Machine Learning Introduction

Thomas Vuillaume, Paul Lagabbe Feb. 2021

table of contents

What is machine learning?

- “Learning is any process by which a system improves performance from experience.” - Herbert Simon

- Definition by Tom Mitchell (1998): Machine Learning is the study of algorithms that
  - improve their performance P
  - at some task T
  - with experience E

- A well-defined learning task is given by <P, T, E>

- Traditional programming:

DATA + PROGRAM → OUTPUT

MODEL

- Machine Learning:

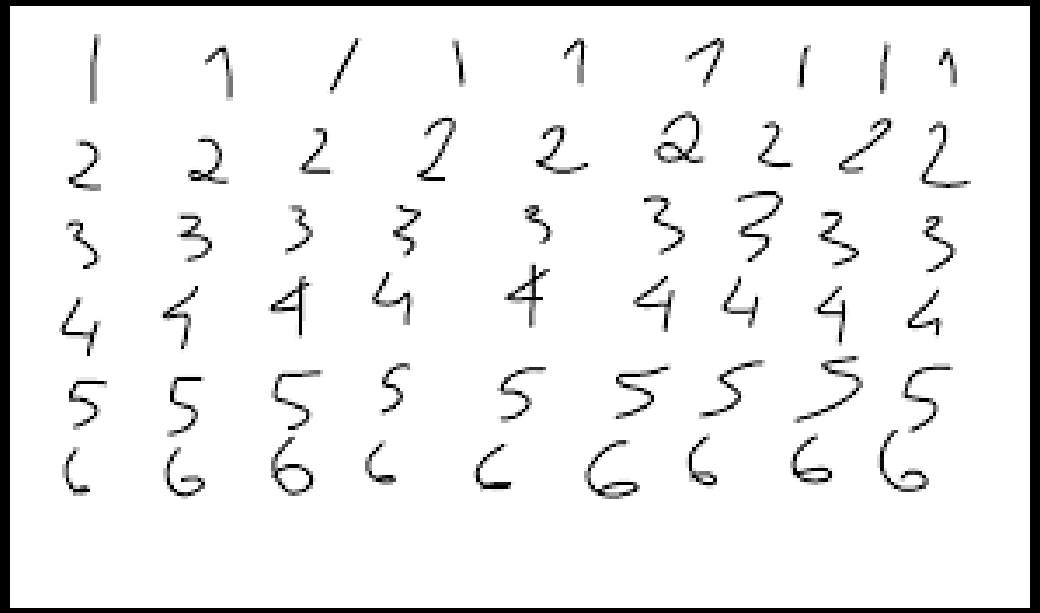
DATA + (OUTPUT) → PROGRAM

When is it useful ?

- We don’t know the underlying physical model
- We don’t know how to program a computer to do a task
- Models must be customized

But machine learning is not always useful. If you can program it or model it with a physical model, it will probably be the best approach.

A classic example: hand writing recognition



- What makes a « 2 » ?
- What is the difference with a « 1 » ?
- How do you program these differences ?
- How do you program to make the difference with huge variability ?

More examples

- Recognizing patterns:
  - Facial identities or facial expressions
  - Handwritten or spoken words
  - Medical images
- Generating patterns:
  - Generating images or motion sequences
- Recognizing anomalies:
  - Unusual credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
- Prediction:
  - Future stock prices or currency exchange rates

State of the art applications

- Autonomous cars
- Translation
- Speech recognition (automatic subtitles)
- Scene labeling
- Face generation
- ...

Data

Data are a collection of samples. Each sample has “features” and one “label”,
The goal of ML is to guess the label of the sample, thanks to his features.
We will call “X” the features matrix,
and “y” the label vector

	eventNumber	label	met_et	met_phi	lep_n	lep_pt_0	lep_pt_1	lep_eta_0	lep_eta_1	lep_phi_0	lep_phi_1	lep_E_0	lep_E_1
0	402756	1	25.609	0.42452	2	48.295	15.214	0.73991	2.27420	-2.316400	-1.39410	62129.0	74721.0
1	101274	0	196.560	1.31140	2	69.459	21.081	-0.52666	0.22380	0.023132	-0.67855	79317.0	21611.0
2	468437	1	45.653	-2.76860	2	45.927	22.822	-1.61910	-2.00770	1.901700	-0.11248	120480.0	86498.0
3	272337	1	49.415	-0.57805	2	45.929	14.263	1.55280	0.12809	2.254700	-2.54810	113360.0	14380.0
4	354546	1	71.988	-2.60390	2	62.029	21.453	-0.51082	-0.38177	0.236200	1.40820	70300.0	23036.0

