARNING

**Lorenza Saitta**
*Università del Piemonte Orientale, Dipartimento di Scienze e Tecnologie Avanzate, Alessandria, Italy*

**Keywords:** Learning, Basic Knowledge Representation, Learning Methods

**Contents**





**Summary**

*Machine Learning* (ML) investigates ways in which a computer can learn from data or past experience; its main goal is to help humans in constructing programs that cannot be built up directly, and programs that learn and improve from experience. Another goal of Machine Learning is to provide computational models for human learning, supporting thus cognitive studies of learning.

Research in Machine Learning dates back to the 40's, but ML was established as a modern discipline in 1980, whit the first of a series of workshops (later on transformed into international conferences) took place. Since then the field has substantially, covering now a large variety of tasks, approaches, and applications.

ML can be analyzed along different directions. The first one is representation of the learned knowledge. From the first attempts at representing knowledge as simple propositional “concepts”, consisting of the values of a set of attributes, more and more sophisticated language have been used, ranging from subsets of First Order Logics to Bayesian networks. Also attention has been devoted to representation of continuous variables.

A second dimension of analysis is the learning task: supervised learning requires a teacher that is able to label some examples of a concept; usually, the goal of learning, in this case, is to acquire knowledge for classification. On the contrary, unsupervised learning only requires data, and the learner has to find structures in the data, if they

exist. Currently, methods are available to perform semi-supervised learning, in which the burden of the teacher is made lighter. Unsupervised learning is useful for concept formation.

A third dimension is the approach selected. A broad spectrum of methodologies and paradigms is available to researchers and users: symbolic methods of rules and decision trees learning, genetic algorithms, neural networks, reinforcement learning, case-based learning and hybrid methods including more than one approach.

Recently, the explosion of data in databases and the Internet has attracted researchers to develop ML methods for new kind of data and tasks: natural language text, images, video, multimedia data are commonly used by now for new tasks such as information retrieval, information extraction, robotics, business intelligence, bioinformatics and many others.

## 1. Introduction

To act on the environment, at the same time trying to understand the motivations underlying the actions and to trace their effects, to analyze the reasons for success and failure, and to discover laws and relations hidden in apparently disconnected pieces of information are all characteristics of what we believe to be an "intelligent" behavior. They are part of that cognitive activity that is *learning*.

Learning concerns vital functions at different levels of consciousness, starting with recognizing sensory stimuli up to acquiring complex notions for sophisticated abstract reasoning.

With the technological evolution of the information processing tools, machines and programs devoted to solving complex tasks have been built up, tasks whose computational complexity set them out of reach of a manual approach. However, these artificial systems require a great effort from the part of the designer or programmer, and, sometimes it is even not possible to come out with a solution. Moreover, even though a program could be written to solve a problem, it is usually rigid and cannot be extended to cases which are similar but not the same as solved ones, and, in addition, programs do not usually improve their performances with experience in much the same way as humans do.

Machine Learning has its roots in several disciplines, notably statistics, pattern recognition and control theory. Its main goal is precisely to help humans in constructing programs that cannot be built up directly, and programs that learn and improve from experience. Another goal of Machine Learning is to provide computational models for human learning, supporting thus cognitive studies of learning.

In Machine Learning the term "concept" is frequently used. There are two ways of considering concepts: the *extensional* and the *intentional* ones. In the extensional approach, a concept is simply a set of objects (instances). In the intentional approach there are at least three different views. In the *classical* view, which goes back to Aristotle, a concept is a "name" corresponding to a set of necessary and sufficient conditions. The instances are not given explicitly, by they are all those objects that satisfy the conditions. This definition, well suited for mathematical concepts, is inadequate for everyday-life domains. Then, weaker definitions have been introduced. The *heuristic* view of a concept only keeps sufficient conditions, whereas the *exemplar* view considers a concept as a prototype, i.e., a real or virtual example with the “typical” features occurring in the instances. Even though there are some prototype-based Machine Learning approaches, the heuristic view dominates the field. Then, when speaking of a concept, we always mean, an intentional description of it.

Given a concept description φ and an object *e*, it is fundamental to define a procedure to ascertain whether *e* is an instance of φ, i.e., to define a *covering* predicate. If φ is defined as a set of (sufficient) conditions, than *e* is an instance of φ if it satisfies the conditions. On the other hand, if the concept is represented as a prototype, we need to define a *distance* measure and a threshold, and we say that *e* is an instance of φ if its distance from the prototype is less than the threshold. The set of all examples that satisfy a hypothesis φ is said the *extension* of φ, denoted by EXT(φ).

