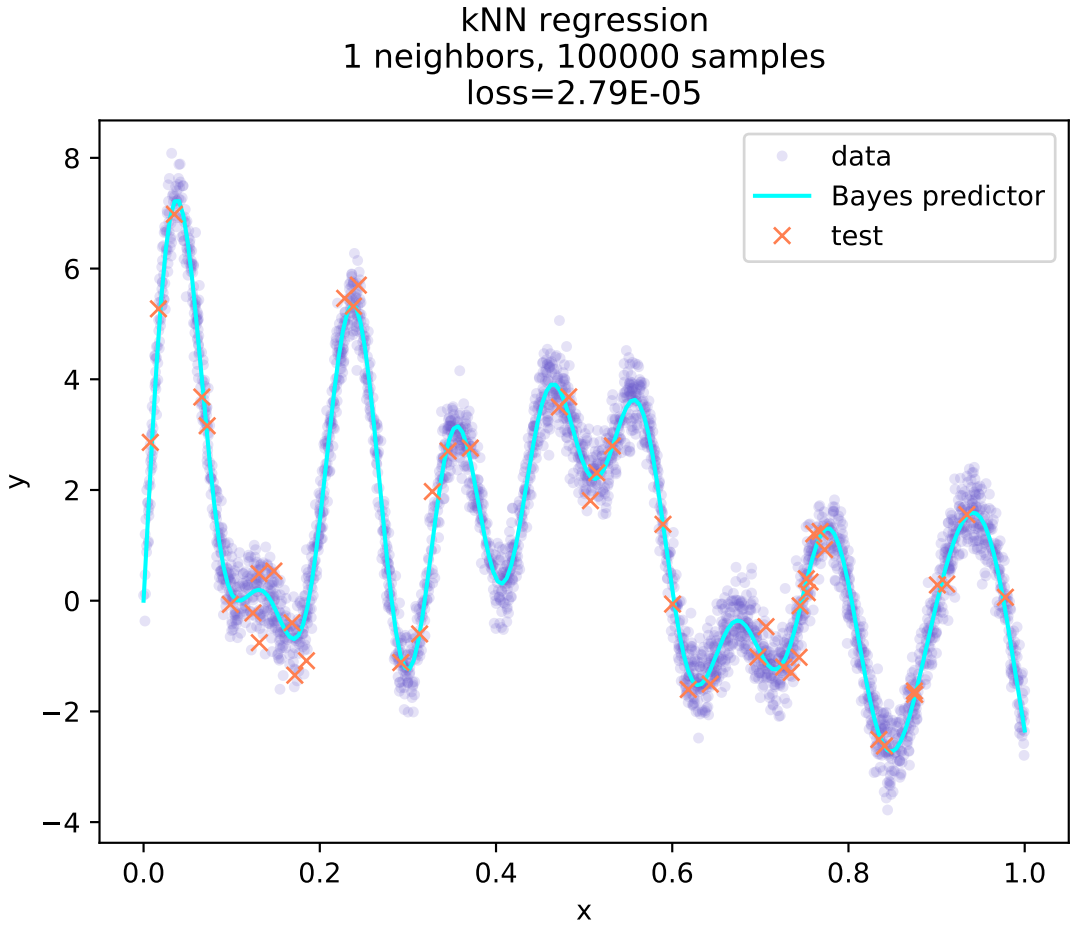
Machine learning II, unsupervised learning and agents: AI and ML



Definitions

What are AI and ML ?

Example 1 : digit classification

Figure – MNIST database [LeCun and Cortes, 2010]

Example 2 : robots

- Boston Dynamics robots (video)
- https://www.youtube.com/watch?v=LikxFZZO2sk
- https://www.youtube.com/watch?v=g0TaYhjpOfo
- https://www.youtube.com/watch?v=tF4DML7FIWk&t=57s
- https://www.youtube.com/watch?v=EezdinoG4mk

Example 3 : go player

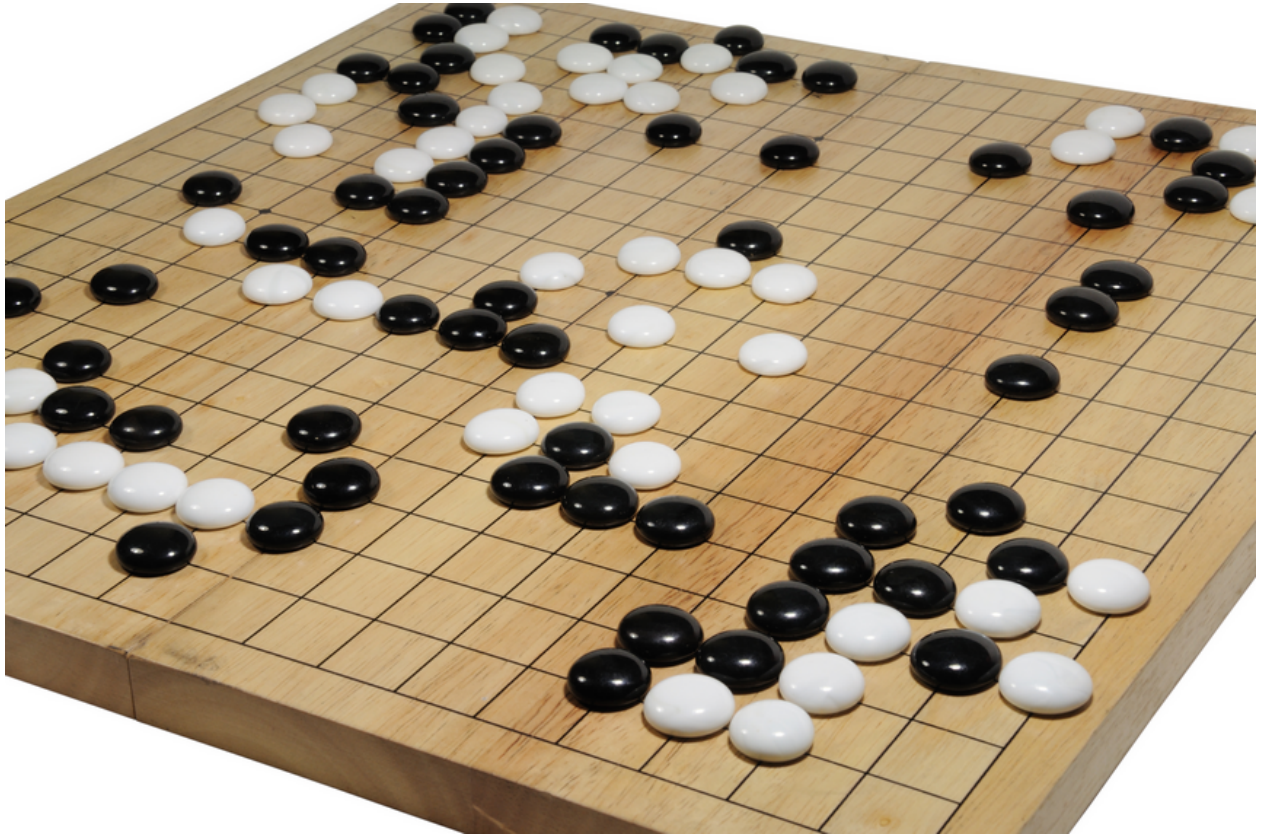


Figure – Go game, won by AlphaGo in 2017 [Silver et al., 2016]

Example 4 : audio engineering helper

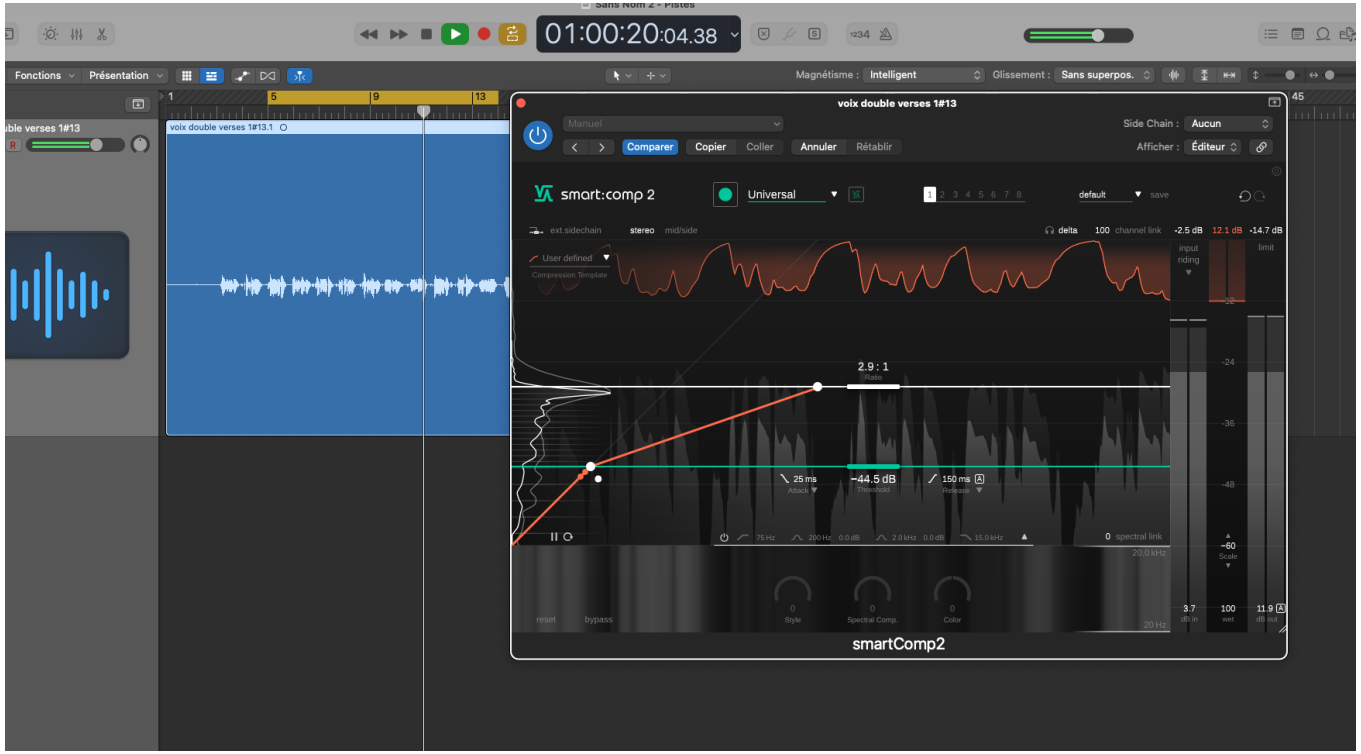


Figure – Semi-automatic compressor : https://www.sonible.com/smartcomp2/

Example 5 : coffee machine

Figure – Coffee machine (https ://www.stockresto.com/fr/machine-a-cafe/83-machine-a-cafe-conti-cc100-2-groupes.html)

Introduction examples

-
  - Boston Dynamics robot
  - MINST classification
  - AlphaGo
  - AI compressor
  - Coffee Machine
- All do different things but could be "reasonably" gathered under the term "AI" (or are branded as AI).

Definition

- Example definition of AI : "the theory and techniques aiming at emulating intelligence".
- Hence defining AI requires definining "intelligence".

Turing test

- It is reasonable to say that the problem of AI was born at the same time as that of **Computer Science**. One of the founders of Computer Science is Alan Turing (1912-1954).
- In 1950, with the article **"Computing machinery and intelligence"** [Turing, 2009], Turing introduces the **Turing test**.
- One of the forms of a Turing Test is a game in which a computer tries to behave like a human by answering questions.

Turing, A.M. (1950). Computing machinery and intelligence. Mind, 59, 433-460.

COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in

Turing test

- Importantly, Turing stresses that the terms "think", and "machines" are far from being **unambiguous**.
- Many (most ?) questions regarding our own "intelligence" and "consciousness" as humans remain open today.
- In that respect, it is tricky to give a indisputable definition of AI.

Moravec Paradox

- It is harder to emulate or simulate simple sensorimotor capacities than level abstract reasoning.
- ie : the sensorimotor system of an ant is harder for us to emulate than beating the world champion at chess (Deep Blue, 1997).

Strong AI, Weak AI

- There exists a distinction between **Weak AI**, and a **hypothetical** AI called **Strong AI**.

Strong AI, Weak AI

- There exists a distinction between **Weak AI**, and a **hypothetical** AI called **Strong AI**.
- **Weak AI** (also known as **narrow AI**) designed and trained for a particular task (e.g. Siri).
- **Strong AI** (also known as Artificial General Intelligence AGI) : able to find a solution, faced with an **unfamiliar task**. This is **hypothetical**, it does not exist yet, and may never exist.

Conscious reasoning

Broadly speaking, the first phases of research in AI were more dedicated to emulating the **conscious brain.**

- logical programming (Prolog)
- expert systems

History : AI Winter (1970's)

- In 1974 fundings for AI reseach started being cut by goverments because of unsuccessful projects.
- there were actually several so-called "AI winters".

Famous AI : Deep Blue (1997)

- Beat the world chess champion Gary Kasparov in 1997.
-

Figure – Deep Blue

- Despite its name, Deep Blue is **not related** to Deep learning !

Decision trees

One of the underlying mechanisms of Deep Blue were **Decision Trees** / minimax algorithm.

2000's and 2010's

- Around the 2000's, there has been a shift in the focus of AI.
- Machine learning and data science have become prominent research and industrial topics.
- We could argue that emulating the **unconscious brain** is more of a focus today (as opposed to formal, conscious reasoning).

