Chapter 1

Introduction to Machine Learning (ML)

1.1. Introductory concepts *

Machine learning is composed by a set of methodologies aiming to develop computer programs capable of improving their performance with their own experience.

(Let us use simply the word learning as machine learning)

* Adapted from Mitchell, Chap. 1 e 2

Introductory concepts

- The learning algorithms are trained with a dataset of known cases prepared by the human user (for example thousand of credit card transactions, some legitimate, some fraud). This is the training dataset.

- The training data contains embedded information (features) that have been used by the training algorithm to find parameter values of its decision function (in the example this function has two values, legitimate-fraud).

- When the following transaction comes, the algorithm classifies it as legitimate or fraud.

- When the algorithm gives an erroneous response, this error, detected a-posteriori, is used to change the values of the parameters of the decision function in a way to decrease the number of errors, improving by this way, with time, the performance of the algorithm.

Introductory concepts

We can then define learning as (Mitchell, p. 2):

Learning: “ 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.”*

Introductory concepts

- identification of people in crowd images,
- fraud detection in banking and finance,
- word identification in automatic translation,
- medical diagnosis
- industrial production

….. are some among many applications of machine learning .

Introductory concepts

Learning of a concept: to infer a boolean function using a training set of examples. Each example contains inputs and outputs of the function to be inferred.

Introductory concepts

Instances: a set X of items on which the concept is defined. For example, let X be the set of all possible days. Each day (instance) is represented by a set of attributes such as *sky outlook, temperature, humidity, wind, rain.*

The target concept, to be learned, can be for example "*good day to play tennis*".

Introductory concepts

- Using a training set of examples, the *target concept* is learned

- The training examples must include positive and negative cases; they must be representative of all possible *instances.*

- But the target concept will only be exactly known if the training set contains all possible instance in the problem framework.

- This is impossible because on one side it would be very big and on the other side future instances are not known at present.

- As a consequence, the best we one can have is a hypothesis or an estimation of the target concept .

Introductory concepts

The Inductive Learning Hypothesis (Mitchell, p. 23):

- " Any hypothesis found (after training) to approximate the target function well over a sufficiently large set of training examples, will also approximate the target function well over other unobserved examples. "

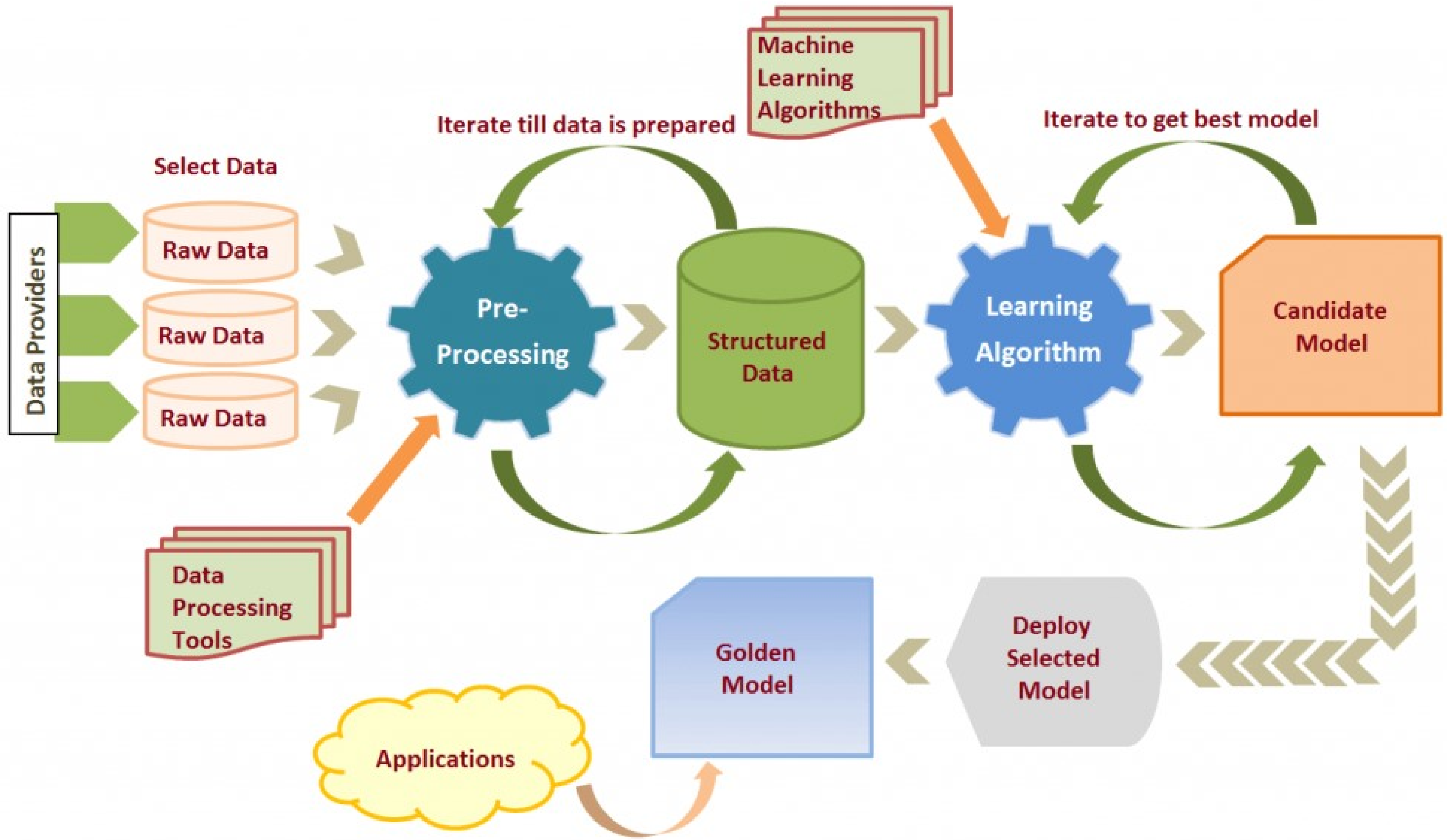
This is the basis of all machine learning algorithms.