Given a concept and a set of objects, we can classify the objects into *examples* (positive instances) of the concept, and *counter examples* (negative instances) of the concept. The most frequent setting in Machine Learning is *learning concepts from examples,* which consists in the task of acquiring a description of an unknown concept, starting from a set of instances. Learning can occur in a variety of contexts. An important distinction is the one between *supervised* and *unsupervised* learning. Supervised learning assumes that a teacher is able to supply a set of labeled data, from which the learner must derive general laws. Unsupervised learning does not exploit pre-labeled examples, but searches for spontaneous emergence of patterns in the data. The unsupervised approach is much more difficult than the supervised one.

To acquire declarative knowledge, as described above, is not the only goal of Machine Learning. Another one is to speed up performing some task, or to increase the efficiency and scope of problem solving skills.

Along the years, different natures have been attributed to machine learning. At the beginning it has been considered mainly an algorithmic process. One of the first approaches to automated learning was proposed by Gold in his *learning in the limit* paradigm. This type of learning considers that an infinite sequence of pieces of data is provided to the learner, who generates a model that explains the data. At each new input, the learner updates its current model (a "hypothesis"), hoping, but never knowing for sure, that it is closer to the "correct" one.

A fundamental step in Machine Learning has been the recognition of its nature as a search problem. Given a set of data, and some language(s) to describe the data and the target knowledge, learning consists in the exploration of a hypothesis space, guided by heuristics until a specified termination condition is met. As the search space is usually too large to be explored exhaustively, the learner must have a criterion to evaluate and compare hypotheses. In order to facilitate the search, the hypothesis space is usually internally structured according to a *more-general-than* relation. This relation induces a partial order among the hypotheses, as represented in Figure 1.

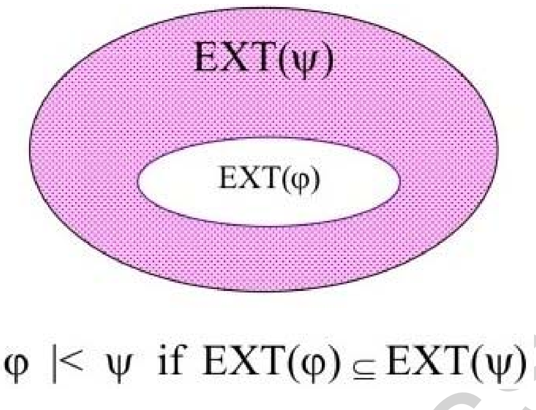


Figure 1**:** Extensional definition of the *more-general-than* relation. Hypothesis $\psi$ is more general than $\varphi$ if $EXT(\varphi) \subseteq EXT(\psi)$

In propositional languages, in which examples are represented as attribute-values pairs, it is also simple to give an intentional definition of the relation, in such a way that it can be evaluated by looking at the syntactic structure of the hypotyheses (formulas). In Figure 2 an example of hypothesis space structured according to the *more-general-than* relation is reported.

