Introduction to Machine Learning

Taras Rashkevych

April 21, 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. feedforward, convolutional and recurrent networks).
- ***Unsupervised Learning***: clustering, dimensionality reduction, autoencoders, deep generative models.
- ***Reinforcement Learning***: these notes only cover a high-level introduction.

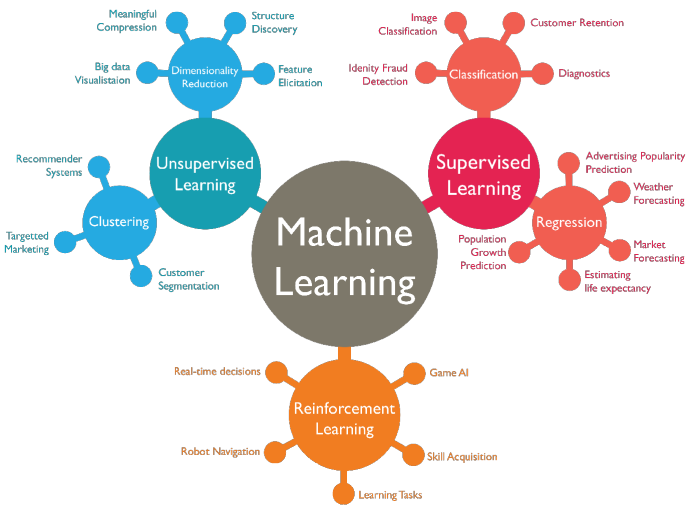


Figure 1: High-level description of the world of Machine Learning

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:

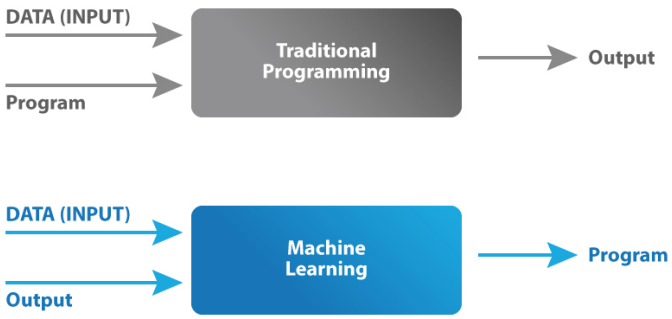


Figure 2: Machine Learning vs Traditional Programming

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 in 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 in point 2 to the data mentioned in 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.

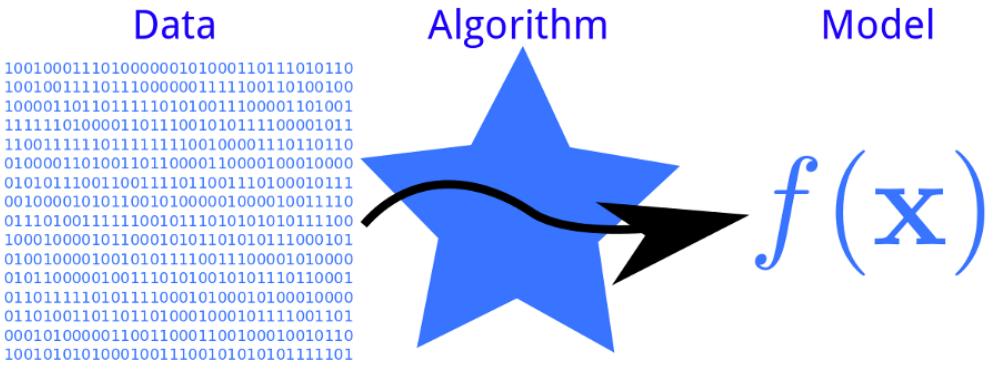


Figure 3: Three fundamental elements in Machine Learning

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.

