Introduction to Machine Learning

Taras Rashkevych

April 12, 2021

Summary

Abstract

My personal notes on the contents of the Introduction to Machine Learning course held by professor Elisa Ricci at the University of Trento. These notes should provide a broad and complete introduction to the world of Machine Learning and Statistical Pattern Recognition and also be the basis for further courses on more deep topics in this area.

1 Introduction

1.1 Overview

The following are the three main families of machine learning methods, which are also the topics of these notes:

- Supervised Learning: parametric/non-parametric algorithms(e.g. nearest neighbors, decision trees and random forests), kernel methods, deep neural networks (e.g. feed-forward, convolutional and recurrent networks).
- Unsupervised Learning: clustering, dimensionality reduction, autoencoders, deep generative models.
- Reinforcement Learning: these notes only cover a high-level introduction.

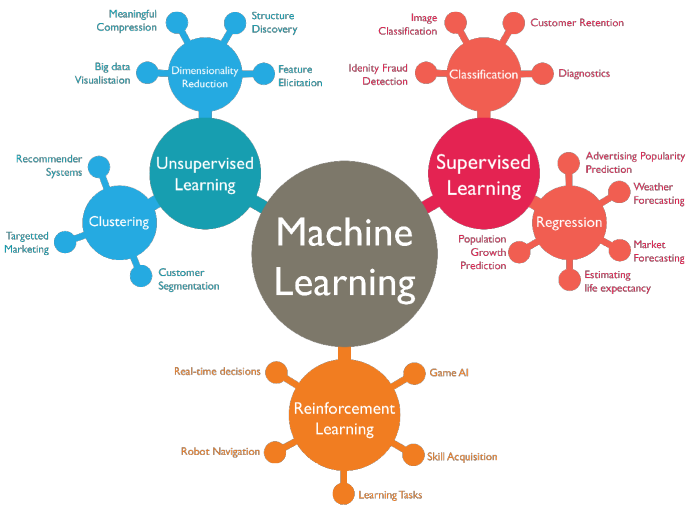


Figure 1: High-level description of the world of Machine Learning

1.2 What is Machine Learning?

There are several definitions of what Machine Learning is, among which we can find the following that perfectly reflect its conceptual nature:

> *"Machine learning is the study of computer algorithms that improve automatically through experience. It is seen as a part of artificial intelligence."*
>
> — Wikipedia

> *"Machine learning is the science of getting computers to act without being explicitly programmed."*
>
> — A. Samuel (1959)

Given all these expressive and equivalent definitions, the general idea of what Machine Learning is and how it differentiates from the traditional way of programming can be summarized by the following image:

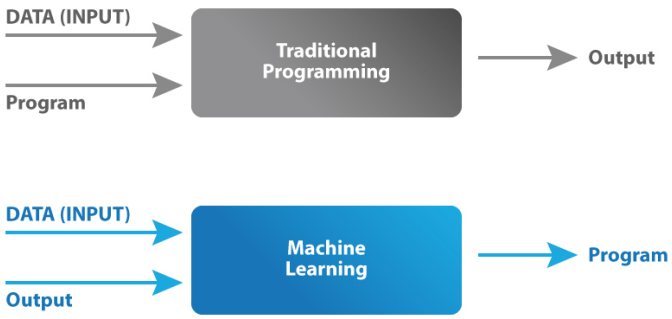


Figure 2: Machine Learning vs Traditional Programming

By observing the image it becomes clear that in traditional programming the programmer writes a program and then gives it some input data in order to produce some output. In machine learning, on the other hand, the conceptual model is totally different because in this case the programmer still gives some input data but also gives some known result (e.g. data with annotations/labels), which in combination with the previous ones are used to produce a program. In this way the machine learns from the input data and the data with annotations to be able write a program.