2000's and 2010's

- Around the 2000's, there has been a shift in the focus of AI.
- Machine learning and data science have become prominent research and industrial topics.
- We could argue that emulating the **unconscious brain** is more of a focus today.
- An important contemporary and future trend in research might be to connect research on emulating the conscious brain and research emulating the unconscious brain.

Machine learning

- **Proposed definition :** "artificial intelligence which can learn and model some phenomena without being explicitely programmed, based on a dataset of samples".
- In Machine learning, some parameters are learned in an *semi-automatic* (programmed) way in order to solve a problem or to optimize a solution.

Machine learning

Machine learning has received attention and funding because it has reached state-of-the-art efficiency on several problems, such as :

- computer vision
- spam classification
- machine translation
- speech recognition
- self-driving cars

Machine learning

Machine learning has received attention and funding because it has reached state-of-the-art efficiency on several problems such as :

- computer vision
- spam classification
- machine translation
- speech recognition
- self-driving cars

Deep learning is involved in several of these setups.
However, machine learning is not all about deep learning.

Example 1 : ImageNet

- A database of images (more than 15M), hand-annotated in order to indicate what objects are present in the image. More than 20000 categories of images.
- Contest : ImageNet Large Scale Visual Recognition Challenge.
- The best classification (top 5 error rate) went from 25% in 2011 to $\simeq$ 15.3% in 2012.

Example 1 : ImageNet

- A database of images (more than 15M), hand-annotate in order to indicate what objects are present in the image. More than 20000 categories of images.
- Contest : ImageNet Large Scale Visual Recognition Challenge.
- The best classification (top-5 error rate) went from 25% in 2011 to $\simeq$ 15.3% in 2012.
- The technology used was deep learning, exploiting GPUs (AlexNet).

`https://paperswithcode.com/sota/image-classification-on-imagenet`

Example 2 : AlphaGo

- In 2015 : beats a professionnal player. In 2017 : beats the world champion.
- Uses several technologies : among them **Deep reinforcement learning.**
- Improvements : AlphaGo Zero, trained without a database of played games. In 20217, AlphaZero beats AlphaGo Zero after 3 days of learning.

Example 3 : AlphaFold

- Goal : to predict the spatial configuration of proteins, from their DNA sequence.
- Achieves a breakthgough performance on the CASP challenge :
  - 2018 : more than 50% GDT (Global distance test), whereas it was $\leq 40\%$ before then.
  - 2020 : $92,4\%$ GDT. At a $\geq 90\%$ score, the method is considered competitive with experimental methods..
- `https://alphafold.ebi.ac.uk/`
- Also based on Deep learning.

AI and Machine Learning

In the four first examples, according to you which one are NOT a Machine Learning system ?

AI and Machine Learning

- Alpha Go : Machine Learning (Reinforcement Learning)

AI and Machine Learning

- Alpha Go : Machine Learning (Reinforcement Learning)
- MNIST : Machine Learning (Supervised Learning)

AI and Machine Learning

- Alpha Go : Machine Learning (Reinforcement Learning)
- MNIST : Machine Learning (Supervised Learning)
- Coffee Machine : no Machine Learning (Automation)
- Compressor : (probably) supervised learning.

AI and Machine Learning

- Alpha Go : Machine Learning (Reinforcement Learning)
- MNIST : Machine Learning (Supervised Learning)
- Coffee Machine : no Machine Learning (Automation)
- Boston Dynamics : no Machine Learning (Robotics)

AI and Machine Learning

- Alpha Go : Machine Learning (Reinforcement Learning)
- MNIST : Machine Learning (Supervised Learning)
- Coffee Machine : no Machine Learning (Automation)
- Boston Dynamics : no Machine Learning (Robotics)
- **edit :** while the robots on the video have no Machine Learning, Boston Dynamics seems to now show interest in the topic. ML technologies might start being implemented in their robots, especially for artificial vision.

AI and scientific domains

- People doing "AI" today can actually come from rather different scientific communities :
  - Statistics and probabilities
  - Optimization
  - Other mathematic fields : graph theory, combinatorics
  - Statistical physics
  - Robotics
  - Cognitive sciences / neuroscience / psychology
- For instance, *data science* is mostly a mix between statistics, optimization, graph theory

ML revolution

- Technical progress in the computing and storage capacities
- Increase in amount of available data. According to IBM, 10^{18} bytes are created each day.
- Progress in algorithmic methods to analyze the data.

Old AI's

- The first industrial robots were introduced in the 1950's. Unimation Inc introduced robotic arms in the US.
- Dendral (1960's) : Assisstant for chemists in order to classify unknown molecules.

Old AI : Deep Blue (1997)

- Beat the world chess champion Gary Kasparov in 1997.
-

Figure – Deep Blue

Younger AIs

- Siri, Google Translate, DeepL : **Natural Language Processing**
- Ross (2010's)"Artificial Attorney" : helps lawyers and legal assisstant to research the law. `https://blog.rossintelligence.com/`

AI in today's industry

- The most prominent usage of AI in today's industry, besides automated processes, is **data mining** : the analysis of data through semi automated algorithms in order to allow some industries to make strategies decisions.
- Data mining is also present in some scientific research areas such as astrophysics, and biophysics.
- A number of companies also explore ML algorithms (e.g. for computer vision) for medical applications.

Introduction to AI and ML

Learning paradigms

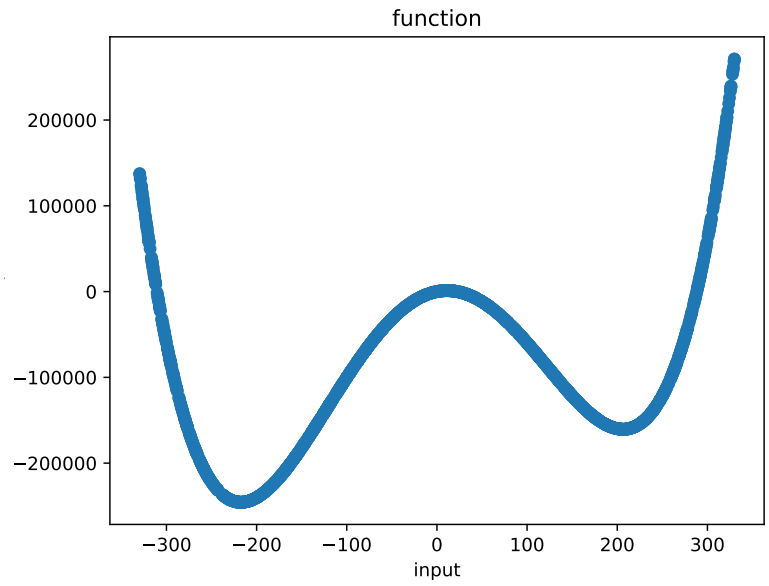
Classical difficulties of machine learning

Machine Learning and Optimization

- We will present the three main **machine learning paradigms**.
- We said earlier that in Machine Learning, some parameters are learned in an semi-automatic way to solve a problem.
- But how are the parameters learned ?

Notion of optimization

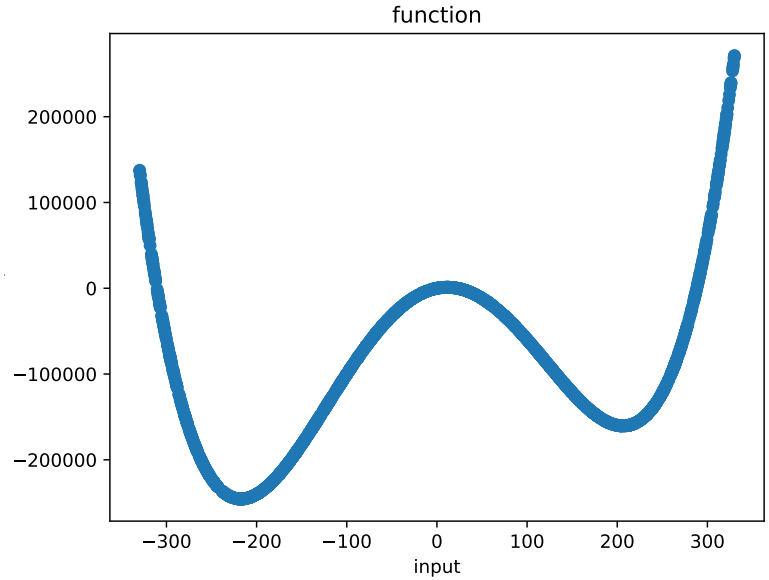
The **quality / relevance** of a set of parameters must be measured with a real number. This measure helps us in finding the best parameters possible.



Notion of optimization

Often, the **quality** of a set of parameters can be measured with a real number. This measure helps us in finding the best parameters possible.

The question is then the choice of this measure, often called a **objective function**, **loss function**, or **cost function**.



Supervised Learning

- For a certain input x, we want to **predict** an output y : for instance,
  - x : contains the age, and the height of a person, so here x is a **vector** containing **two features**.
  - y : best record on a 100 meters track
- To do so, we learn from a number of **labeled examples** (x_i, y_i) track time)

Supervised learning

We assume that there exists a relationship between the inputs x and the output y, given by some function f.

$$y = f(x) \tag{1}$$

f is most of the time a function that contains some **randomness**.

- We learn from a number of **labeled examples** (x_i, y_i)
- In the case where what we want to predict is a **class**, it is a **classification problem** : $y \in \mathbb{N}$. (example : MNIST)
- In the case where what we want to predict is a **real number** (float), it is a **regression problem** : $y \in \mathbb{R}$.
- **Objective :** find a good estimation $\tilde{f}$, of f.

Choices and modelisation

- What subset of functions should $\tilde{f}$ belong to ?
- What **loss function** whould we use ?

Lost function

- What **loss function** whould we use ?
- It should be a measure of the discrepancy between our prediction and the correct label.

Loss function

- What **loss function** whould we use ?
- It should be a measure of the discrepancy between our prediction and the correct label.
- Let us take the example of a regression problem, when the label is a real number.

Loss function

- What **loss function** whould we use ?
- It should be a measure of the discrepancy between our prediction and the correct label.
- Let us take the example of a regression problem, when the label is a real number.
- For an individual sample, a possible discrepancy is the squared loss

$$(\tilde{f}(x_i) - y_i)^2 \tag{2}$$

Loss function

- Taking into account the whole dataset, the **loss function** writes :

$$\sum_{i=1}^{n} (\tilde{f}(x_i) - y_i)^2 \tag{3}$$

- Several other loss functions are possible :

$$\sum_{i=1}^{n} |\tilde{f}(x_i) - y_i| \tag{4}$$

Loss function

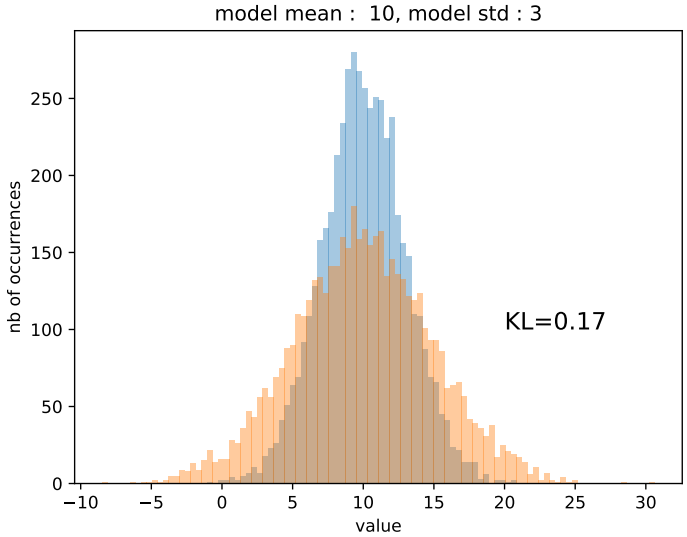
- The loss function is a **real number** measuring the relevance of a **collection** of parameters.
- The number of parameters depends on the situation, and varies between 1 (e.g. for a simple linear model) and millions (e.g. for deep neural networks).

Unsupervised Learning

- From a number of samples x_i, you want to retrieve information on their structure : **modelisation**.
-
  - clustering (this afternoon)
  - density estimation (tomorrow)
  - dimensionality reduction (tomorrow)

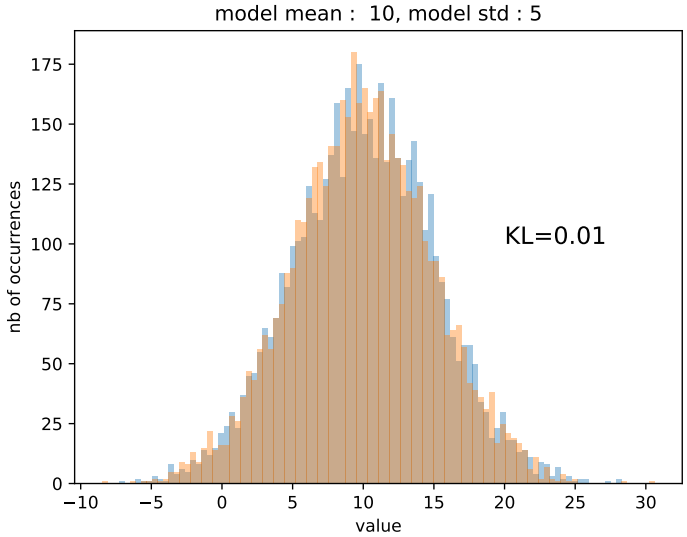
Unsupervised Learning

- From a large number of samples x_i, you want to retrieve information on their structure : **modelisation**.
- density estimation :



Unsupervised Learning

- From a large number of samples x_i, you want to retrieve information on their **structure**
- density estimation :



-

Unsupervised Learning

- From a large number of samples x_i, you want to retrieve information on their **structure**
- density estimation.
- Question : how do you constrain your distribution to fit your data ?
  - parametric models
  - non-parametric models

Reinforcement Learning

- A **more general paradigm** that describes an **agent** in a **world**.
- The standard formalization was the one proposed by Richard Sutton [Andrew, 1998] (pdf available online).

