wjert, 2025, Vol. 11 Issue 3, 149-156. *Review Article* ISSN 2454-695X

World Journal of Engineering Research and Technology
WJERT
www.wjert.org
SJIF Impact Factor: 7.029

INTRODUCTION TO MACHINE LEARNING

***Oleg Namicheishvili, Mikheil Ramazashvili and Zhuzhuna Gogiashvili**

Georgian Technical University, Tbilisi, Georgia.

Article Received on 09/01/2025 **Article Revised on 29/01/2025** **Article Accepted on 18/02/2025**

***Corresponding Author**
Oleg Namicheishvili
Georgian Technical University, Tbilisi, Georgia.

ANNOTATION

Machine learning is a very interesting field. Derived from many disciplines, such as statistics, optimization, algorithms or signal processing, it is an ever-changing area of research that has an incredibly large role in society today. For several decades it has been used in automatic character recognition or spam filters, and now it is used to protect against bank fraud, to recommend books, movies or many other products that suit our tastes, to identify people through our cameras or to automatically translate texts from one language to another. In the coming years, machine learning will probably allow us to significantly increase.

- Road safety (among other ways, mass use of autonomous vehicles.
- Response to emergency situations, natural disasters.
- Production of new medical products.
- Energy efficiency of our buildings and industrial sectors.

Machine Learning fascinates computer scientists and mathematicians no less than neurologists, teachers, philosophers and artists. A definition that can be applied to a computer program as well as to a robot, pet or human is given, for example, by Fabien Benureau: «Learning is the change in behavior based on experience».

Machine learning can serve to solve problems when.

- It is not known how to solve it (like the above task about shopping predictions).
- It is known how to solve it, but it is not known how to formalize it in algorithmic terms (this is the case, for example, in face recognition and natural language comprehension tasks).

- It is known how to solve it, but with very demanding procedures for computer resources (this is the case, for example, when predicting the interactions of large molecules, since the difficulties in the modeling process are practically insuperable).

Therefore, machine learning is used when the amount of *data* is relatively large, but the *knowledge* is quite inaccessible or underdeveloped. Thus, machine learning can also help humans learn : Models created by machine learning algorithms can show the relative importance of certain information or how it interacts with other information to solve a particular problem. In the shopping prediction example, understanding the model can allow us to analyze the characteristics of past purchases that predict future purchases. This aspect of machine learning is widely used in scientific research : Which genes and how are involved in the development of certain types of tumors? Which brain imaging areas predict behavior? What characteristics of molecules give them medical properties for specific indications? Which aspects of images captured by telescope provide the opportunity to identify a specific astronomical object?

Machine learning is built on two main pillars. These are the following.

- ❖ On the one hand, data in the form of examples on which the algorithm will train;
- ❖ And secondly, the learning algorithm, which is the procedure to be performed on this data to build a model. Running a learning algorithm on a set of data is called training.

These two pillars are equally important. On the one hand, no learning algorithm can create a good model using irrelevant data - this is the concept of garbage in, garbage out (GIGO) in action. The conclusion is apparent: a machine learning algorithm fed with low-quality data will not be able to make anything but similarly low-quality predictions. On the other hand, a model trained on relevant data with an inadequate algorithm will not have good quality. Although these two terms are often used interchangeably, it is important to distinguish between a *machine learning algorithm* and *a learning model*. The former uses the data to create the latter, which can then be used as a normal classical program. Thus, the learning algorithm allows for event modeling based on examples. It is believed that it is necessary to define the goal and optimize it. For example, the case may concern the minimization of the number of errors made by the model on the training examples. Machine learning can be seen as a branch of artificial intelligence. Indeed, a system that cannot learn is unlikely to become intelligent. The ability to learn and to learn from experience is extremely important for a system designed to adapt to a changing environment. Artificial intelligence, defined as the collection of methods used to create machines capable of exhibiting intelligent—so-called intellectual behavior, also encompasses cognitive science, neurobiology, logic, electronics, engineering, and much more. Perhaps that is why the term "artificial intelligence" arouses more excitement in the imagination of human society and is therefore increasingly used instead of "machine learning".