Some problems are so hard to solve that it becomes not impossible but extremely difficult to write programs that solve them and even if such programs are produced by writing them manually then it is likely that these programs will not be able to provide sufficiently satisfiable results. So in these cases it is preferable to make the machine learn from the data and write the program that will provide some better results. An excellent example is for instance a robot that has to learn how to walk because it is extremely difficult to write a suitable program for this kind of task but it is extremely useful to make the robot learn from the data because in this way it will also capture the real-world noise and in general all the unexpected things, which could not be modelled by a program written manually. Unfortunately there are many other examples in which the machine learning approach seems to be the most promising. Nevertheless all the solutions obtained with the machine learning approach present a common structure, in the sense that there are always *three fundamental elements* which are the following:

1. ***Data***: *a lot of data* is required in order to build complex and more or less reliable models.

2. ***Algorithms***: there are *a lot of algorithms* that can process the data mentioned at the point 1, each of which has its own peculiarities. The choice of a particular algorithm always depends on the task to be addressed.

3. ***Model***: this is the result of applying the algorithms mentioned at the point 2 to the data mentioned at the point 1, which will be used on future data *similar* to the ones that were used for its training. So this result is somehow the summary of the knowledge that has been extrapolated from the data and represents the so-called *"knowledge"* that computers acquire by processing the data through the algorithms.

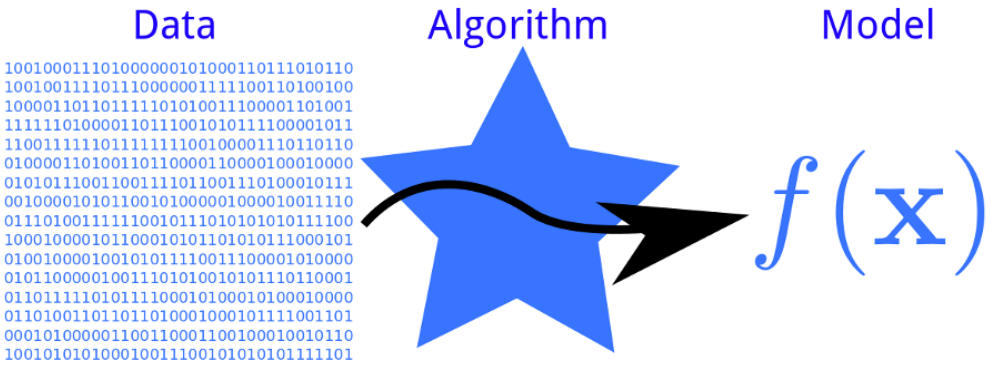


Figure 3: Three fundamental elements in Machine Learning

So the main reason why machine learning is so important and so widely used consists in the **Hardness** of manually writing programs that should represent satisfiable solutions to a particular class of problems. This hardness is then instantiated based on the particular nature of the problem:

- **Lack of Expertise**: Human beings have no idea how they should program the robot to navigate the surface of Mars because until now no human being has had the opportunity to explore this planet directly. So it is better for the machine to learn what to do autonomously.

- **Lack of Expressiveness**: Sometimes it is too difficult to explain the human experience such as sight or hearing and so to define a precise set of rules to follow. A great example is speech recognition where deducing the person's name from a particular waveform becomes significantly complex, therefore the best solution is to learn from the data and improve performance accordingly.

- **Personalization**: There are situations in which some kind of personalization is required and is not feasible for a manually written program to provide such personalization to all customers. Excellent examples are personalized medicine and the recommendations provided by streaming services.

- **Big Data**: In almost all situations in which machine learning is used a big amount of data is required for model training, but there are some situations in which the amount of data to be processed and to be reasoned about in order to achieve a particular goal is exceptionally large and therefore could not be handled by a manually written program. A great example is genomics where the amount of data to be used is extremely large.

It should be mentioned though that there are situations in which machine learning **is not useful.** Generally this happens when the rule, which determines how the machine is supposed to act, is perfectly known. So if the final goal is to calculate a payroll or perform a calculation using a well-known physical law, then traditional programming is perfect here.

After having seen the main reasons why machine learning is so widely used, it is mandatory to mention some of its most important practical applications:

- **Pattern Recognition**: This classic application consists in recognizing patters, which means that, given a certain input such as an image of handwritten digits, facial identities, facial expressions or a medical image, it is possible to recognize certain similarities and continuous repetitions of some structures seen before and be able to deduce the targeted content of that input. It is useful to mention that images are not the only available data type used in pattern recognition.