Reinforcement Learning

- A **more general paradigm** that describes an **agent** in a **world**.
- The standard formalization was the one proposed by Richard Sutton [Andrew, 1998]
- At each time, the world is in a state s. An agent perfoms an **action** a according to a **policy** π. When performing an action, the agent receives an **reward** r.

Reinforcement Learning applications

- Example : some Chessplayers, AlphaGo, automatic vacuum cleaner, self-driving cars (trajectory optimization, parking, etc.)
- healthcare : `https://arxiv.org/pdf/1908.08796.pdf`
- `https://deepmind.com/blog/article/alphago-zero-starting-scratch`

Reinforcement Learning

- At each time, the world is in a state s. An agent perfoms an **action** a according to a **policy** π. When performing an action, the agent receives an **reward** r.
- The agent wants to learn an **optimal policy**, meaning the policy that maximises its reward.

This paradigm has many variants

- State s, action a, policy π, reward r.
- Is the policy **deterministic** ? Is it **stochastic** ?
- Does the agent have a **model** of its environment ?
- How many steps ahead whould the agent look ?

This paradigm has many variants

- State s, action a, policy π, reward r.
- Is the policy **deterministic** ? Is it **stochastic** ?
- Does the agent have a **model** of its environment ?
- How many steps ahead should the agent look ?
- All these conditions change the way the problem should be addressed and solved. The **Bellman equations** rule the updates of the optimal policy.

Application of the paradigms

Let's discuss the introductory examples in terms of learning paradigms.

Example 1 : web browser and page ranking

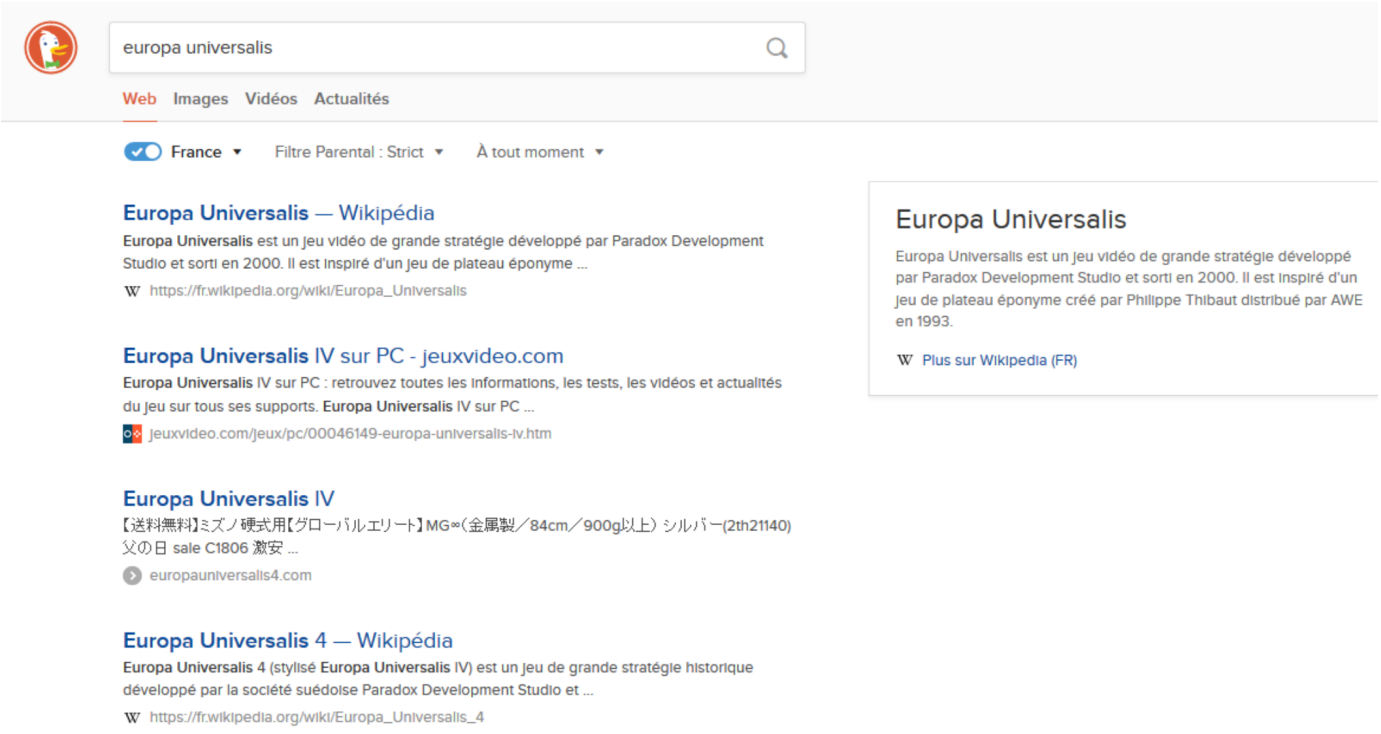


Figure – Duckduckgo web browser

Example 1 : web browser and page ranking

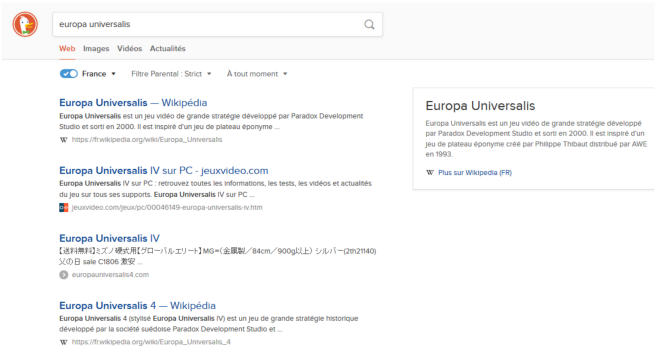


Figure – Duckduckgo web browser

- User i types a query $\tilde{x}_i$, and finally chooses a web page $\tilde{y}_i$.
- The web browser wants to propose the most relevant output
- **Supervised learning**