Introductory concepts

There are many realities, many types of concepts, many performance functions, many target functions

Many techniques and methodologies

1.2 The process of machine learning



from Akhil Mittal , https://elearningindustry.com/machine-learning-process-and-scenarios

1 Raw data gathering

In ML, algorithms run through a big quantity of data (big data) to be trained and tested, before they are applied to new data in the problem domain (prediction, classification, decision).

The quantity and quality of the raw data, namely the good coverage of the problem domain determine the final quality of the ML system.

2 Raw data pre-processing

Preprocessing is needed to

- eliminate "abnormal" data, such as outliers,
- filtering to eliminate noise, for example,
- normalization of all dimensions of the data to obtain a mathematically well posed problem,
- other problem-specific needs, for example class balance.

3 Features extraction

Usually ML algorithms do not process raw data, but features.

Features are characteristics embedded (directly invisible) in the data that must be extracted.

Features are problem dependent. They may be in time domain (ex. statistical moments), frequency domain (ex. power spectrum parameters), time-frequency domain (ex. wavelet coefficients), or other inherent to the problem.
If they are in very high number, dimension reduction may be needed for computational reasons.

Structured data results from pre-processing and features extraction, eventually after an iterative process with feedback.

4 Select the ML model (algorithm)

As you will study in MEI, there are many algorithms (models) for ML.

There is no law to select one for a given problem. Depends on experience, trial and error, art and science.

Data can be classified in two main types: numerical and symbolic.

In this course of ML we will work with numerical data. Symbolic data will be studied in the Artificial Intelligence course.

Train and test a candidate model: analyze its performance and chose another if it is not good enough. Patience, resilience and enthusiasm are the keys for success in ML.

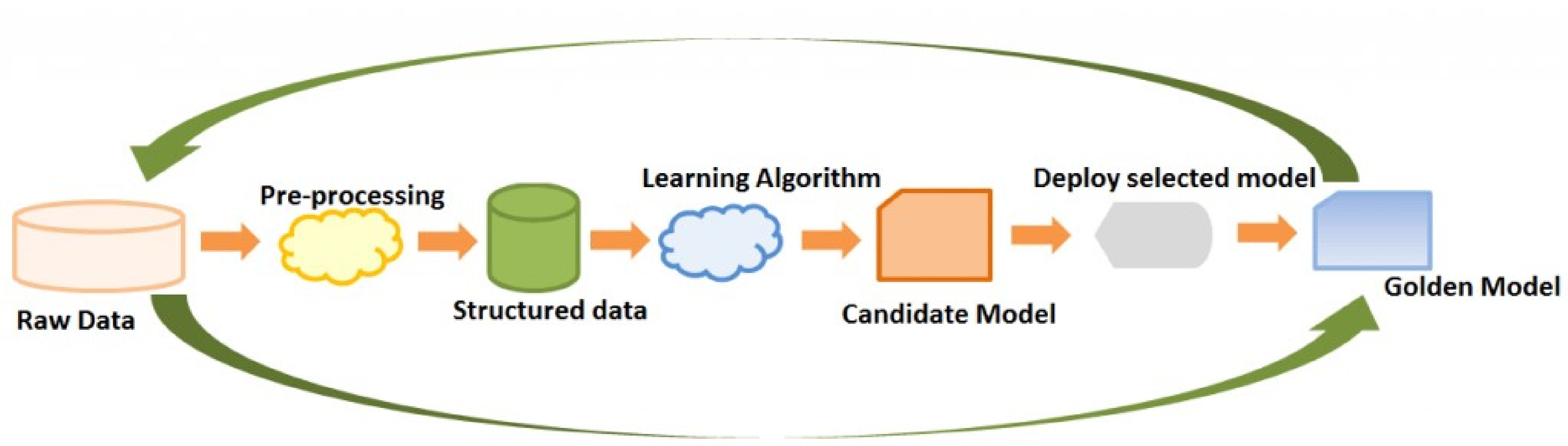
5 Apply the selected (gold) model to the problem