KEYWORDS: Machine learning, supervised learning, unsupervised learning, incentivized learning.

Content

1. Types of machine learning tasks

Machine learning is a very broad field, and in this subsection we will list only the largest classes of problems that are of interest to us.

1.1. Supervised learning

Among machine learning tasks, supervised learning is probably the easiest type to understand : Its purpose is to learn to make *predictions* based on a list of *labeled* examples, that is, based on the value which has to be predicted (see Figure 1.1). Labels play the role of a "teacher" and control the learning process of the algorithm.

Definition 1.1 (Supervised Learning) *Supervised learning* is a branch of machine learning interested in problems formalized as follows: given n number of $\{\vec{x}^i\}_{i=1,\dots,n}$ observations, which are described in X space, and their $\{y^i\}_{i=1,\dots,n}$ labels, that are described in Y space, it is implied that labels may be received from observations via fixed and unknown function $:\phi:\mathrm{X}\to\mathrm{Y}:\ y^i=\phi\left(\vec{x}^i\right)+\epsilon_i$, where ϵ_i is random noise. Then it is necessary to use the data to determine a function$:f:\mathrm{X}\to\mathrm{Y}$, such that for any pair$\left(\vec{x},\phi(\vec{x})\right)\in\mathrm{X}\times\mathrm{Y},\ f(\vec{x})\approx\phi(\vec{x})$. The space on which the data is defined is most often of $\mathrm{X}=\mathrm{R}^p$ type. But we will also be introduced to working with other types of representations, such as binary, discrete or categorical variables, and even strings or graphs.

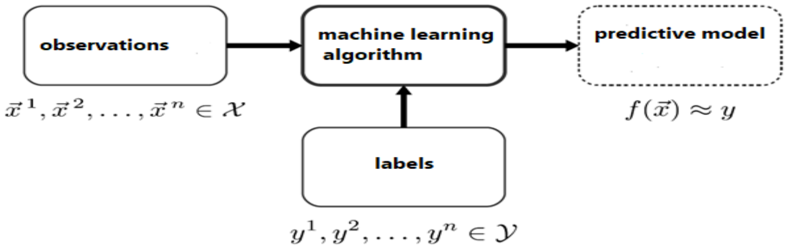


Figure 1.1 – supervised learning.

Binary classification. When the variables are *binary*, they indicate belonging to one *class*, then we speak of *binary classification*.

Definition 1.2 (binary classification) A supervised learning problem in which the space of variables is binary, in other words $Y \equiv \{0,1\}$, is called a binary classification problem.

Multi-class classification. When the variables are discrete and thus correspond to strictly more than two (in Burbak's sense, «multiple») classes, then we speak of a multi-class classification.

Definition 1.3 (Multi-class Classification) A supervised learning problem in which the space of variables is discrete and finite, in other words $Y = \{1, 2, ..., C\}$, is called a multi-class classification problem. C is the number of classes.

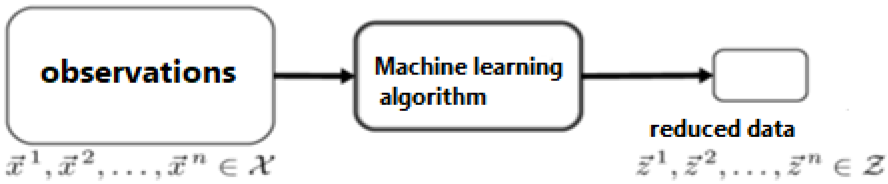
Regression In the case when the variables are represented by *real* values, we talk about *regression*.

Definition 1.4 (regression) The supervised learning problem, in which the space of variables is used, is called the *regression* problem.