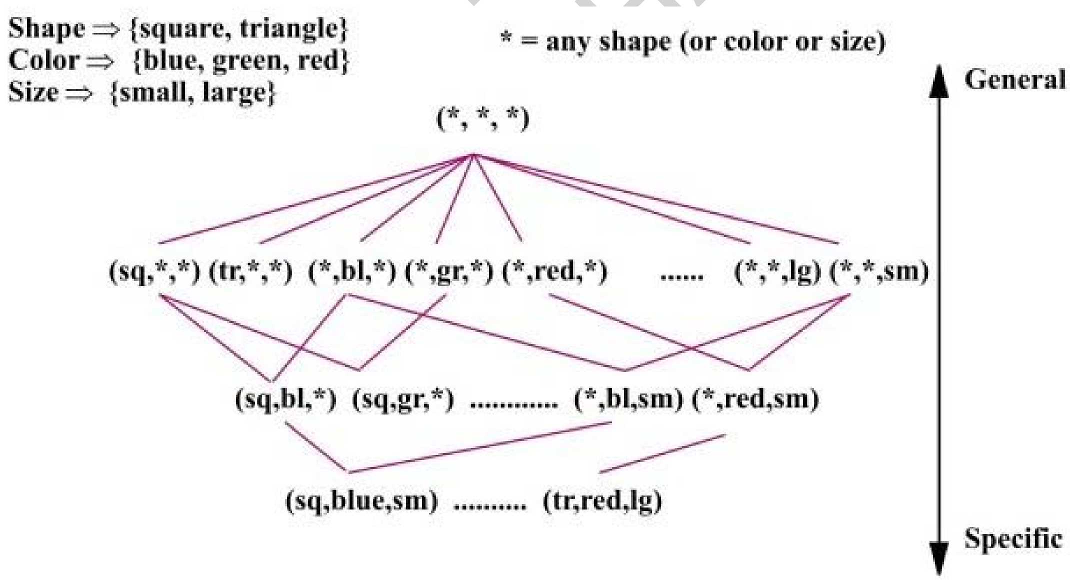


Figure 2**:** Example of hypothesis space structured according to the *more-general-than* relation. Hypotheses describe scenes with two objects

In predicate logic there is no unique intentional definition of the relation. One of the best known is Plotkin's $\theta$-subsumption. Given two formulas $\varphi$ and $\psi$, we say that $\psi$ $\theta$-

subsumes $\varphi$ if there exists a substitution $\theta$ such that $\psi\theta \subseteq \varphi$. Let us consider an example:

$$\varphi = \text{red}(a) \wedge \text{small}(a) \wedge \text{on}(b,a)$$
$$\psi = \text{red}(x) \wedge \text{on}(y,x)$$
$$\theta = \{x/a,\ y/b\}$$
$$\psi\theta \subseteq \varphi$$

In addition to the generality level of a hypothesis, two basic domain-independent criteria, which are widely used, are *accuracy*, and *simplicity* of the learned knowledge. Accuracy refers to an evaluation (error rate, effectiveness, speed up,…) of the behavior of the acquired knowledge on unseen cases. Simplicity is usually related to some syntactic measure of the format of the knowledge, such as number of rules, number of conditions, and so on. An additional criterion, even though difficult to be quantified, is *comprehensibility*, usually based on a human expert's opinion.

## 2. Basic Knowledge Representation and Learning Methods

The first approaches developed, which also produced a number of practical applications, were devoted to infer concept descriptions for classification purposes. The learned concept descriptions were represented as decision trees, or logical formula, either propositional or in a restricted first order logic. The type of representation of the target knowledge strongly influences the learning algorithms and their complexity. The most basic methods for learning induce decision trees or discriminating rules, in the context of supervised learning for classification tasks, and characterizing rules or prototypes, in the context of unsupervised learning for concept formation.

-
-
-



**Bibliography**

Anglano C., Giordana A., Lo Bello G., and Saitta L. (1998). "An Experimental Evaluation of Coevolutive Concept Learning". In *Proc. 15th Int. Conf. on Machine Learning* (Madison, WI), pp. 19-27.

Bellman R.E. (1957). *Dynamic Programming.* Princeton University Press, Princeton, NJ.

Botta M. and Giordana A. (1993). "Smart+: a Multi Strategy Learning Tool". In *Proc. 13th Int. Joint Conference on Artificial Intelligence* (Chambéry, France), pp. 937-943.

Brazdil P. and Michalski R. (Eds.) (2000). *Proc. of the 5th Int. Workshop on Multistrategy Learning* (Guimaraes, Portugal).

Breiman L. (1996). "Bagging Predictors". *Machine Learning , 24,* 123-140.

Carbonell J.G. (1983): "Learning by Analogy: Formulating and Generalizing Plans from Past Experience", in R. S. Michalski, J. Carbonell & T. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach, Vol. I,* Tioga Publishing Company (Palo Alto, USA), pp. 137-161.

De Jong K. A. (1975). "Analysis of the Behavior of a Class of Genetic Adaptive Systems". *Doctoral Dissertation,* Dept. of Computer and Communication Sciences, University of Michigan, Ann Arbor, MI.

Esposito F., Michalski R., and Saitta L. (Eds.) (1998). *Proc. of the 4th Int. Workshop on Multistrategy Learning* (Desenzano del Garda, Italy).

Fayyad U., Piatetsky–Shapiro G., and Smyth P. (1996). "The KDD Process for Extracting Useful Knowledge from Volumes of Data". *Communications of ACM, 39,* 27–34.

Fisher D. H. (1987). "Knowledge Acquisition via Incremental Conceptual Clustering". *Machine Learning, 2,* 139-172.

Freund Y. and Schapire R.E (1999). "A short introduction to boosting". *Journal of Japanese Society for Artificial Intelligence, 14,* 771-780.