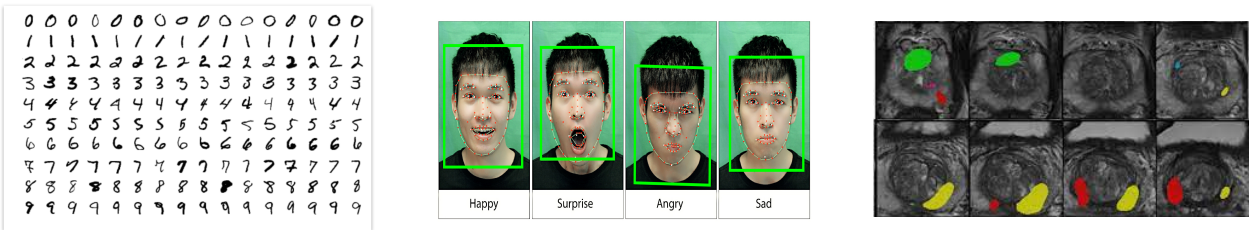


Figure 4: Some examples of Pattern Recognition

- **Pattern Generation**: This is another application which is extremely used nowadays and consists in generating patterns, which means that, given a certain distribution of data to learn from, it is possible to produce samples based on these data. For example with this technique it is possible to generate fake images and artificial motion sequences.

Figure 5: Some generated portraits of famous people

- **Anomaly Detection**: This is another important application which consists in detecting anomalies, which means that, given a certain amount of data, it is possible to produce machine learning models that will be able to predict unusual behavior. Some relevant examples are unusual credit card transactions, unusual patterns of sensor readings, unusual video surveillance frames and unusual system logins.

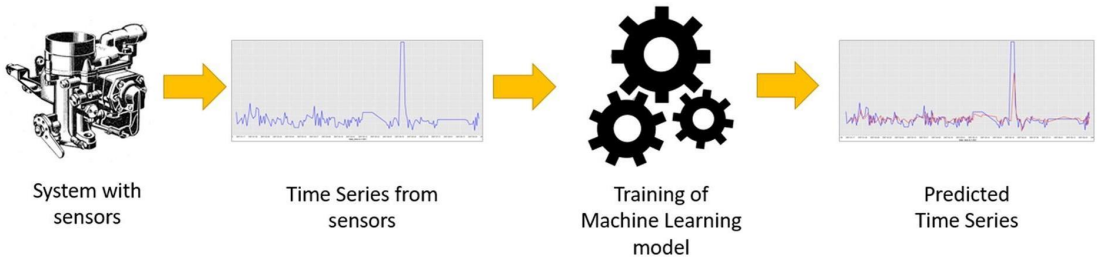
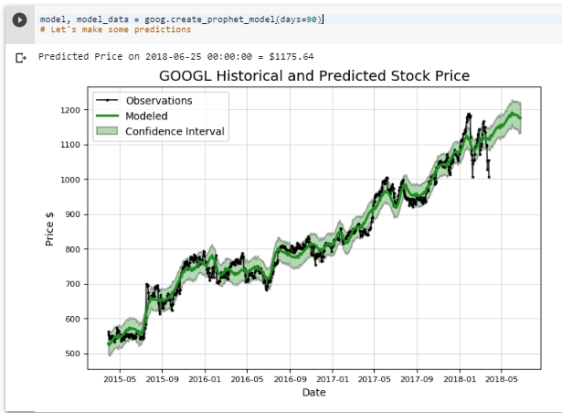


Figure 6: Anomaly detection in a system with sensors

- **Prediction**: This is an extremely important application in finance since it provides predictions on future stock prices or currency exchange rates. Nowadays it is also used in autonomous driving to predict future moves of people and other vehicles and to avoid possible accidents. Moreover, it is also used in gaming to predict future best moves, as it has been demonstrated by the AlphaGo program developed by the Google DeepMind group.