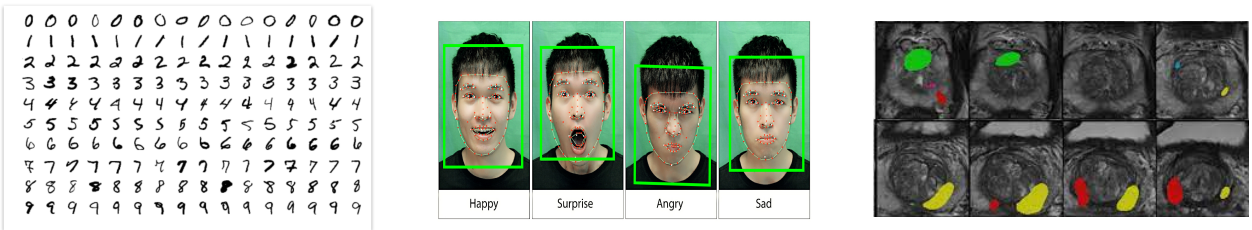


Figure 4: Some examples of Pattern Recognition

- **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.

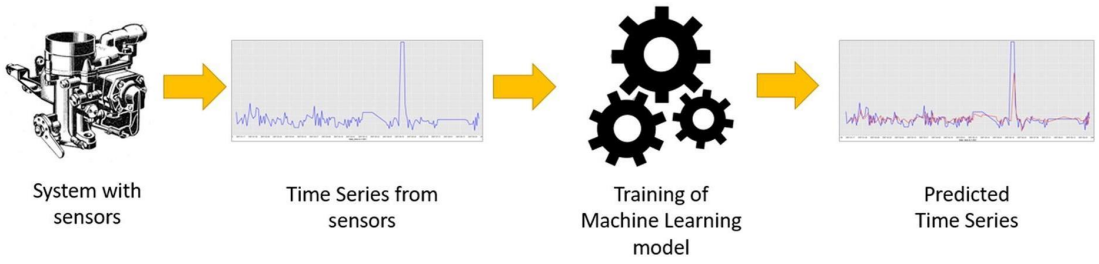
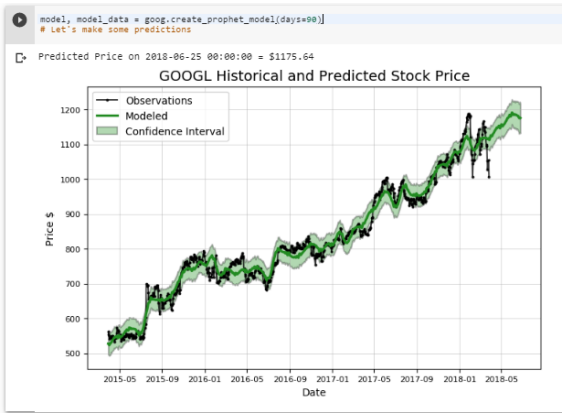


Figure 6: Anomaly detection in a system with sensors

- **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.



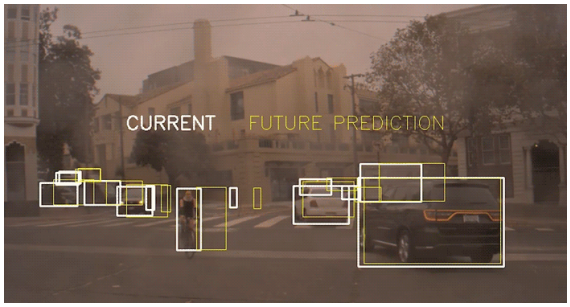


Figure 8: Some examples of Prediction

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(i.e. 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*. During this stage, the so-called *feature extraction* also takes place, which basically consists in extracting from the raw data of the training set some kind of representations, called *features*, which are essential for adequate processing. These features will also be used by the aforementioned model to process future data.

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.