Example 2 : targetted ads

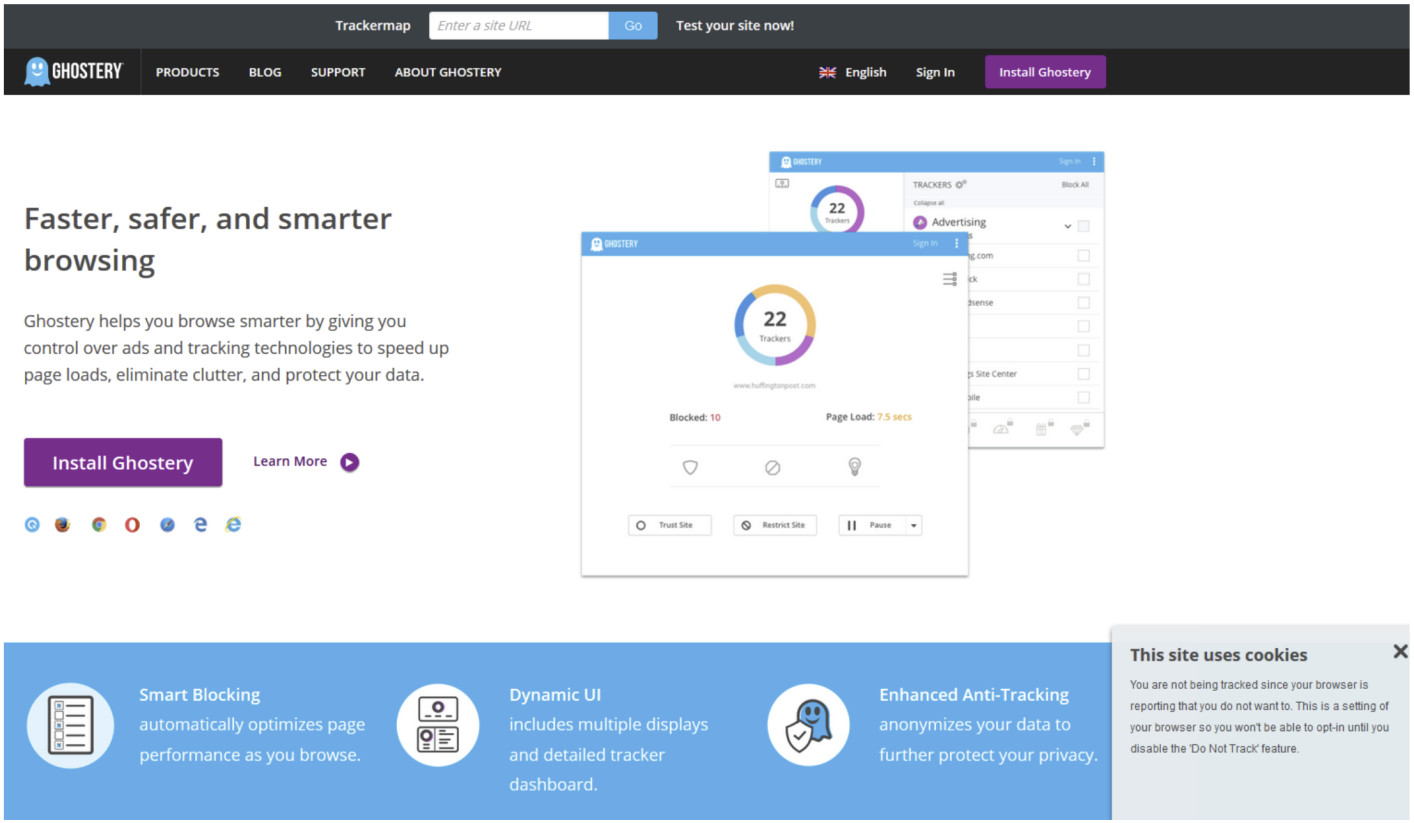


Figure – Ghostery

Example 2 : targetted ads

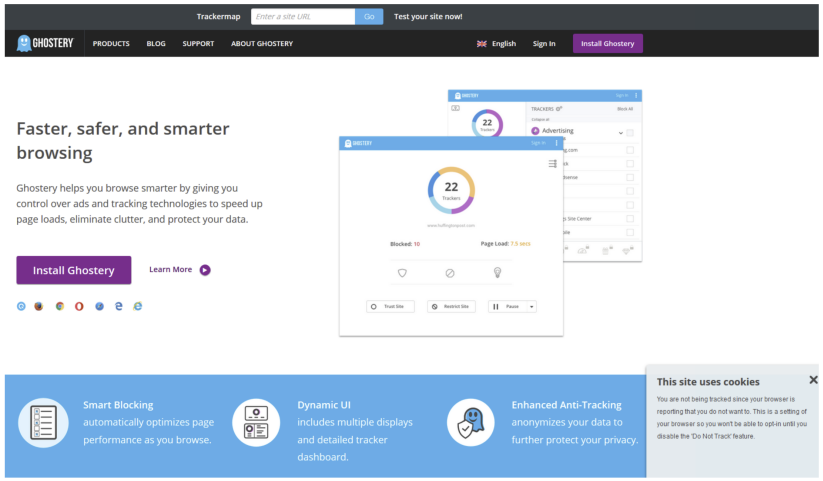


Figure – Ghostery

- Propose products (output) to a user (input)
- **Supervised**

Example 4 : Image segmentation

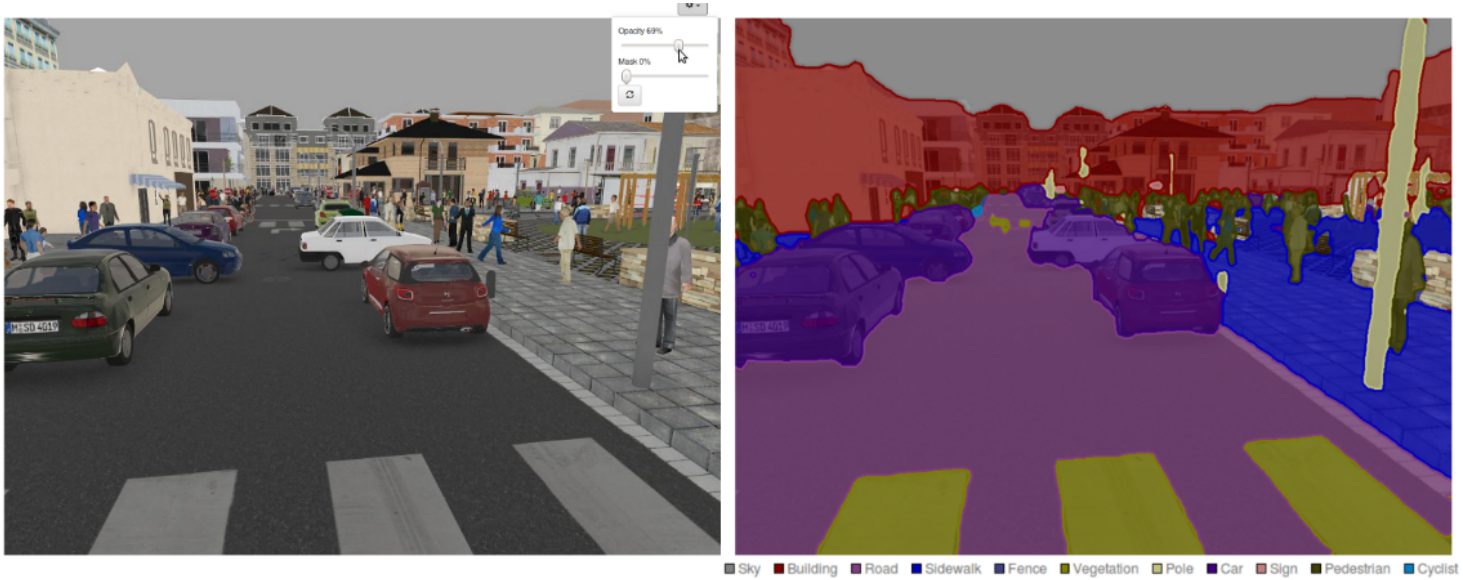


Figure – Source : Nvidia dev blog

Example 4 : Image segmentation

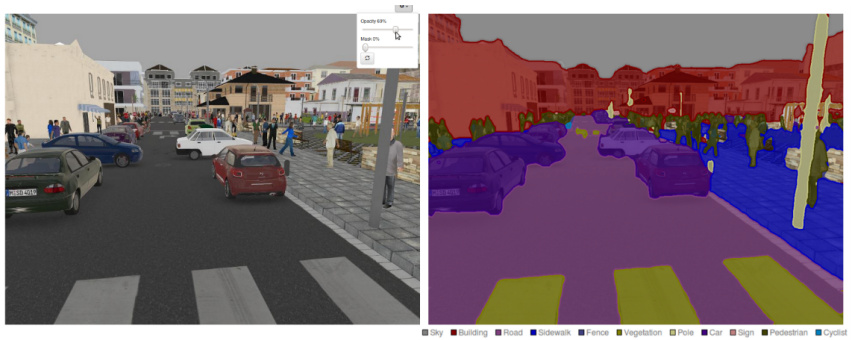


Figure – Source : Nvidia dev blog

- Often people give labeled examples of segmented images
- **Supervised**

Example 4 : Image segmentation

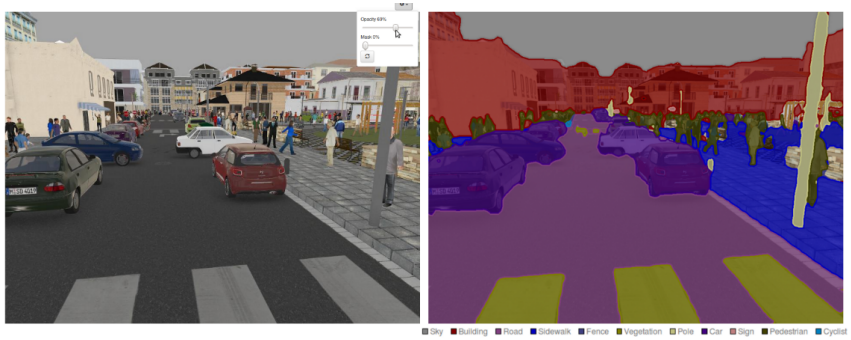


Figure – Source : Nvidia dev blog

- One could also group pixels together based on their properties
- **Unsupervised** (but most of the time the supervised method is used)

Example 5 : Digit classification

Figure – MNIST dataset [LeCun and Cortes, 2010]

Example 5 : Digit classification

Figure – MNIST dataset [LeCun and Cortes, 2010]