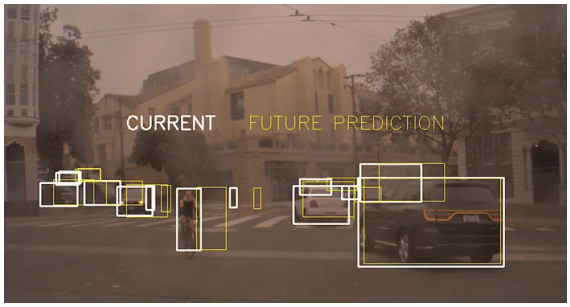


Figure 8: Some examples of Prediction

At this point, it is worth mentioning some other definitions of what Machine Learning is, provided by some big experts in this field:

> *"It is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions."*
>
> — Christopher M. Bishop

"The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest."

— Kevin P. Murphy

"Machine learning is about predicting the future based on the past."

— Hal Daume III

Given these last three definitions, it is possible to detect two fundamental steps of machine learning consisting in automatically discovering some regularities(e.g. patterns) and applying these regularities to take some actions, such as predicting future data. At this point, it becomes natural to introduce the typical machine learning pipeline:

1. ***Training***: At this stage of the process the past knowledge is represented by the so-called *training data*, which will be passed through a particular *algorithm* to produce the so-called *model*, also called *predictor*.

2. ***Testing(Inference)***: At this stage, on the other hand, a component of the future knowledge is represented by the so-called *testing data*, which will be passed to the *model*, generated during the first stage, to produce some *predictions*, which are the other component of the future knowledge.

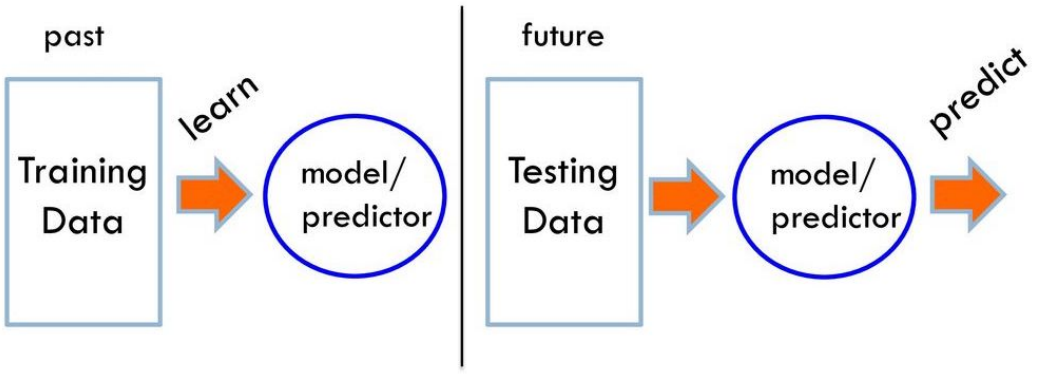


Figure 9: Typical machine learning process

The last definition of Machine Learning, which will be seen in this section and which introduces the concept of performance measure, is the following:

> *"A computer program is said to learn from **experience E** with respect to some class of **tasks T** and **performance measure P**, if its performance at tasks in T, as measured by P, improves with experience E."*
>
> — T. Mitchell (1970)

Basically what this definition states is that machine learning is the study of algorithms that:

- improve their performance P
- at some task T
- with experience E

At this point, thanks to this definition, it is possible to describe any machine learning problem by defining the triplet $\langle T, P, E \rangle$, which is then instantiated based on the particular nature of the problem:

- **Example 1.1** *T: Recognizing handwritten words. P: Percentage of words correctly classified. E: Database of human-labeled images of handwritten words.*
- **Example 1.2** *T: Driving on four-lane highways using vision sensors. P: Average distance traveled before a human-judged error. E: A sequence of images and steering commands recorded while observing a human driver.*
- **Example 1.3** *T: Categorize email messages as spam or legitimate. P: Percentage of email messages correctly classified. E: Database of emails, some with human-given labels.*

Given all these high-level definitions, reasons of usage, practical applications and some more detailed explanations of the fundamental elements of machine learning, it is worth mentioning some of the real-world success stories:

- *Face Detection*: First working system in 2002.
- *Pedestrian Detection*: First working system in 2005.
- *Body Tracking(RGB-D)*: Used for example in gaming.