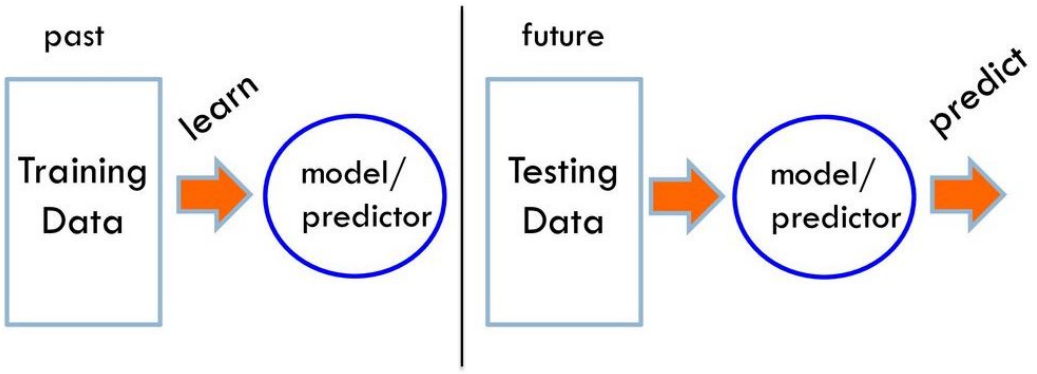


Figure 9: Typical machine learning process

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*: The first working system in 2002.
- *Pedestrian Detection*: The 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 in chronological order the popularity growth of Artificial Intelligence and its two most famous subsets, which are Machine Learning and Deep Learning.

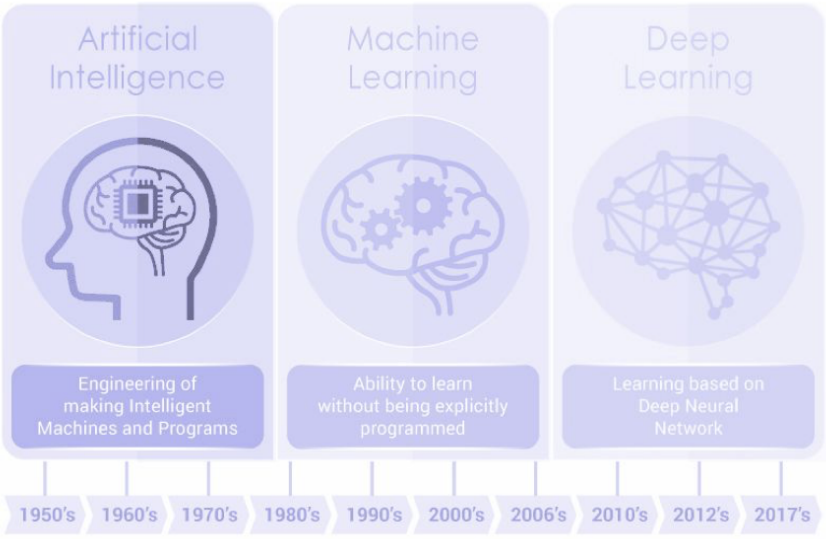


Figure 10: Artificial Intelligence, Machine Learning and Deep Learning

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:

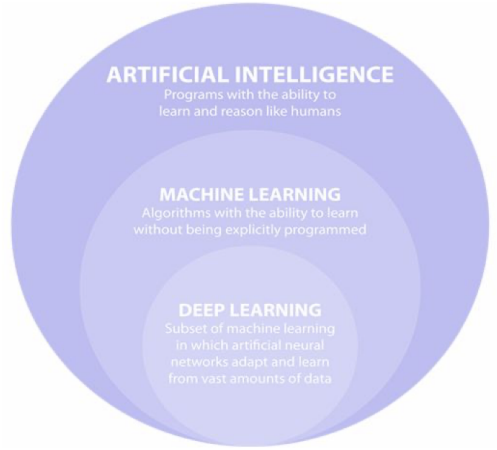


Figure 11: Artificial Intelligence, Machine Learning and Deep Learning

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. These neural networks actually consist of several layers of nodes between input and output that apply, instead of using the feature extrapolation process, a hierarchical processing so that the network itself automatically learns a mapping between the initial raw data and the final output. In fact, in order to produce this mapping, some form of representation of the input data is computed on each individual layer, with the abstraction gradually increasing(i.e. from particular details to general concepts) by aggregating the information from the lowest layers near the input layer towards the highest layers near the output layer. A summary of this process can be seen in the image below:

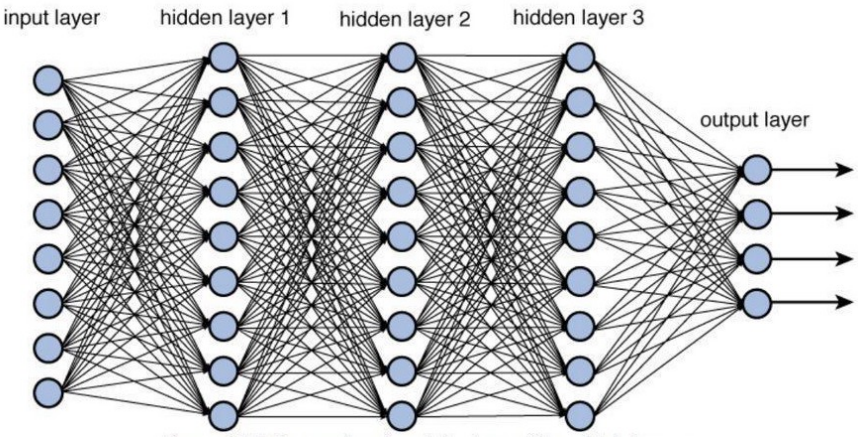


Figure 12: Deep Neural Network

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(they won the 2018 Turing Award for conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing) can be found at this address: `https://www.nature.com/articles/nature14539`.

At this point it is useful to mention some extremely important real-world success stories that made deep learning so popular and so widely used:

- *Object Recognition*: This is a general machine learning task that is composed of other two machine learning tasks called Image Classification and Object Detection. The former consists in assigning to each object present in an image or in a frame of a video a label indicating the category to which that particular object belongs and a parameter indicating the degree of accuracy. The latter consists in assigning the position of that particular object by indicating the so-called bounding box, which defines the rectangular boundaries of where the object is located. The first actually working deep learning system that solves this general task on the dataset called ImageNet was created in 2012. Furthermore, in 2015, the error rate of another system on the same dataset has decreased to the level of a human being, which once again underlines the power of deep learning.

- *Image Captioning*: This is a natural extension of Image Classification and a much more challenging task, which consists in assigning a textual description to an image or to a frame of a video. The first actually working deep learning system that solves this kind of task was created in 2015 and an excellent example of a similar system can be seen in this youtube video. Nowadays the systems that solve this kind of task have been further improved in performance.

- *Image Synthesis*: This is another machine learning task that consists in automatically generating plausible images associated with the sketch of a scene passed as input. Such a system has been successfully developed by Nvidia.

- *Speech Recognition*: This a self-explanatory machine learning task that has been studied for years and has been significantly improved in performance by using some deep learning techniques starting from 2009.

- *Neural Machine Translation*: This a machine learning task that consists in translating sentences from one language to another and that has also been significantly improved in performance by using some deep learning approaches starting from 2014.

- *Real-Time Voice Translation*: This is a machine learning task that consists in translating a person's speech in real-time by generating an artificial voice that speaks in the target language. An excellent example of a system that performs this kind of task can be seen in this youtube video.

- *AlphaGO*: This is the aforementioned program that has been developed by the Google DeepMind group with the purpose of predicting the best future moves in an ancient Chinese game called Go and that has actually beaten the best world player at Go. An interesting fact is that a movie has been made about this story and the trailer can be seen in this youtube video.

To conclude this subsection it is mandatory to mention the main reasons why deep learning has only come into play in the last few years:

- *Flood of available data*: Nowadays the amount of data available is extremely large.
- *Increased computational power*: There have been continuous improvements in hardware manufacturing that have made it possible to produce devices with a more powerful computational ability. In fact, an extremely important role for processing in the field of machine learning is played by GPUs.
- *Research Development*: There are a lot of new machine learning algorithms and theory developed by researchers.
- *Industry Involvement*: There is much more interest and support provided by companies, such as OpenAI, DeepMind and many others.