When (hopefully) a good model is found, it is put into "production", i.e., working with new, unseen, data.

Measure continuously its performance. If it degrades with time, retrain the ML model.

6 Retrain for dynamic changes or concept drift

Eventually (re)training may be a constant task, and all the process must be repeated frequently to catch the dynamic changes of the problem to which ML is being applied.



(https://elearningindustry.com/machine-learning-process-and-scenarios)

1.3. Hard computing and soft computing

The term "soft computing" is often used to name some ML models, in contrast with other possible models of "hard computing".

Development of an automatic pilot for an automobile

- Equations of mechanics
- Equations of fluid dynamics

to obtain a mathematical model of the car-road system

- The computer of the automatic pilot uses these equations to control and guide the car.

- **Which would be the complexity of the model ?**
- **Would it be possible to write exact equations for that ?**

..... "hard computing"

- Does the human pilot use such equations to drive ?
- How do you drive ? How did you learn ?
- Will it be possible to project an automatic pilot that in some way learns and behaves like an human driver ?

How to compute in a similar way to the human driver ??

Methodologies for machine learning capable of

- tolerate the imprecision, the uncertainty, the partial truth, the approximation

with the aim to obtain

- tractability, robustness, and computational solutions with low cost

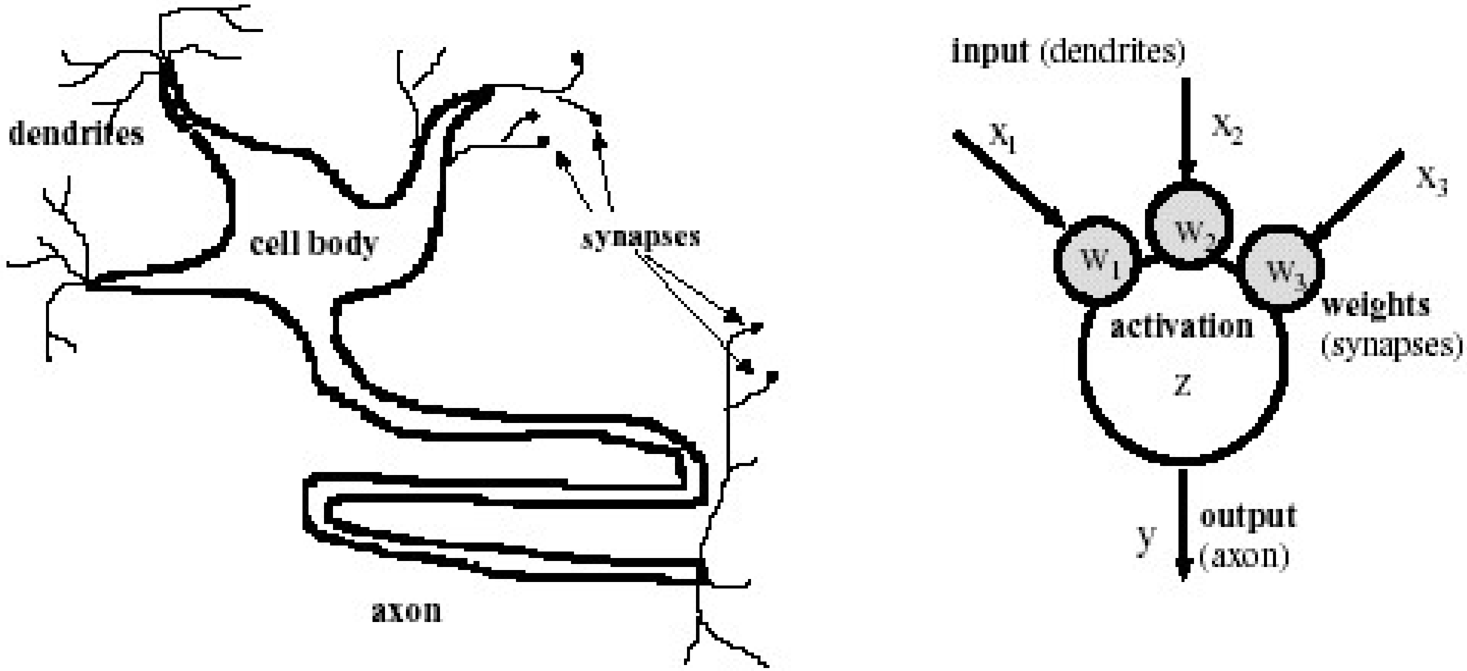
….. **"soft computing"**

...The role model for soft computing is the human mind ....

In A Definition of Soft Computing- adapted from L.A. Zadeh   http://www.soft-computing.de/def.html

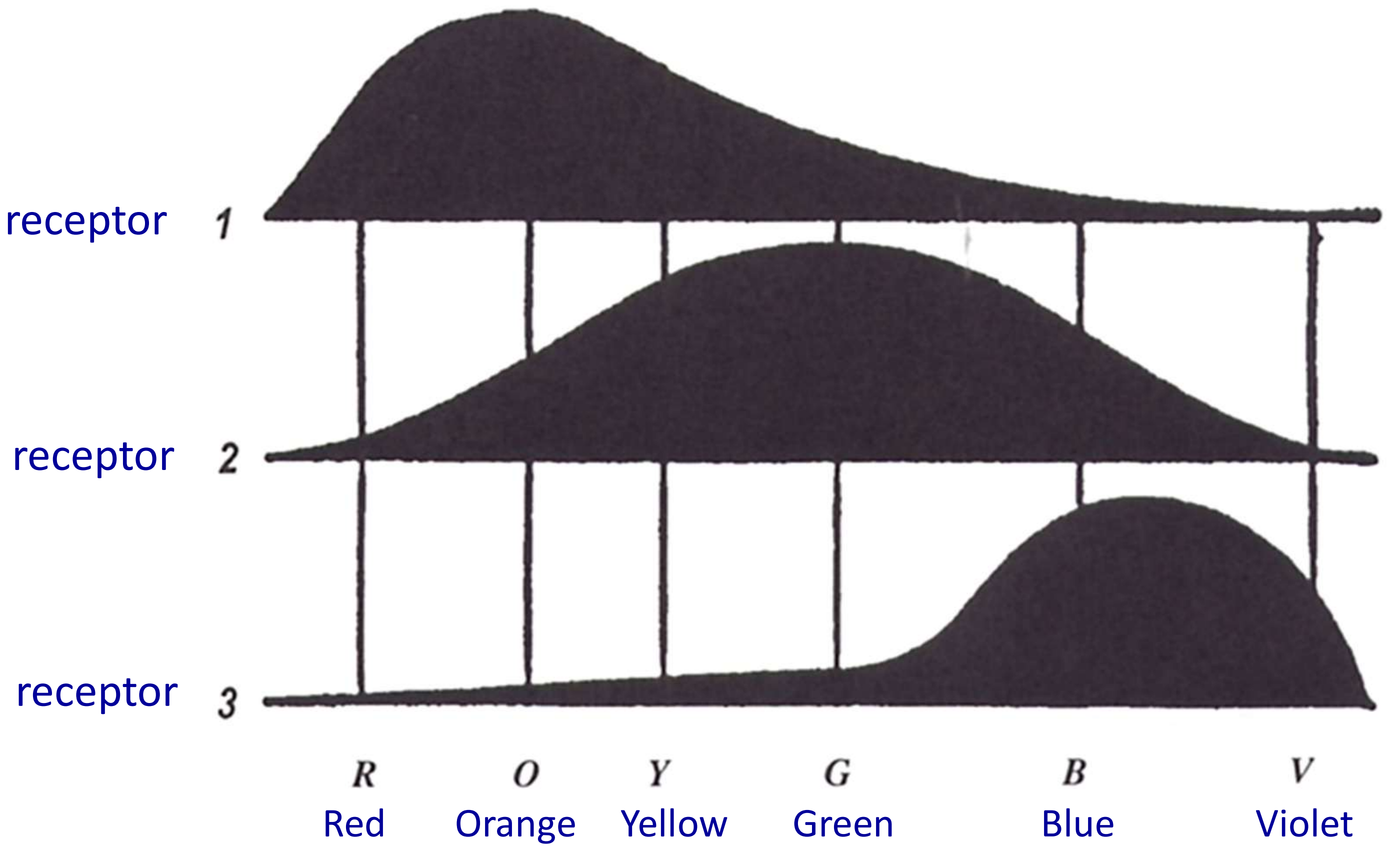
Professor Lofti Zadeh passed away on 7 September 2017