- **input** : pixels
- **output** : digit (class)
- **Supervised**

Example 3 : Text Translation

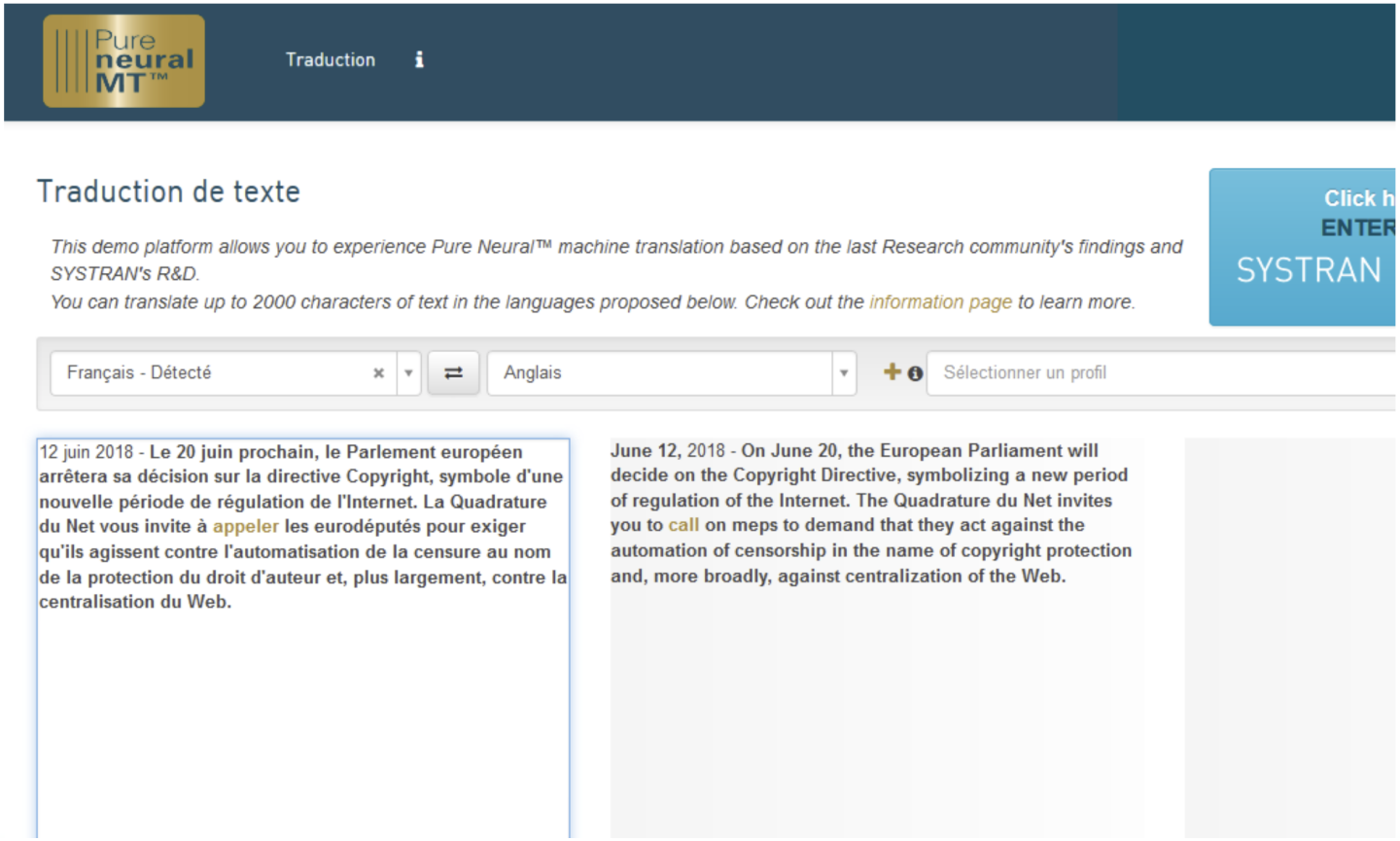


Figure – Text translation

Example 3 : Text Translation

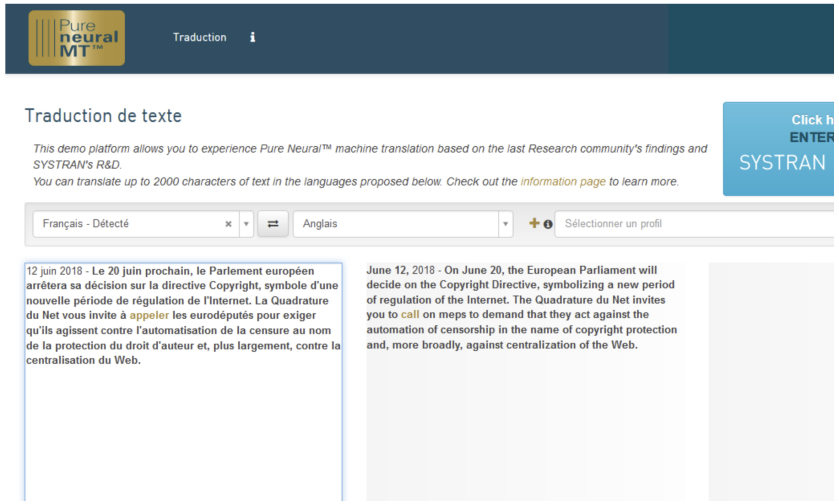


Figure – Text translation

- Based on translated examples
- **Supervised**

Example 7 : Time Series Analysis

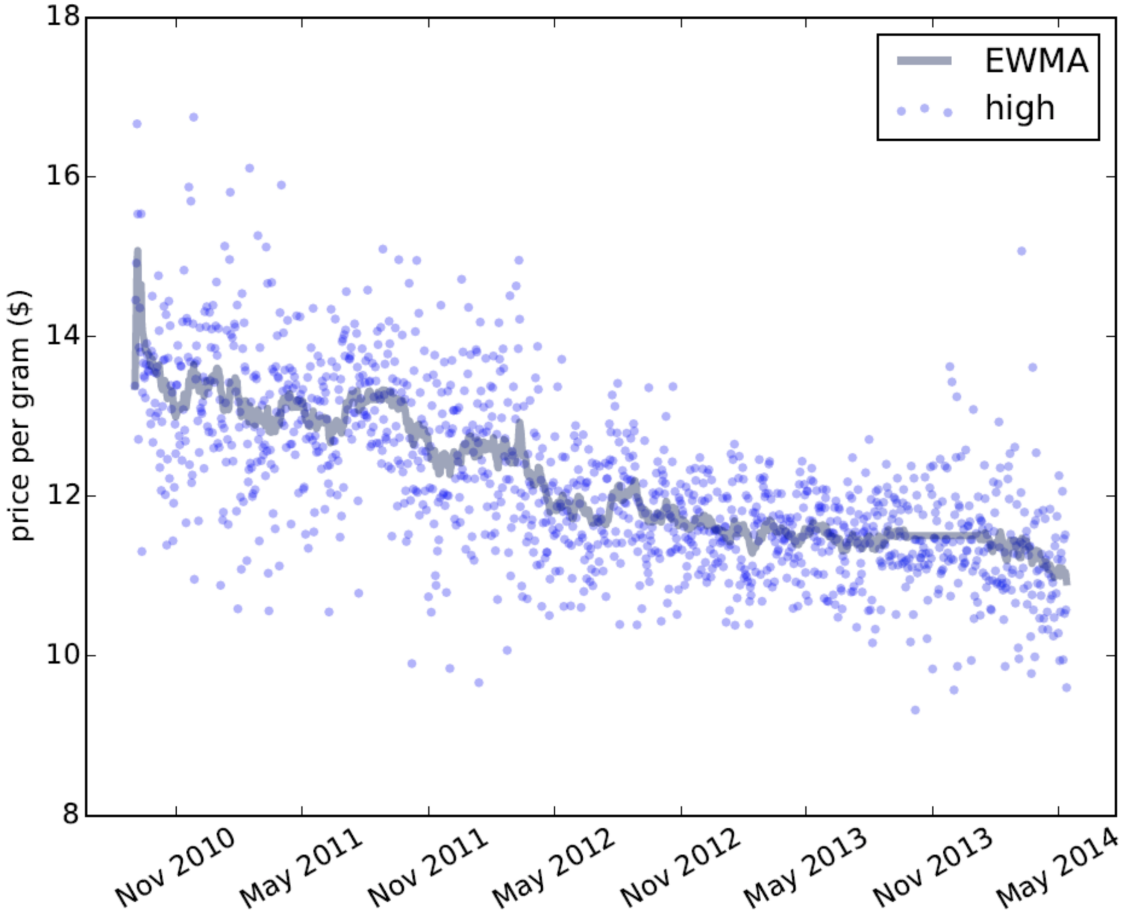


Figure – Price of some plant

Example 7 : Time Series Analysis

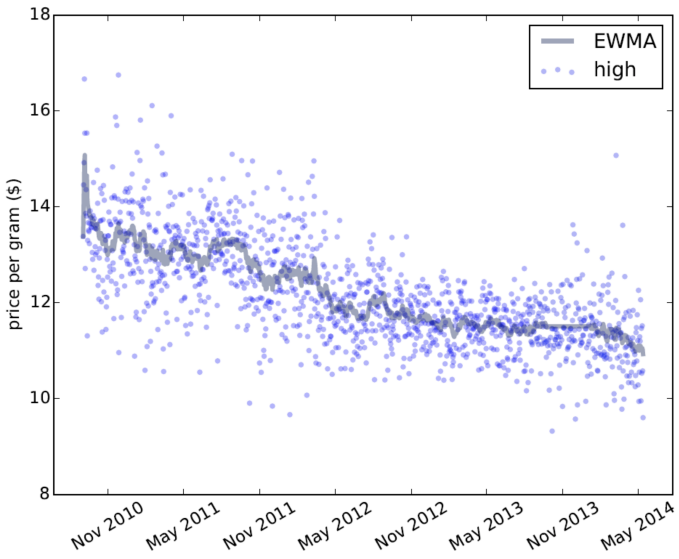


Figure – Price of some plant

- **Supervised**

Example 8 : Clustering

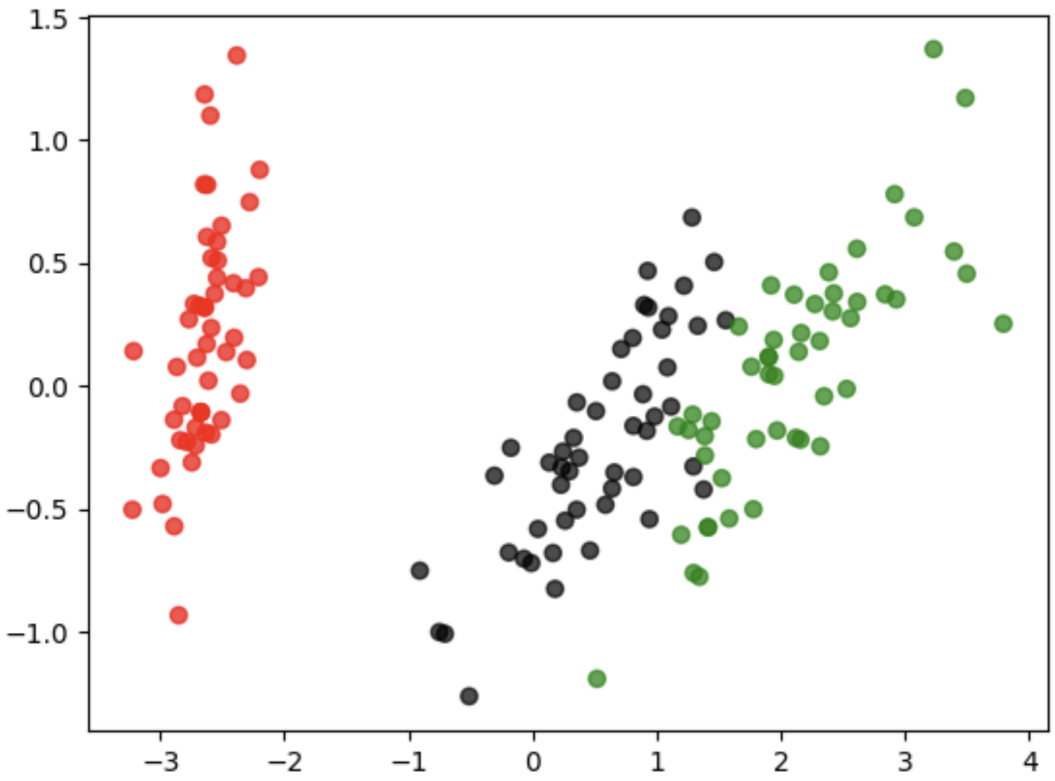


Figure – Iris dataset

Example 8 : Clustering

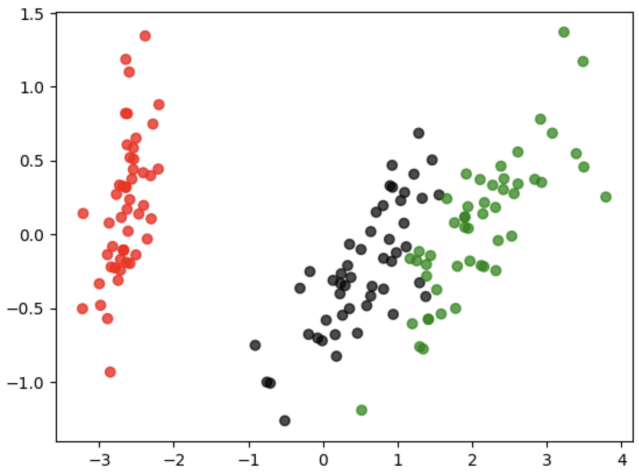


Figure – Iris dataset

- Grouping together points as a function of their features
- **Unsupervised learning**

Example 9 : game AI

Figure – Europa Universalis

Example 9 : game AI

Figure – Europa Universalis

- Designing an **agent** performing **actions**
- **Reinforcement learning**

Conclusion

Sometimes there is an ambiguity between RL, supervised and unsupervised.

- unsupervised learning can be used in a supervised learning problem (semi supervised learning)
- unsupervised learning and supervised learning can be used in a reinforcement learning problem

Introduction to AI and ML

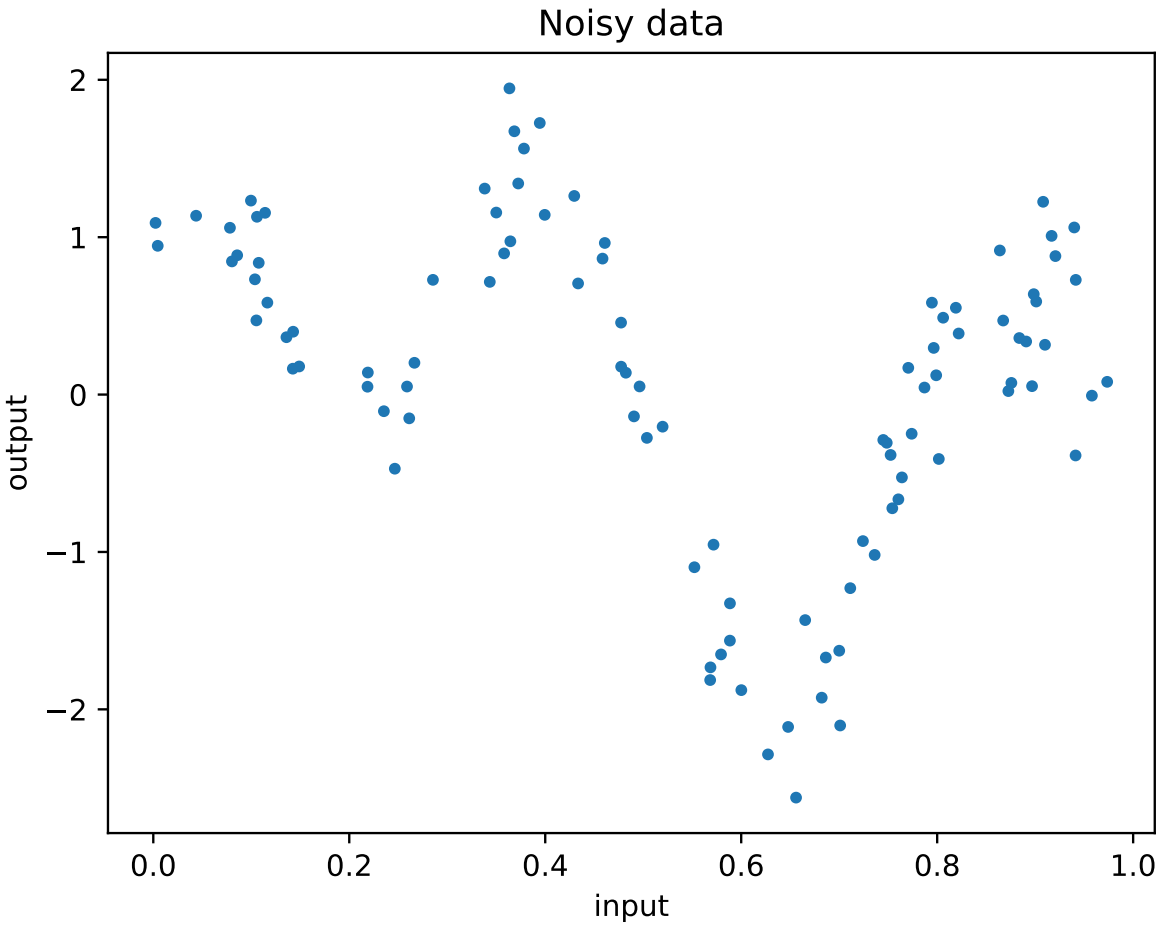
Learning paradigms

Classical difficulties of machine learning

Overfitting

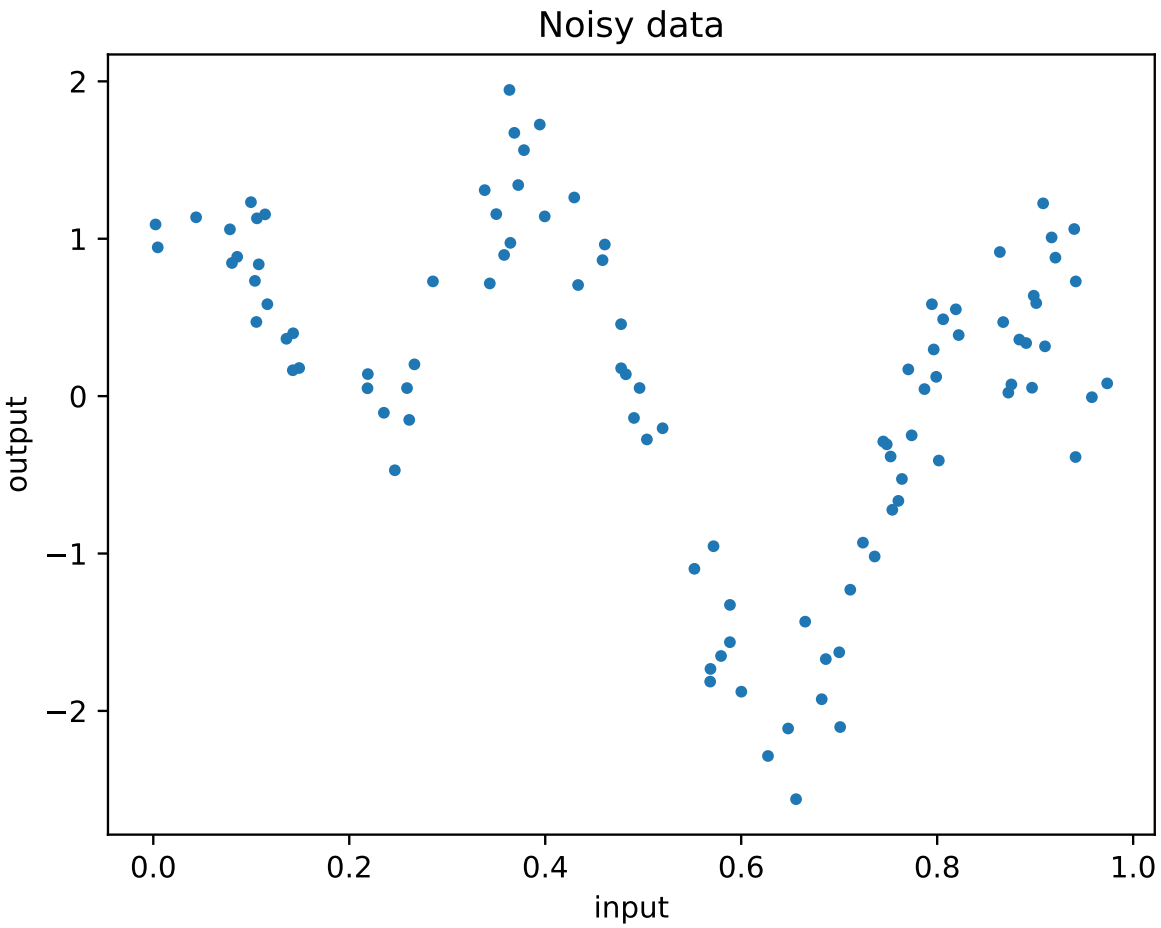
- Overfitting is a statistical problem that comes from the fact that we learn based on a finite dataset.
- We do not know if thie finite dataset contains enough statistical information about the physical process that generated it.