2 Data, Features, Models

This section will mainly focus on data, features, and models. As has been mentioned in the previous section the first key element of machine learning is *data*. Then, by processing these data, the second main key will be extracted, the so-called *features*, which are the most relevant part of the data that will be used for learning. Finally, based on the features extracted previously, it will be possible to build the so-called *models* that will represent the summary of the knowledge acquired through the learning process and will be used to perform some actions.

2.1 Learning Process

Based on the results provided in the previous section, it is time to introduce a more complete description of the machine learning pipeline, as illustrated in the image below:

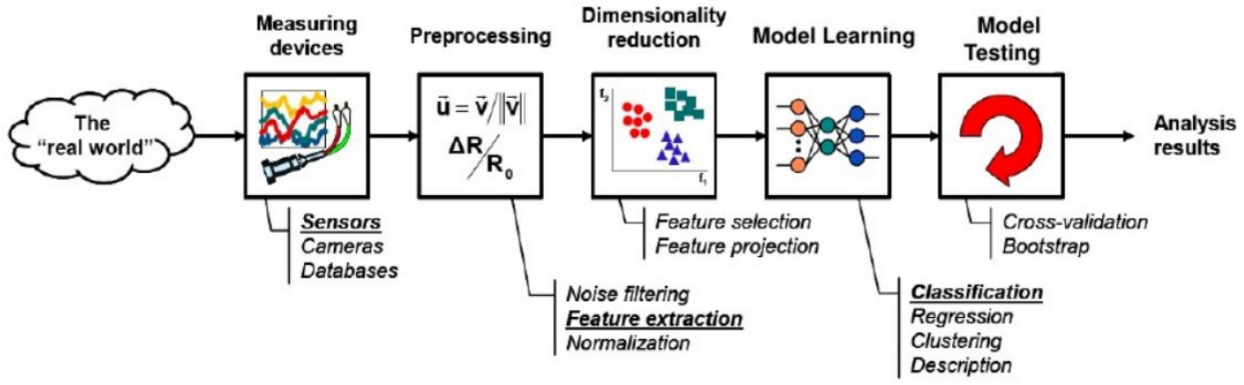


Figure 13: Learning process in detail

So, as can be observed in the image above, the learning process in real-world applications where some kind of machine learning is used is actually quite complex and it is also necessary to understand how to perform each step of the pipeline shown below in order to build an actually working machine learning model:

1. ***Data Source***: This is the actual place where all data are generated. Normally this place is represented by the real world itself.

2. ***Data Collecting***: This is the process by which all data are collected from the data source mentioned in point 1 by using different devices, such as sensors, cameras, and databases.

3. ***Data Preprocessing***: This is a process that depends on the type of the problem which is being addressed. Normally this process consists in some noise filtering, if the data are images or sensor signals, in normalization, if the data are seen as vectors, or in the so-called *feature extraction*, which is achieved by transforming the data into a set of vectors containing relevant information.

4. ***Dimensionality Reduction(Optional)***: Sometimes the features from the previous step are not used directly to produce a model, but are carefully selected to extract the most relevant part through the process called *feature selection*.

5. ***Model Learning***: This is the core step of the machine learning process in which, by applying a particular *learning algorithm*, the actual *model* is generated. The actual type of algorithm that will be applied depends on the particular nature of the problem being solved, such as classification, regression, clustering, description, and many others.

6. ***Model Testing***: Once the model has been generated through a particular learning algorithm, a particular model testing protocol is applied to validate the accuracy of the generated model.

It is also important to mention that another element that adds complexity to the design of machine learning models is the choice of a particular algorithm to use. This is because of the fact that there is a huge number of machine learning algorithms that could be used. Fortunately, there is always a guide that helps in making the right decision, that is the particular nature of the problem to be solved.

At this point, given the general machine learning pipeline described above, it is time to dive into the first most important element of the process which is *data*. The first problem to solve is the fact that the concept of data is quite abstract, but the concrete data that are used in real-world applications can be extremely different form each other. For this reason, it is necessary to define a general procedure that will allow representing data independently of their actual structure. The answer to this problem is called *feature extraction*: that is the process by which each *example* in the *training set* is associated with a data structure, which is normally a simple vector of numbers with cardinality n, that *represents* the relevant information about the example and indicates the *actual form* that is seen by the algorithms. Each such vector takes the name of *feature*.