Biological Neurons



(from Brause, R., Medical Analysis and Diagnosis by Neural Networks, 2001)

Fuzzy brain: Young theory of visual receptors



Intensity of the activity of each receptor in each color (wavelength). Each receptor is maximally excited by one color, and progressively less by the adjacent colors.

(from Erikson, R., M.I. Chalaru, C.V. Buhusi, The Brain as a Fuzzy Machine: a Modelling Problem, in Theodorescu et al. (Editors), Fuzzy and Neuro-Fuzzy in Medicine, CRC Press, 1999)

Methodologies in this course of ML

Unsupervised learning:

- clustering

Given a dataset, find topographic structures embedded in it (that are not directly visible) based o the similarities and dissimilarities of the data points, using some measure.

The result is an organization of the data into clusters, that may be faced as classes in the problem domain.

Methodologies in this course of ML

Supervised learning:

The algorithms are trained by trying to reproduce known solutions for a set of instances of the problem - the training set.

Then they are tested in new data: the testing set.

In this course:

Decision trees

Artificial Neural Networks & Deep Learning

Fuzzy Logic and Fuzzy Systems

.... complementary, used in a combined way ...

Neuro-Fuzzy Systems:

A very used combination of neural networks and fuzzy logic in many applications

- Air conditioning, washing machines, photo and video machines, etc.
- In industry, namely car industry
- In services (bank, commerce, etc.))
- In internet
- In medical applications

... The employement of soft computing techniques leads to systems which have a high **MIQ** (**Machine Intelligence Quocient**). In large measure, it is the high MIQ of SC-based systems that accounts for the **rapid growth in the number and variety of applications of soft computing....**

In A Definition of Soft Computing- adapted from L.A. Zadeh https://www.soft-computing.de/def.html

Bibliography

Principal

Machine Learning, Tom Mitchell, McGraw-Hill, 1999

Neural Network Design, Hagan, Demuth and Beale, PWS Publishing*, 2nd ed. 2014,* free eBook at *https://hagan.okstate.edu/NNDesign.pdf 2 August2022*

Fuzzy Logic With Engineering Applications, Timothy Ross, 4th Ed., Wiley, 2016.

Deep Learning Toolbox Users´s Guide, The Mathworks, 2022

Fuzzy Logic Toolbox Users´s Guide, The Mathworks, 2022.

Complementary

Machine Learning, An Algorithm Perspective, Marsland, Stephen, CRC Press 2nd Ed 2014

Machine Learning, A Probabilistic Perspective, Murphy, Kevin P., MIT Press 2012.

Fundamentals of Artificial Neural Networks, Hassoun. M. H.,MIT Press, 1994.

Neural and Adaptive Systems, J.C. Príncipe, N.R. Euliano, W. C. Lefevre, Wiley, 2000

Introduction to Neuro-Fuzzy Systems, Robert Fullér, Springer Verlag 2000.

Bibliography

Complementary

Introduction to Neural Networks, Kevin Gurney, UCL Press, 1997.

*Neural Networks: A Comprehensive Foundation,*Simon Haykin,Prentice Hall,1999.

Introduction to Neural Networks, Kevin Gurney, UCL Press, 1997.

Fundamentals of Neural Networks, Laurene Fausette, Prentice Hall, 1994.

Fuzzy Modelling and Control, Andrzej Piegat, Springer Verlag, 2001.

Fuzzy Engineering, Bart Kosko, Prentice Hall, 1997.

Fuzzy Sets and Fuzzy Logic, Theory and Applications, George J. Klir and Bo Yuan, Prentice Hall, 1995.

Fuzzy Systems Theory and its Applications, T. Terano, K. Asai and M. Sugeno, Academic Press, 1987.

Fuzzy Logic, H. Zimmermann, Springer Verlag, 1997.

An Introduction to Fuzzy Sets Analysis and Design, Witold Pedrycz and Fernando Gomide, MIT Press, 1998.

Heaven in a Chip, Fuzzy Visions of Society and Science in the Digital Age, Bart Kosko, Three Rivers Press, NY, 1999