samples labels Features space

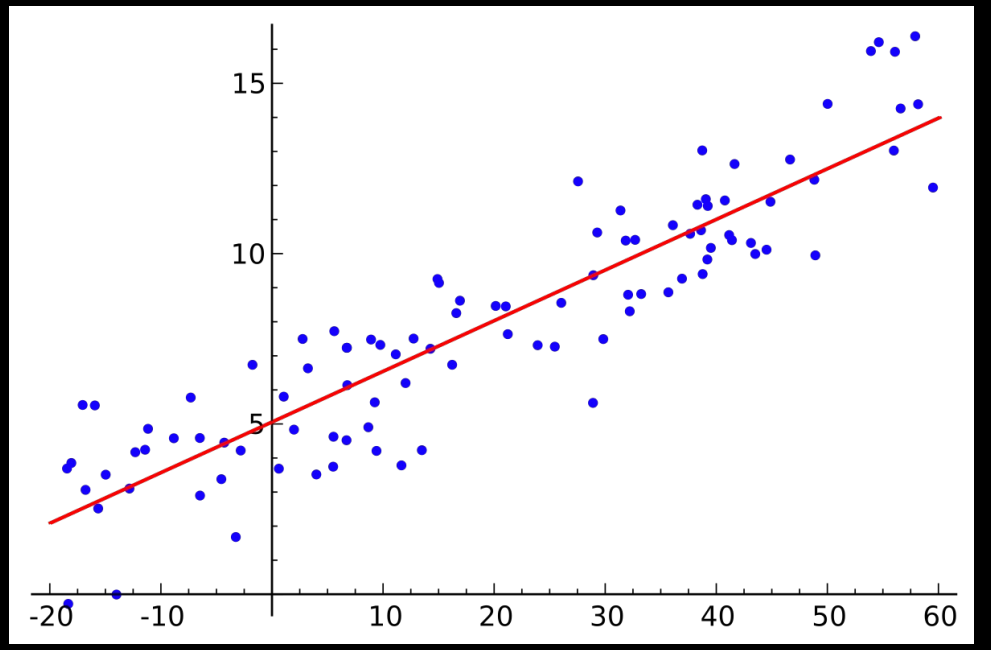
Types of learning

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Supervised Learning

- Regression: given X and y,
  find the continuous function f,
  so that f(X) = y