One way of extracting these features is to consider them as *questions that can be asked* about the example, like for instance in the image below with the fruits:

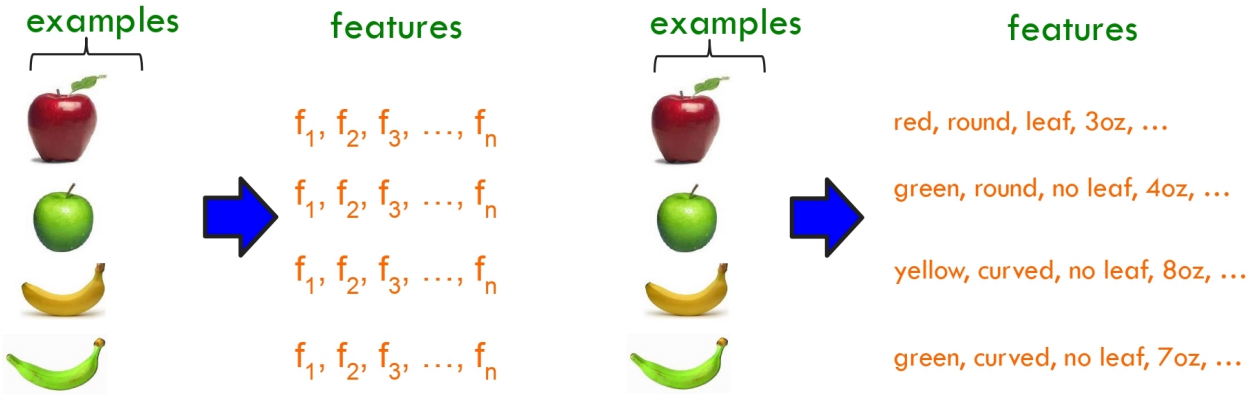


Figure 14: Example of features

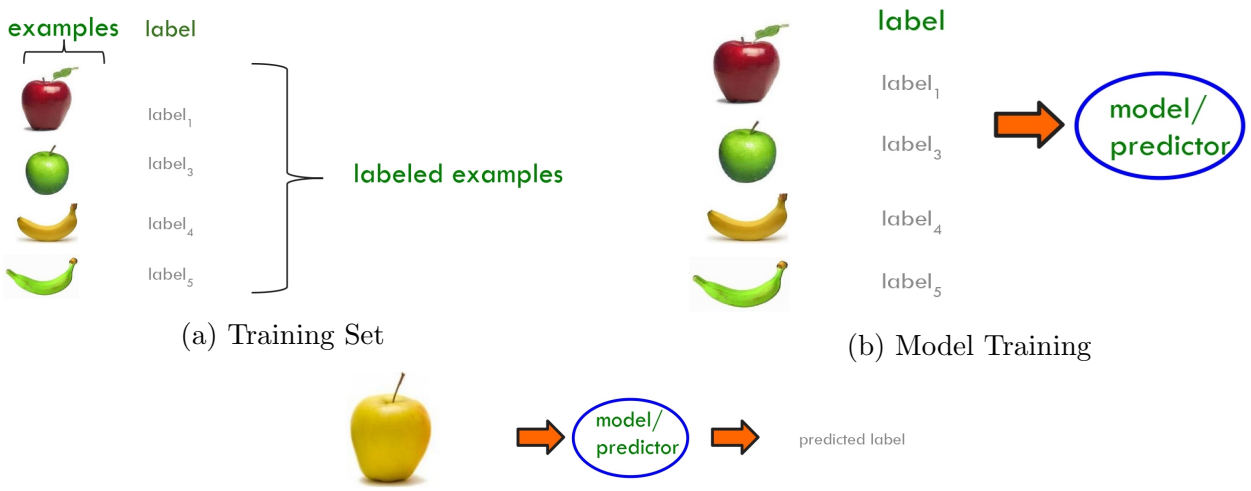
Therefore, the final result of the feature extraction process is the actual *training set*, which is a set of numerical vectors, each of which is the same dimension as the others. Unfortunately, there is a problem with this approach, which is actually the most delicate part of the design of the machine learning pipeline: *how* are these features chosen? There are actually many answers to this question and the best answer should be the one that will ideally produce a set of features that represent as well as possible the original data without any loss of information. Therefore, there is always a risk that, after processing the input data, some information associated with the real data may be lost, which in the case could cause the machine learning algorithm to work improperly.