Polynomial regression

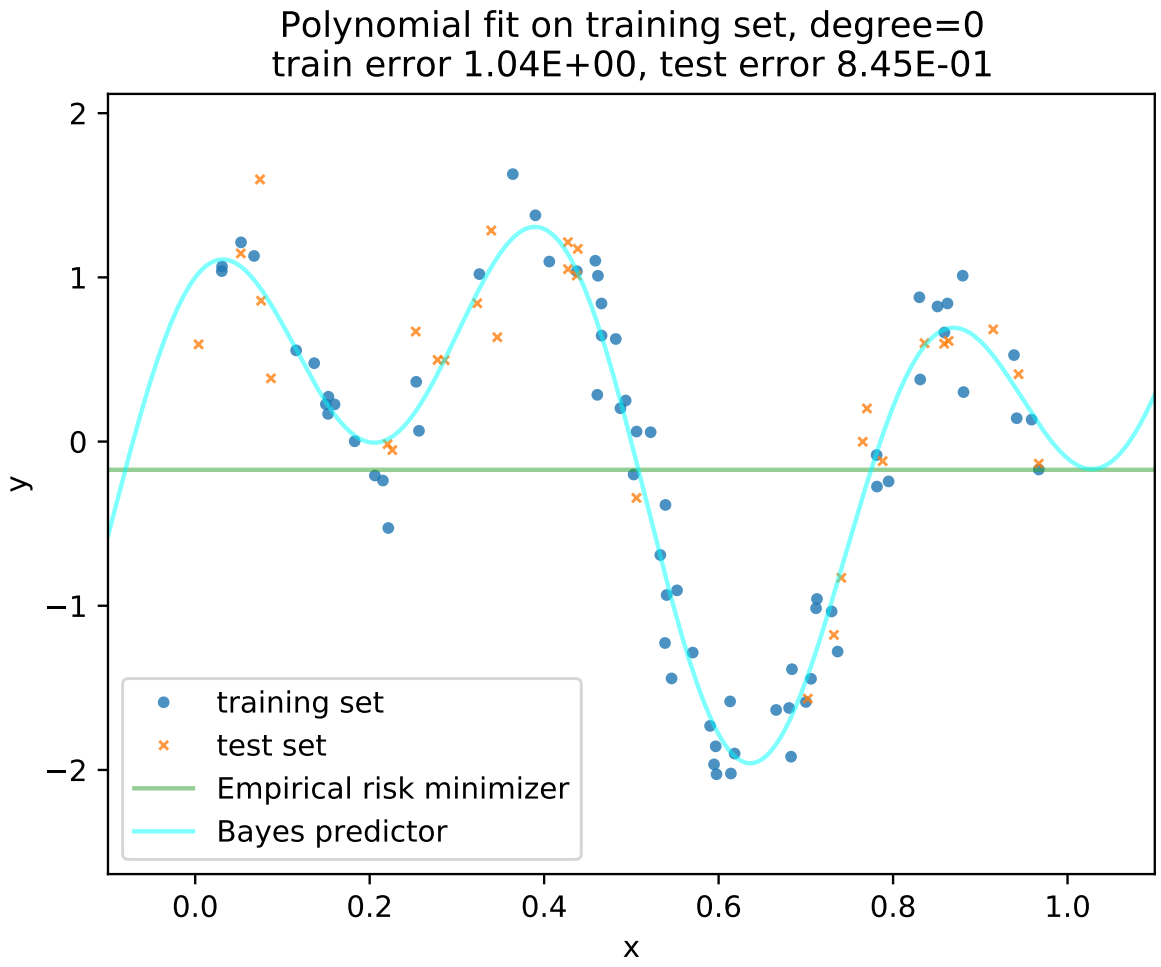


Polynomial regression

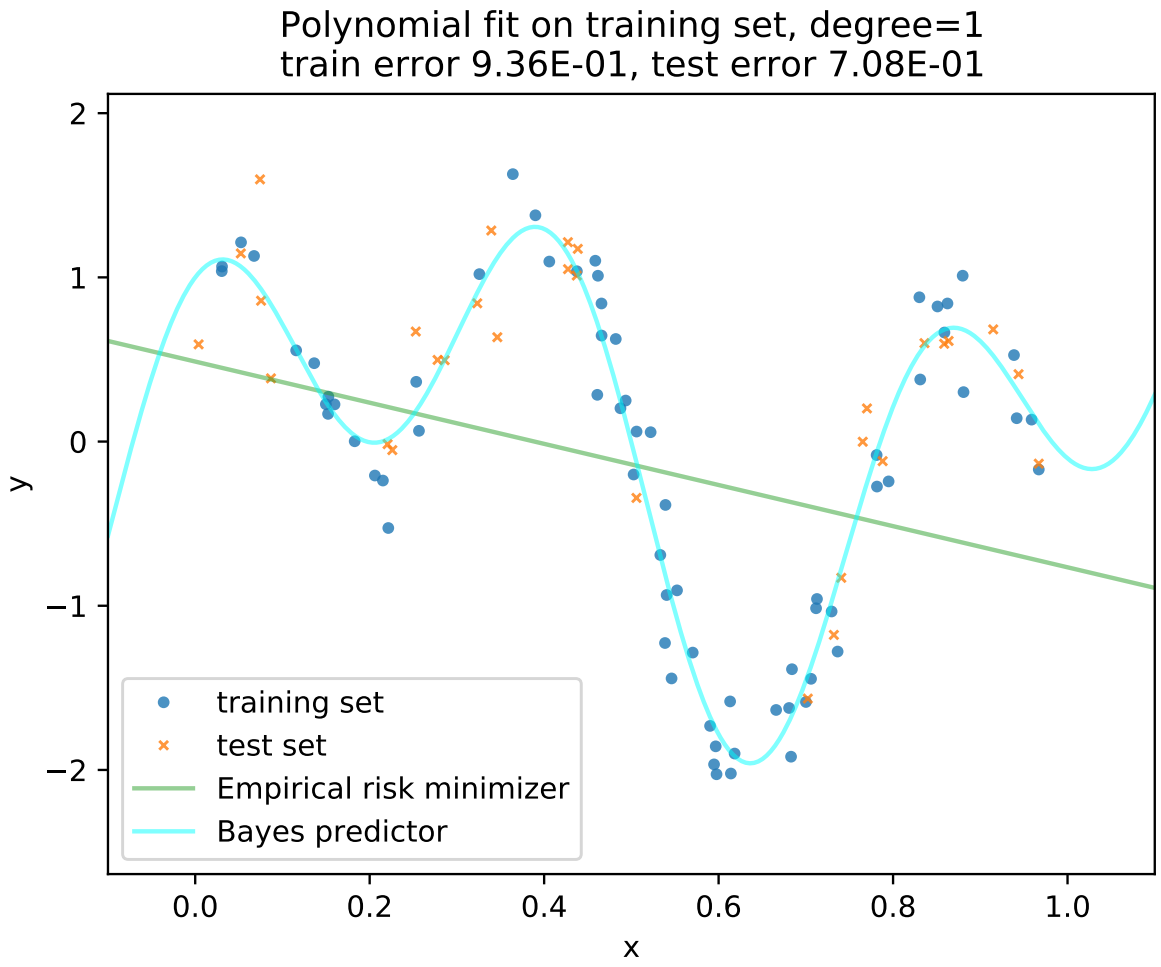
Exercice 1 : is there a polynom that interpolates the data ?