Structured regression When the space of variables represents a more complex structured area than the spaces mentioned above, then it is called *structured regression or structured output prediction*. It can be used, for example, to predict vectors, images, graphs or sequences. Structured regression can be used to formalize many problems, such as machine translation or speech recognition problems (for example, text-to-speech and speech-to-text).

1.2. Unsupervised learning

During *unsupervised learning*, data is not labeled. In such a case, it is about modeling the observations for better understanding (see Figure 1.2).



?

Figure 1.2 – Unsupervised learning.

Definition 1.5 (Unsupervised Learning) *Unsupervised learning* is a branch of machine learning concerned with problems that can be formalized as the following: given in X space n number of $\{\vec{x}^i\}_{i=1,\ldots,n}$ *observation*, and the goal is to study a function defined on X space that verifies certain properties.

This definition is rather vague and will, of course, become clearer and clearer on examples. **Clustering** First of all, *clustering*, *or partitioning*, involves identifying groups in the data (see Figure 1.3). This allows us to distinguish between their general characteristics and possibly infer the properties of an observation based on the group to which the observation belongs.

Definition 1.6 (Partition) *Partitioning or clustering* is one of the unsupervised learning problems that can be formalized as n number of $\{\vec{x}^i\}_{i=1,\ldots,n}$ *observations* to find some $\bigcup_{k=1}^{K} C_k$ distribution (grouping). This division (grouping) must correspond to one or more precisely specified criteria.

Dimension reduction. *Dimension reduction* is another important family of unsupervised learning problems. t involves finding a representation of the data in a lower dimensional space than their original representation space (see Figure 1.4). This reduces computation time and the amount of memory needed to store the data, and often improves the characteristics of an algorithm subsequently trained on the same data from supervised learning.

Definition 1.7 (Dimension Reduction) *Dimension reduction* is called an unsupervised learning problem, which can be formalized as a search for a space with a smaller Z dimension than the X dimension in which the n number of $\{\vec{x}^i\}_{i=1,\ldots,n}$ observations is represented. The projections of these $\{\vec{z}^i\}_{i=1,\ldots,n}$ data on the Z space must meet certain, precisely specified, properties.

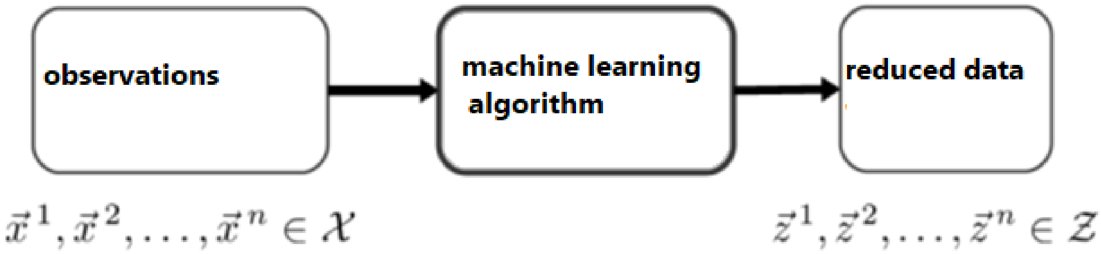


Figure 1.4 – Dimension reduction.

Note. Some dimensionality reduction methods are *controllable:* it is therefore a matter of finding the most relevant representation for a *given label prediction*.

Density estimation Finally, a large family of unsupervised learning problems is actually a traditional problem in statistics: it deals with the estimation of a probability distribution given the assumption that the data set is randomly selected.

1.3. Semi-supervised learning

As you might expect, semi-supervised learning, also called *semi-automated learning or partial learning*, consists in getting guesses from a partially labeled dataset. The first advantage of this approach is that it avoids the need to mark the entire teacher set, which is relevant when the data collection is not difficult, but requires a certain amount of human work to mark them.