Giordana A. and Saitta L. (2000). "Phase Transitions in Relational Learning", *Machine Learning, 40.*

Koza J.R. (1992). *Genetic Programming.* The MIT Press, Cambridge, MA .

Langley P. and Simon H. A. (1995). "Applications of Machine Learning and Rule Induction". *Communications of ACM, 38,* 55–64.

Li M. and Vitányi P. (1997). *An Introduction to Kolmogorov Complexity and Its Applications,* Springer-Verlag, New York, NY.

Michalski R. (1983). "A Theory and Methodology of Inductive Learning". In R. Michalski, J. Carbonell & T. Mitchell (Eds.), *Machine Learning: An AI Approach, Vol. I.* Morgan Kaufmann (Los Altos, CA), pp. 83-134.

Mitchell T. M. (1982). "Generalization as Search". *Artificial Intelligence, 18,* 203-226.

Muggleton S. (1991). "Inductive Logic Programming". *New Generation Computing, 8,* 295-318.

Neri F., Saitta L., and Tiberghien A. (1997). "Modeling Physical Knowledge Acquisition in Children with Machine Learning". In *Proc. of 19th Annual Conf. of the Cognitive Science Society,* Erlbaum Associates, London, UK, pp. 566–571.

Quinlan J.R. (1986). "Induction of Decision Trees". *Machine Learning, 1,* 81-106.

Rumelhart D. and McClelland J. (1986). *Parallel Distributed Processing,* MIT Press, Cambridge, MA.

Saitta L. and Neri F. (1998). "Learning in the 'Real World'". *Machine Learning, 30,* 132-160.

Vapnik V. (1995). *The Nature of Statistical Learning Theory,* Springer Verlag, New York, NY.

Wnek J. and Michalski, R.S. (1994). "Hypothesis-driven Constructive Induction in AQ17-HCI: A Method and Experiments". *Machine Learning, 14,* 139-168.

Wolpert D.H. and Macready W.G. (1995). "No Free Lunch Theorems for Search". *Technical Report SFI-TR-95-02-010,* Santa Fé Institute, Santa Fé, CA.

**Biographical Sketch**

**Lorenza Saitta** is Full Professor of Computer Science at the Dipartimento di Scienze e Tecnologie Avanzate (DI.S.T.A) of the University Amedeo Avogadro, an offspring of the University of Torino recently founded in Estern Piemonte (Italy).

She started her research activity in Pattern Recognition, moving soon to AI, specifically in the area of Fuzzy Logic for Expert Systems. In 1984 she started working in Machine Learning, thus initiating the reaserch in the field in Italy. Herfirst interests have been in inductive symbolic approaches. In this period, she and her research group developed the systems ML-SMART and RIGEL, which learn first-order logic concept descriptions, and have been applied to real-world industrial problems. Later, she worked on the integration of different learning strategies, involving more complex reasoning schemes, such as deduction and abduction Of this period is the implementation of the system WHY, which exploits examples and a causal model of the domain to acquire and revise first-order logic based diagnostic knowledge. More recently, she became interested in Genetic Algorithms (system REGAL) and in Cognitive Sciences.

She is an Action Editor for the Machine Learning Journal, the Convener of the Research Technical Committee of the CEE founded Network Excellence for Machine Learning and Co-Director of the European Science Foundation project on Learning in Humans and Machines. She has also been responsible of, or participated to, several European Research projects.

She gave an invited survey on Machine Learning at ECAI-92, and has been Invited Speaker to the Int. Joint Conf. on Artificial Intelligence (IJCAI-93), the European Conf. on Machine Learning (ECML-94), the Int. Workshop in Inductive Logic Programming (ILP-94) and the Int. Workshop on Artificial Intelligence and Cognitive Science (1994) and will give an invited talk at the 3rd Multistrategy Learning Workshop (1996).

She authored (or edited) three books and more than 100 papers in journals, books and international conferences. She is (or has been) member of various journals Editorial Board and of many International Conferences Program Committees, as, for instance, the Machine Learning Conference 1992, 1993, 1994, the European Machine Learning Conference 1991, 1993, 1994, and ECAI-92. She has been Co-Chairperson of the Int. Conf. on Information Processing and Management of Uncertainty (IPMU-88) and of the Int. Symposium of Methodologies for Intelligent Systems (ISMIS-88). She has also been Chairperson of the International Conference on Machine Learning (ICML-96).