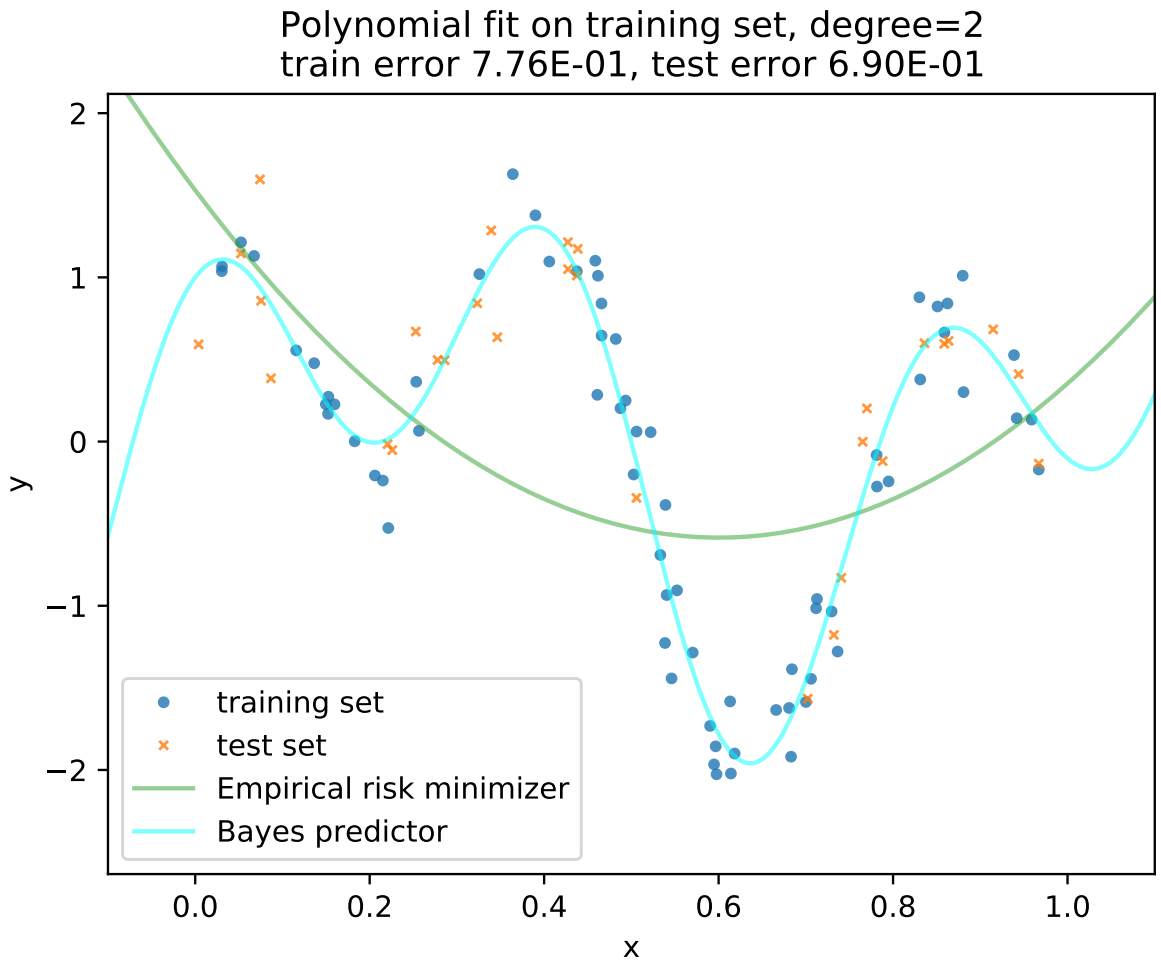
Polynomial regression



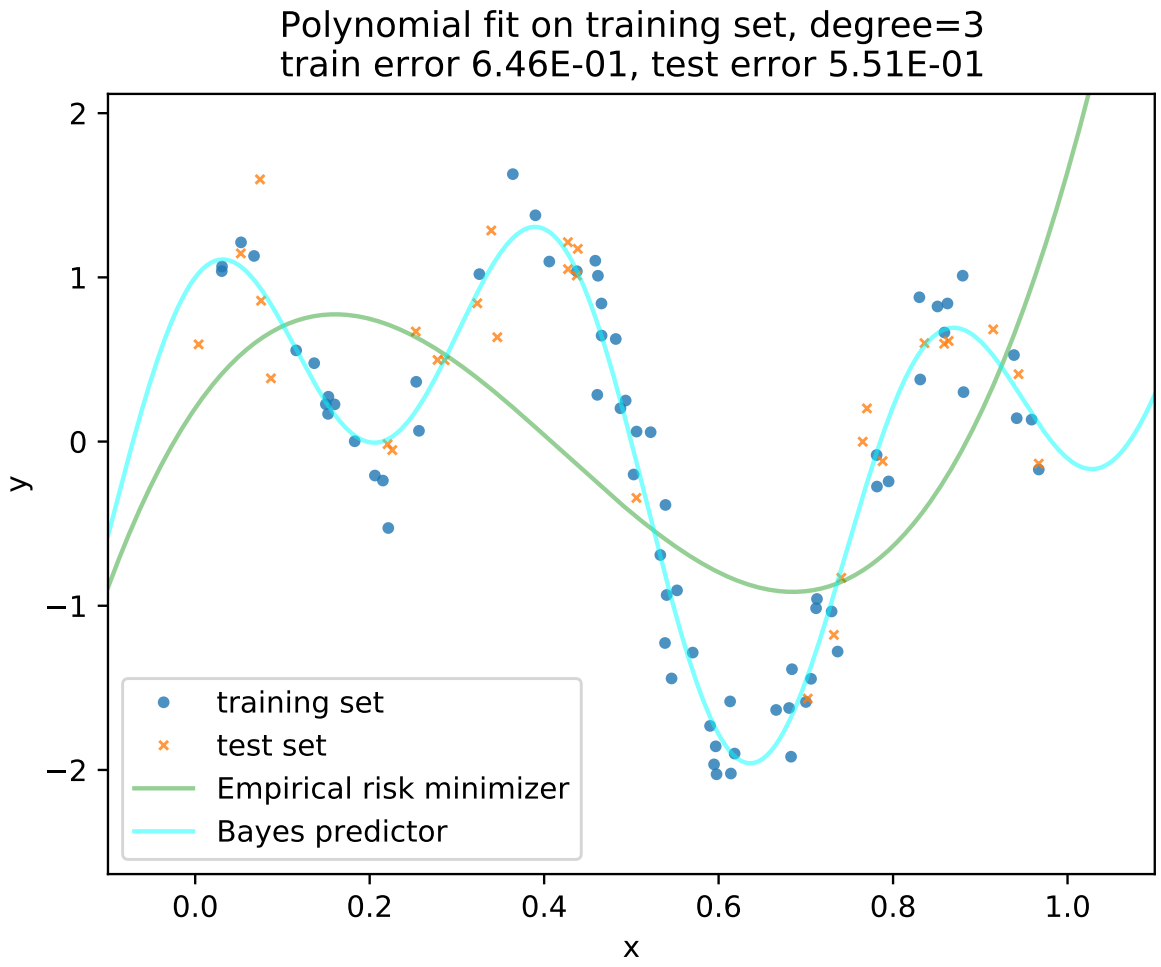
Polynomial regression



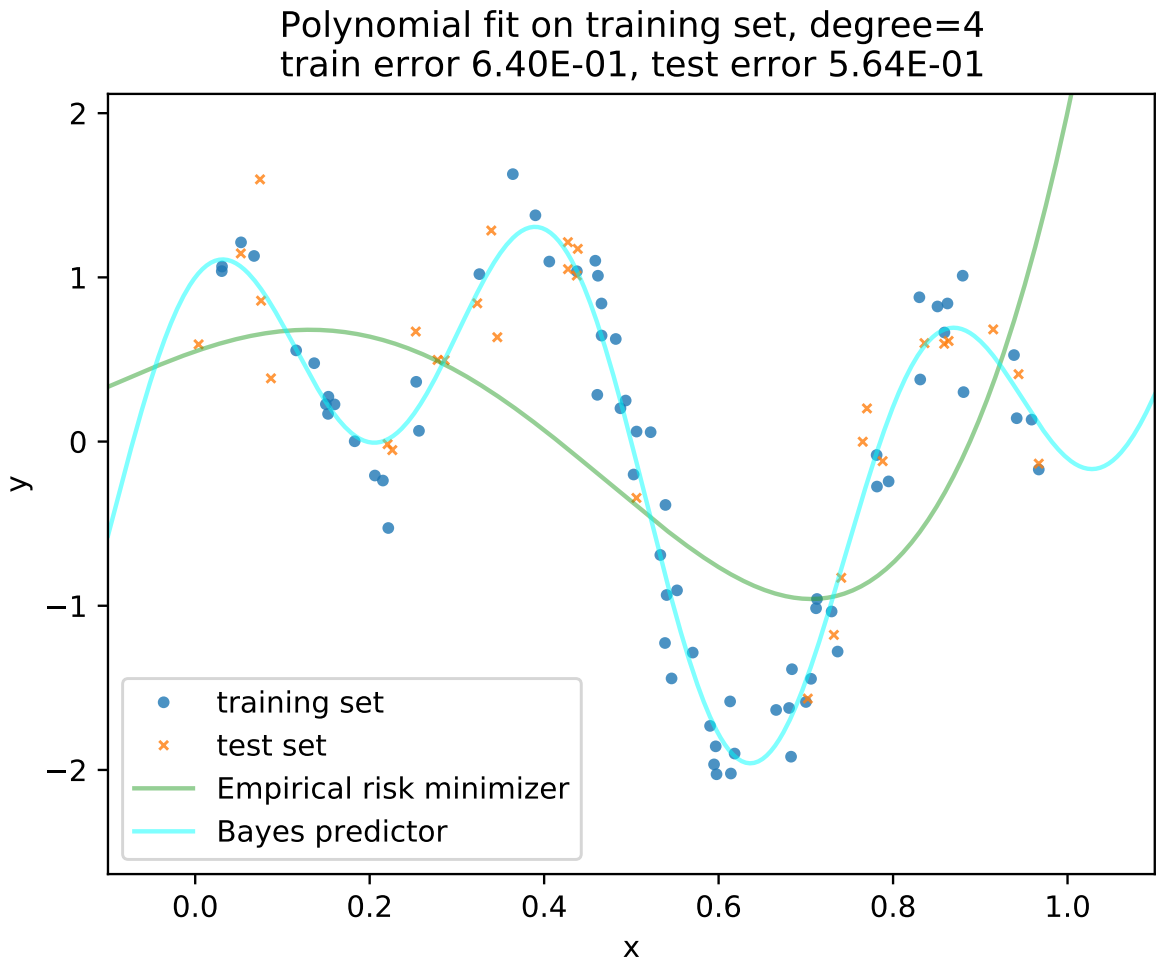
Polynomial regression



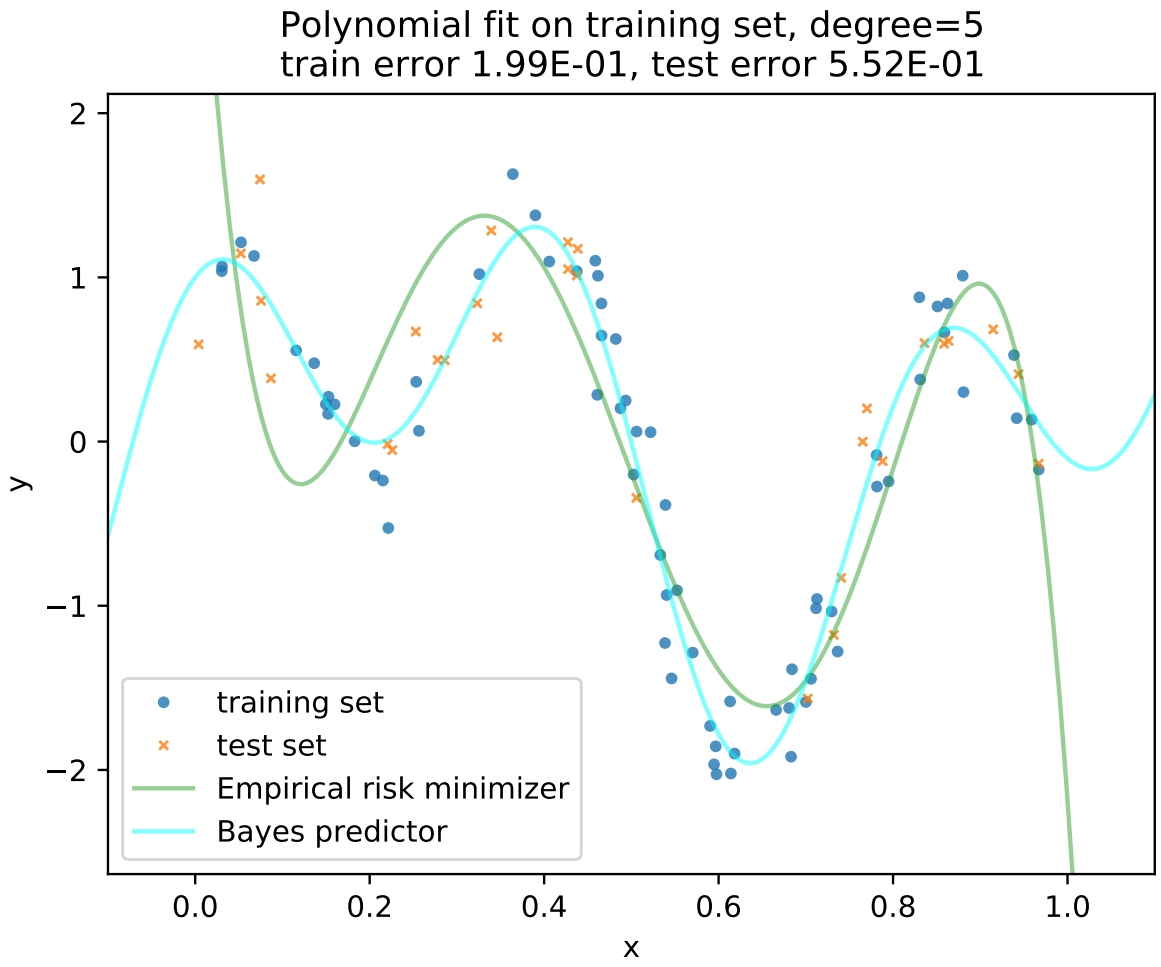
Polynomial regression



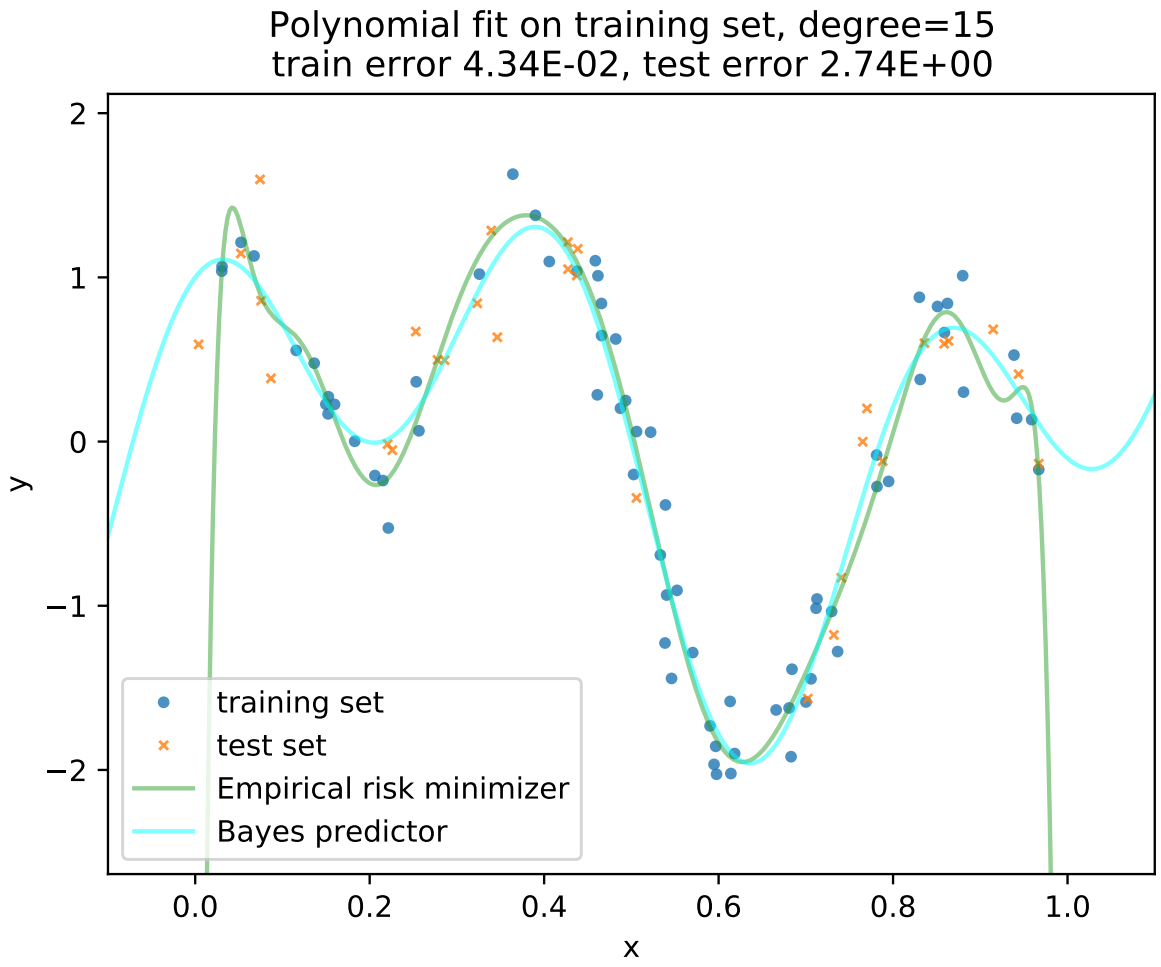
Polynomial regression



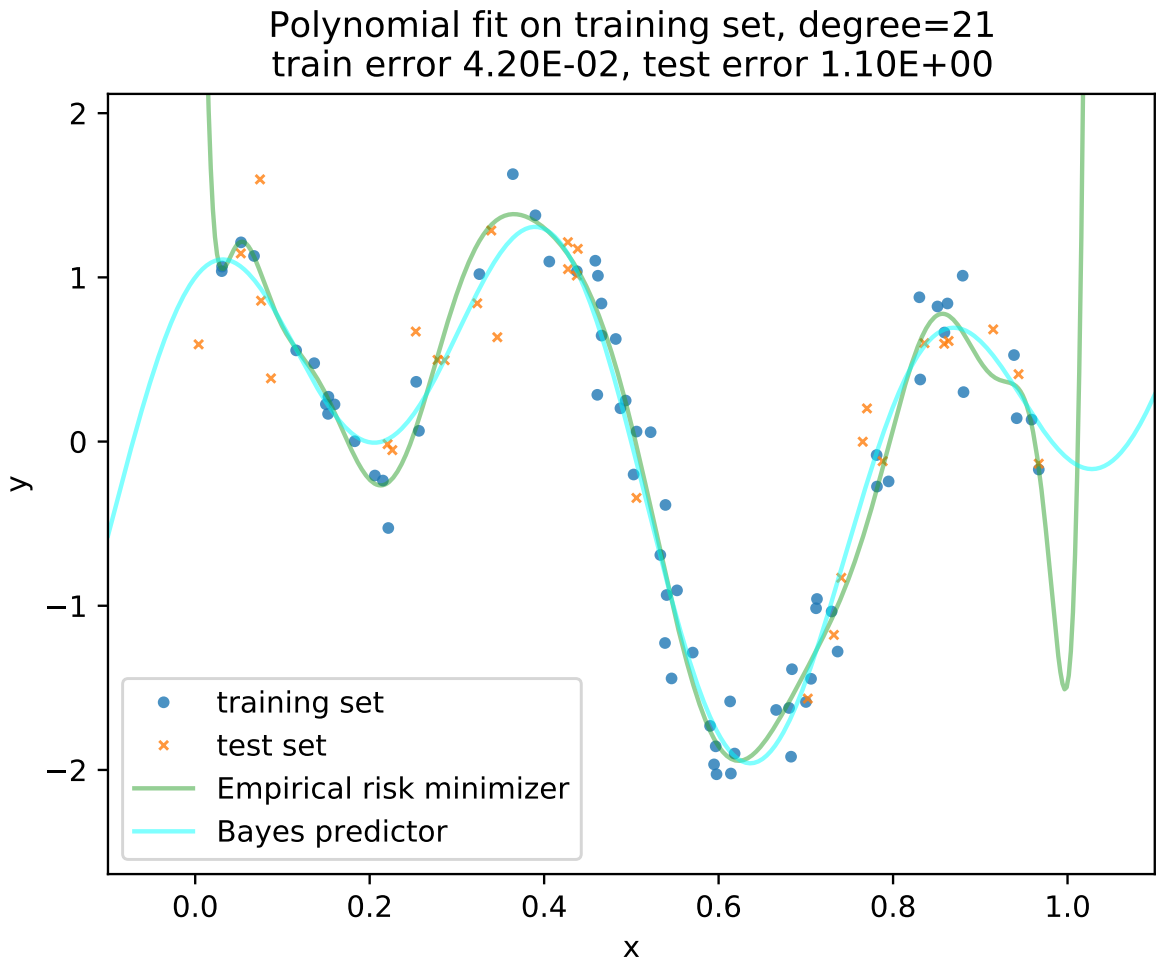
Polynomial regression



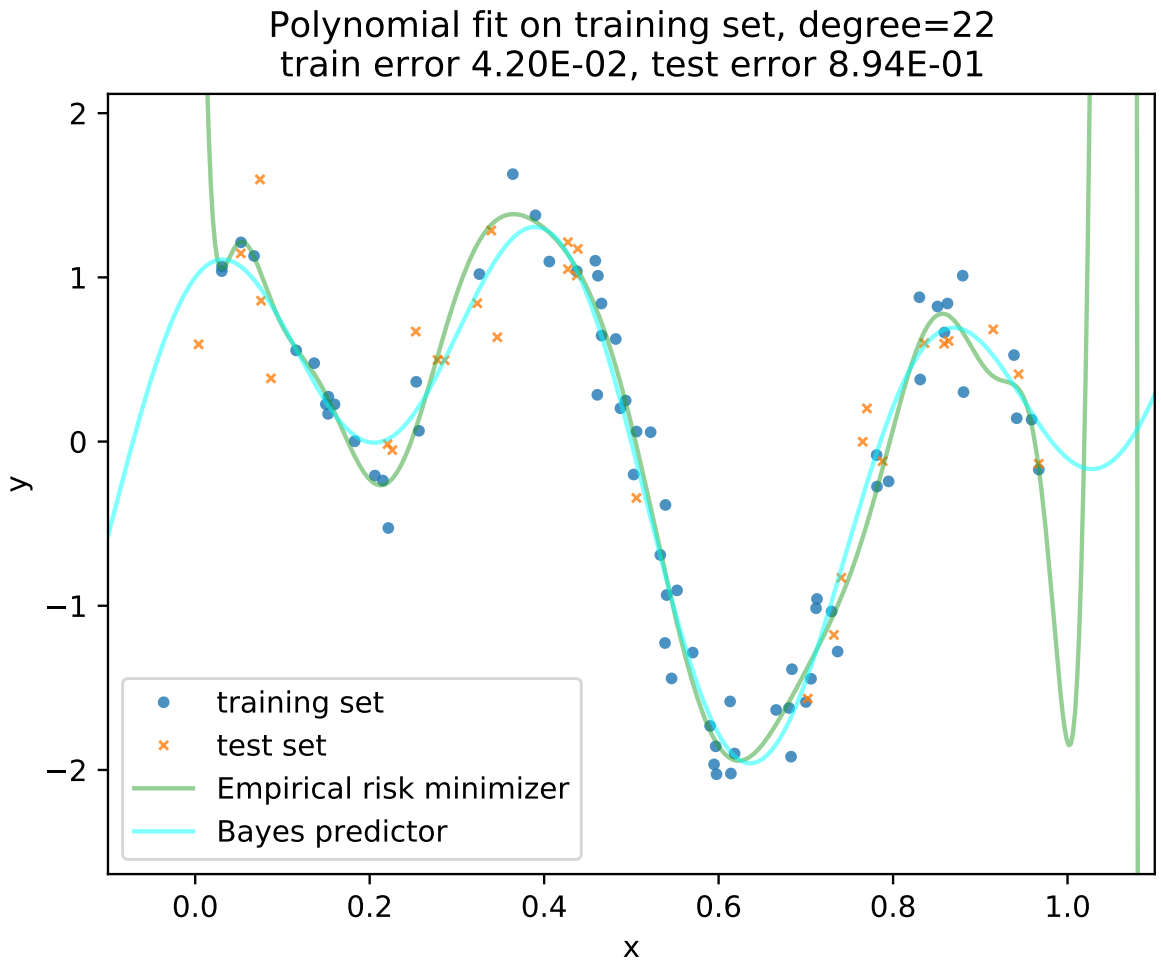
Polynomial regression



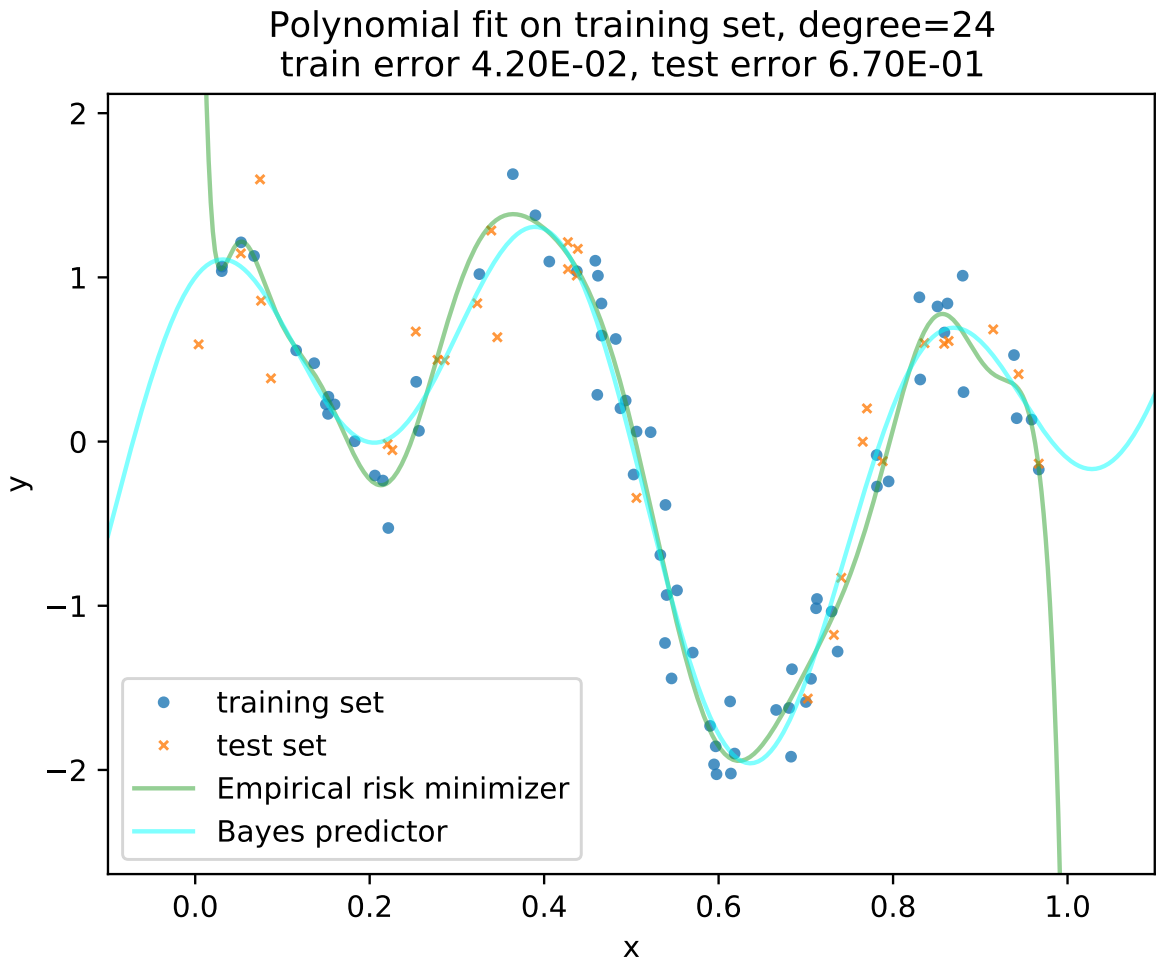
Polynomial regression

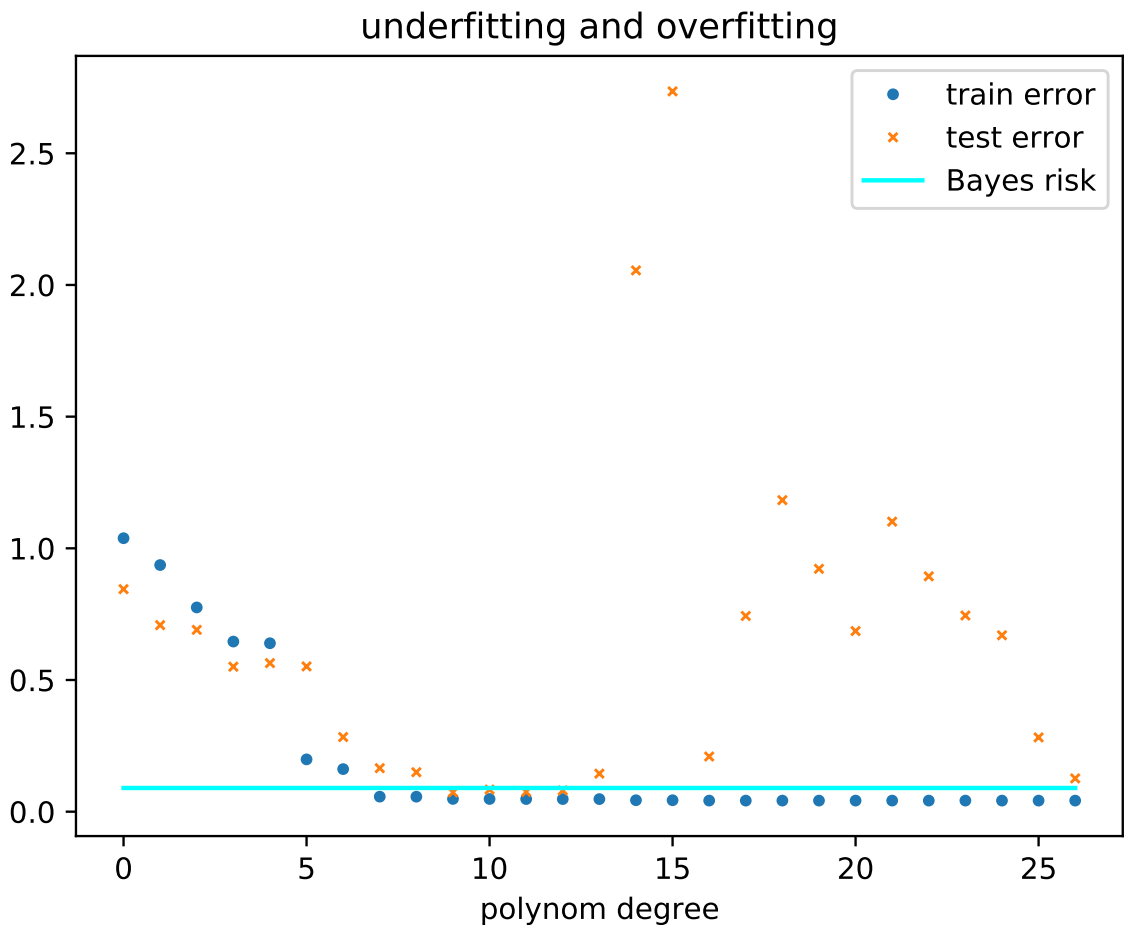


Polynomial regression



Polynomial regression




underfitting and overfitting
train error
test error
Bayes risk
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
polynom degree

Trying to prevent overfitting

- The problem of overfitting is linked to that of **generalization** : to what extent are we allowed to extrapolate the knowledge obtained on the training samples to new samples ?
- To improve generalisation, one can use heuristics like **regularization**.

Example regularization methods

- use smooth functions (fonctions régulières)