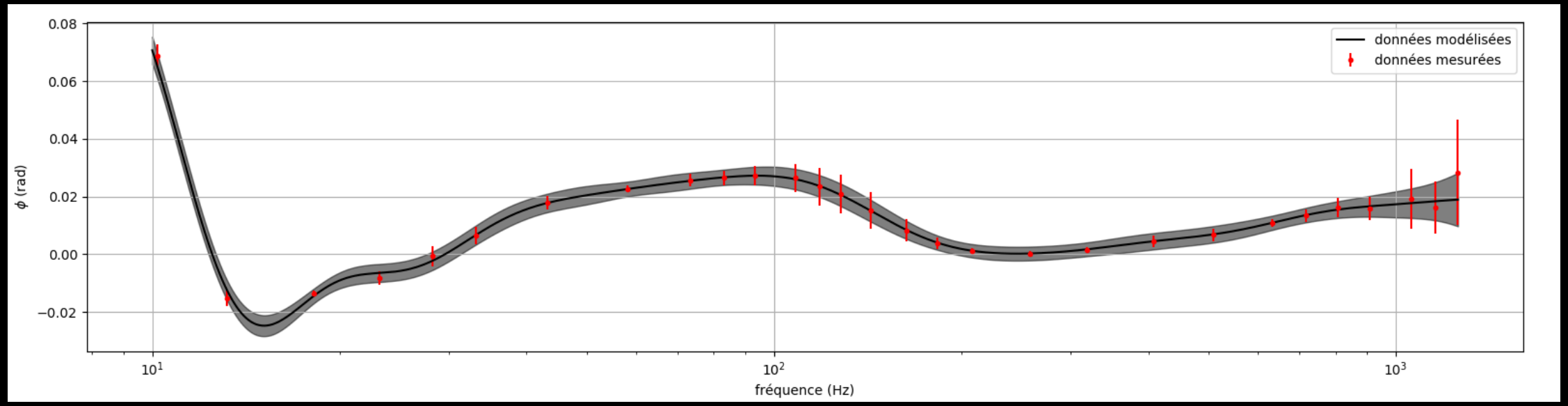
e.g.: linear regression : y = A.X + b



Supervised Learning

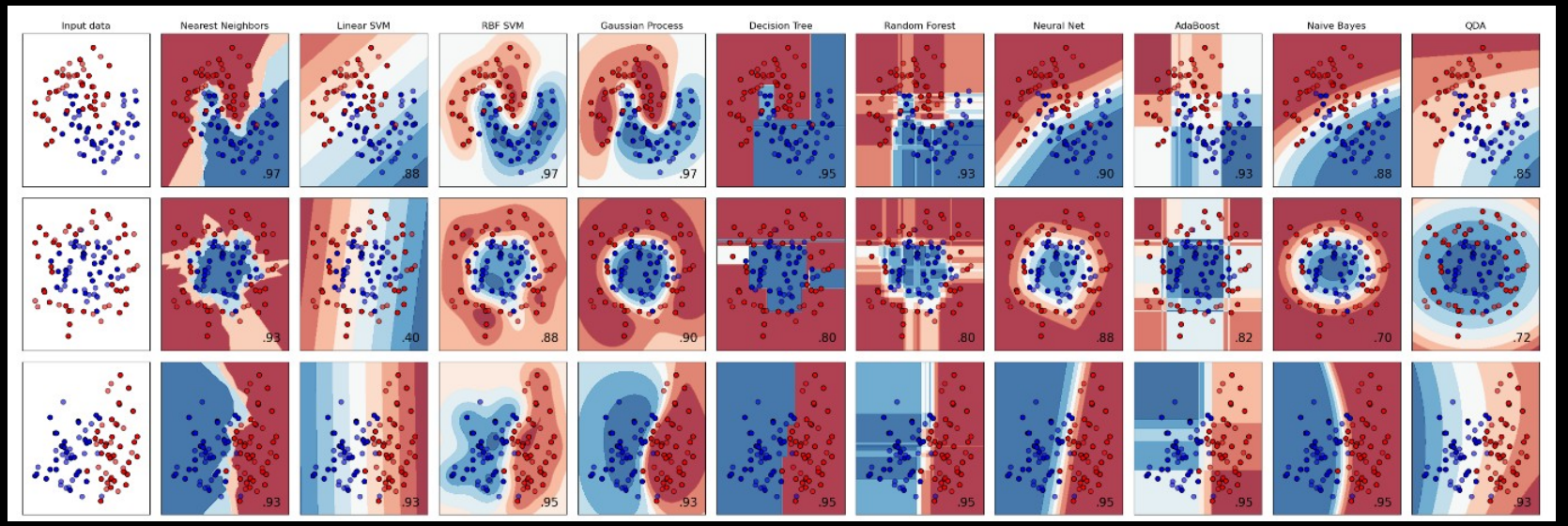
- Regression: given X and y, find the continuous function f, so that f(X) = y

More generally...



Supervised Learning

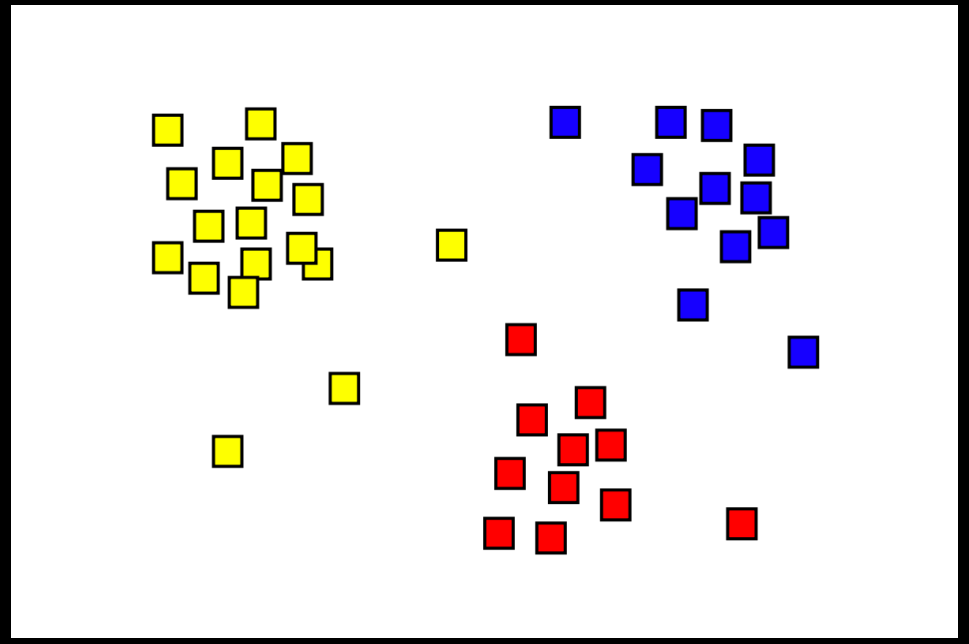
- Classification: given X and y,
  find the model to classify in a **discrete** number of categories
  e.g.: Given the position of the point X = (x1,x2), what is the color of the point ? y = red or blue ?



Unsupervised Learning

- Given X=(x1,x2,x3,x4,...) – find a structure behind the x's
  y labels (colors) are not given, but we have to find it

e.g.: clustering



Reinforcement learning

- Given a sequence of states and actions with (delayed) rewards, output a policy (= what to do at each state)

Basic example: learning a game, such as GO. There is a state (the board), some possible actions (dictated by the rules) and a the end a reward (win or loose).

See learning hide and seek: https://www.youtube.com/watch?v=kopoLzvh5jY

Machine learning is not new.

- Machine learning is actually quite old !
  - First developments in the 50’s and 60’s with Samual’s checker player
- Even neural network and deep learning basic ideas go back to the 60’s (perceptron)
- It is the aggregation of huge quantities of data in the latest years, combined with an increase of computing capacities that allowed it to boom

Disclaimer

At the end, machine learning algorithms are powerful statistical algorithms,
we still don’t know how to make « true » learning

References

- Eric Eaton, https://www.seas.upenn.edu/~cis519/fall2017/lectures/01_introduction.pdf
- David Kirkby, https://github.com/dkirkby/MachineLearningStatistics
- Wikipedia pages for diagrams
- Scikit-learn website : https://scikit-learn.org/stable/index.html