For example, in figures classification, a database containing hundreds of thousands of figures is easily acceptable, but applying a label to each relevant sample can be incredibly time-consuming. Moreover, human-made labels are likely to reflect human biases and prejudices, which in turn will be reproduced with great accuracy by a completely controlled algorithm. Semi-supervised training sometimes allows avoiding this underwater boulder.

1.4. Incentivized learning

With *incentivized learning (reinforcement learning)*, the teaching system can interact with the environment and perform actions. In response to these actions, it receives an *incentive*, which can be positive if the action was associated with the correct choice, or negative otherwise. Sometimes an incentive can be obtained as a result of a long sequence of actions: This is the case, for example, in a system that teaches Go or chess.

Thus, teaching in this case lies in establishing politics to determine strategy for systematically obtaining the best possible incentive. The main use of reinforcement learning lies in games (for example, chess or Go) and robotics.

2. Practical recourses

2.1. Software implementations

Many open access programs and libraries provide implementations of machine learning algorithms.

2.2. Data repositories

Many data repositories are open access and allow machine learning algorithms to be explored or retested.

3. Notation system

As far as possible, we use the following notations in this course.s

- String letters (x) denote a scalar.
- String letters with an arrow $(\vec{x})$ on them represent a vector.
- Capital letters (X) represent a matrix, a function or a random variable.
- Calligraphic letters $(\mathcal{X})$ represent multitude or space.
- *The subscripts* correspond to the variable, while the superscripts correspond to the observation: x_j^i is $j-$variable of $i-$observation and corresponds to the X_{ij} data contained in the X matrix.
- n is the number of observations, p - the number of variables, C - the number of classes;
- $[a]_+$ Represents the positive part of $a \in \mathrm{R}$, in other words, is $\max(0.a)$.
- $\mathrm{P}(A)$ Represents A probability of occurrence.
- $\mathrm{E}[X]$ Is X mathematical expectation of a random variable.
- $\mathrm{V}[X]$ Is X the variance of a random variable.
- δ is an indicator function

$$\delta_A = \begin{cases} 1 & If\ A\ is\ true \\ 0 & otherwise \end{cases};$$

- $\langle .,. \rangle$ represents a scalar product on R^p.

– $\langle .,. \rangle_H$ represents a scalar product on H.

– $M \geq 0$ Means that M is a symmetric positive semi-definite matrix.

4. Key moments (conclusion)

- A machine learning algorithm is an algorithm that trains a model on examples by reducing problem solving to an optimization task.
- Machine learning is used when it is difficult or impossible to define clear instructions for a computer to solve a problem, but there are many obvious illustrative examples.
- Machine learning algorithms can be classified according to the nature of the problem they are trying to solve. These are: supervised, unsupervised, semi-supervised and incentivized learning algorithms.

5. BIBLIOGRAPHY

1. Bakir, G., Hofmann, T., Schölkopf, B., Smola, A. J., Taskar, B., et Vishwanathan, S. V. N. (2007). *Predicting Structured Data*. MIT Press, Cambridge, MA. https://mitpress.mit.edu/books/predicting-structured-data.
2. Barto, R. S. et Sutton A. G. (1998). *Reinforcement Learning :An Introduction*. MIT Press, Cambridge, MA. http://incompleteideas.net/book/the-book-2nd.html.
3. Benureau, F. (2015). *Self-Exploration of Sensorimotor Spaces in Robots*. Thèse de doctorat, Université de Bordeaux. https://tel.archives-ouvertes.fr/tel-01259955/file/BENUREAU_FABIEN_2015.pdf.
4. Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 44(1.2): 206–226. https://hci.iwr.uni-heidelberg.de/system/files/private/downloads/636026949/report_frank_gabel.pdf.
5. Scott, D. W. (1992). Multivariate density estimation. Wiley, New York. https://www.worldcat.org/title/multivariate-density-estimation-theory-practice-and-visualization/oclc/710751351.