2.2 Machine Learning Methods

Now it is time to talk about different families of machine learning methods to add a deeper understanding to what has been briefly described in the overview subsection of these notes.

2.2.1 Supervised Learning

The first big family of machine learning methods is the so-called *Supervised Learning*. All methods associated with this family have access to the data *and* some annotations that are associated with these data. These annotations are also called *labels*. Therefore, the training set, that is the input to the learning algorithm, in these cases consists of a set of pairs, each of which consists of an *example*(in the form of a *feature vector*) and its relative *label*. For this reason, all the pairs in the training set are called *labeled examples*. Then this set of labeled examples is processed by the learning algorithm to build a model, also called a predictor, that will be used on new data, which have not been seen during the training phase, to produce labels associated with these data. It is necessary to note that all labels generated by the model will belong to the set of all possible labels that have been learned from the training set. All this process can be observed in the images below:



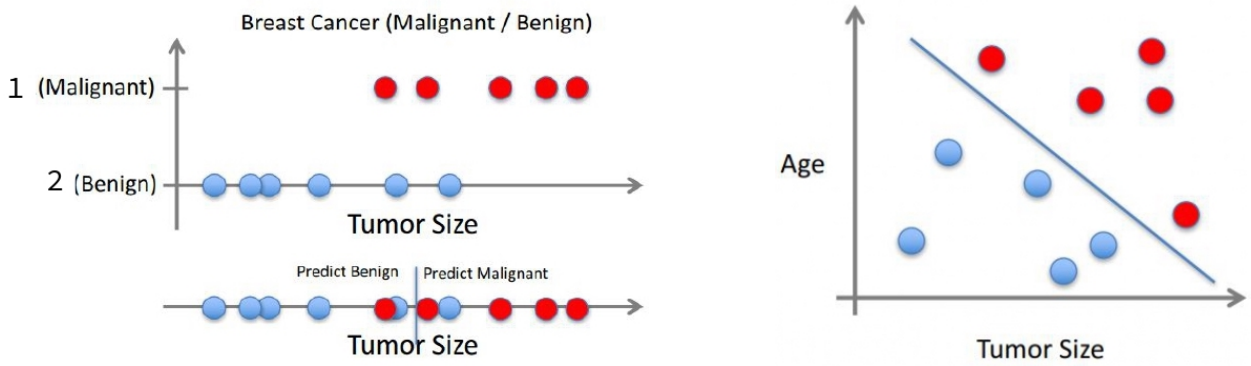
(a) Training Set

(b) Model Training

(c) Model Testing

At this point, it is possible to introduce two major machine learning tasks that use the supervised learning structure defined above:

- ***Classification***: Given a finite training set $\mathcal{T} = \{(x_1, y_1), ..., (x_m, y_m)\}$ where m is the total number of pairs of labeled examples, the task is to learn a function f, defined as $f : \mathbb{R}^d \rightarrow \{1, 2, ..., k\}$ where d is the dimension of the input space and k is the total number of labels, to predict the label y given the input x. So the dimension of the x component(feature vector) in the pairs of labeled examples and as input to the function f is determined by the value of d, as can be observed in the examples below:



(a) x is one-dimensional($d = 1$)

(b) x is multidimensional($d = 2$)

This classification framework of supervised learning is of course extremely general and can be instantiated for several problems, such as Face Recognition, Character Recognition, Spam Detection, Medical Diagnosis, and Biometrics.