1.3 What is Deep Learning?

In order to understand what Deep Learning is, it should be useful to observe the image below that represents the popularity growth in chronological order of Artificial Intelligence and its two most famous subsets, which are Machine Learning and Deep Learning.

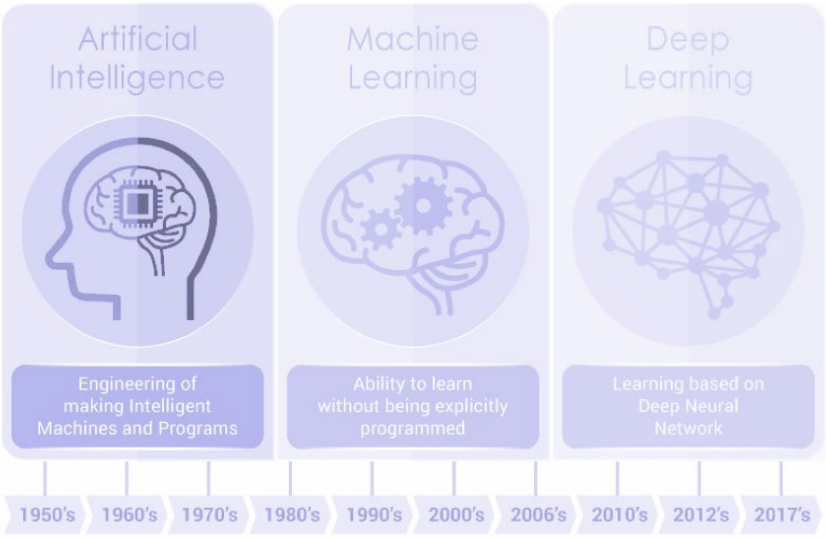


Figure 10: Artificial Intelligence, Machine Learning and Deep Learning

It immediately becomes clear that deep learning is now the most studied field(even if the idea of neural networks on its own is quite old) of the entire artificial intelligence area and is also the most promising one. In fact, most systems nowadays that apply some kind of machine learning are actually based on deep learning models. Furthermore, to understand even more deeply the relationships between the three areas mentioned above, the next image could be useful:

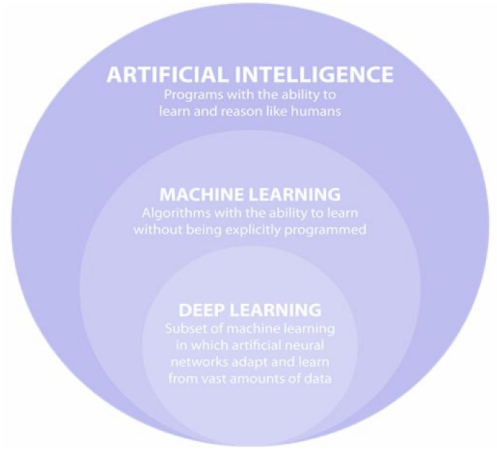


Figure 11: Artificial Intelligence, Machine Learning and Deep Learning

So, as the above images suggest, deep learning is the most recent actually working accomplishment of machine learning and artificial intelligence in general, which allows constructing complex and efficient machine learning models that are based on the so-called artificial neural networks, which adapt and learn from particularly large amounts of data.

At this point, it is worth mentioning some definitions of artificial intelligence and deep learning, provided by big experts in these areas:

> *"Our ultimate objective is to make programs that learn from their experience as effectively as humans do."*
>
> — J. McCarthy(1958)

> *"Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction"*
>
> — G. Hinton — Y. LeCun — Y. Bengio

The actual article about deep learning, its fundamental idea, and its applications written by the fathers of deep learning can be found at this address: `https://www.nature.com/articles/nature14539`.