- penalize the magnitude of the weight vector in a linear regression or a neural network.
- remove neurons in a neural network (pruning)

Curse of dimensionality

Two numbers are important in machine learning :

- n : number of samples
- d : dimension (number of features) of a unique sample

Both can be large and prohibitive for some algorithms.

Curse of dimensionality

Two numbers are important in machine learning :

- n : number of samples
- d : dimension (number of features) of a unique sample

Both can be large and prohibitive for some algorithms.
Exercice 2 : **Example :** what is the algorithmic complexity of the inversion of an invertible matrix $A \in \mathbb{R}^{d,d}$?

Curse of dimensionality

- n is large when the dataset has many samples.
- d is large is each sample has many features :
  - image
  - DNA sequence
  - text
  - audio/video file

Curse of dimensionality

- The objects considered are in high d and high n
- Even "simple" situations grow very complex because of algorithmic complexity (Atari games).

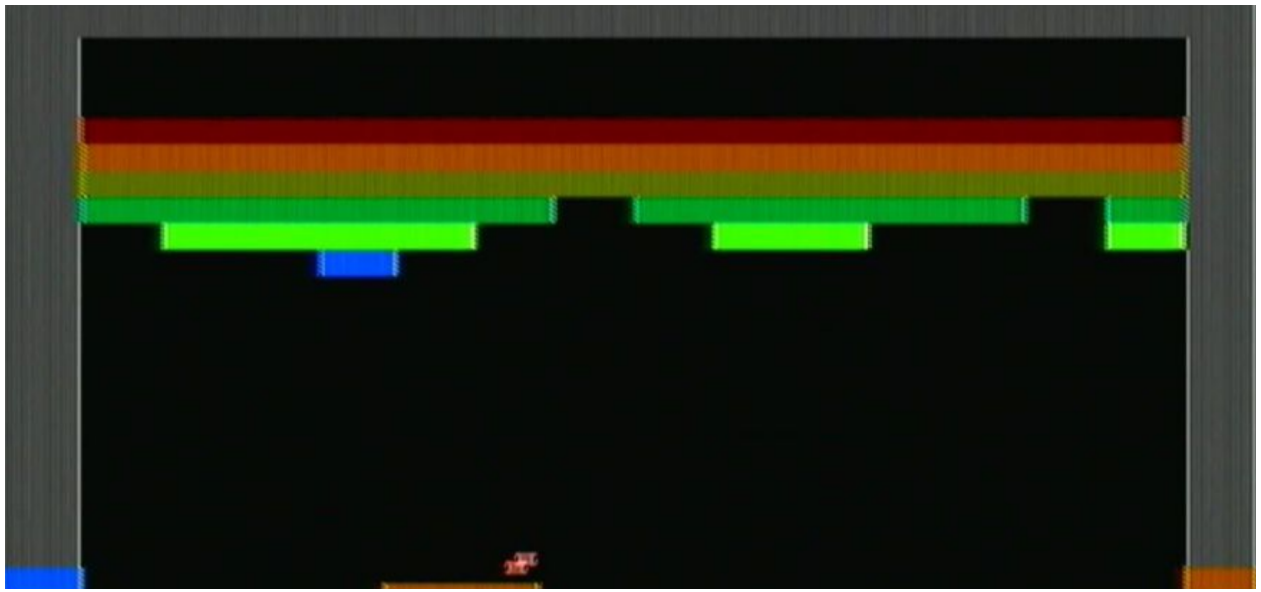
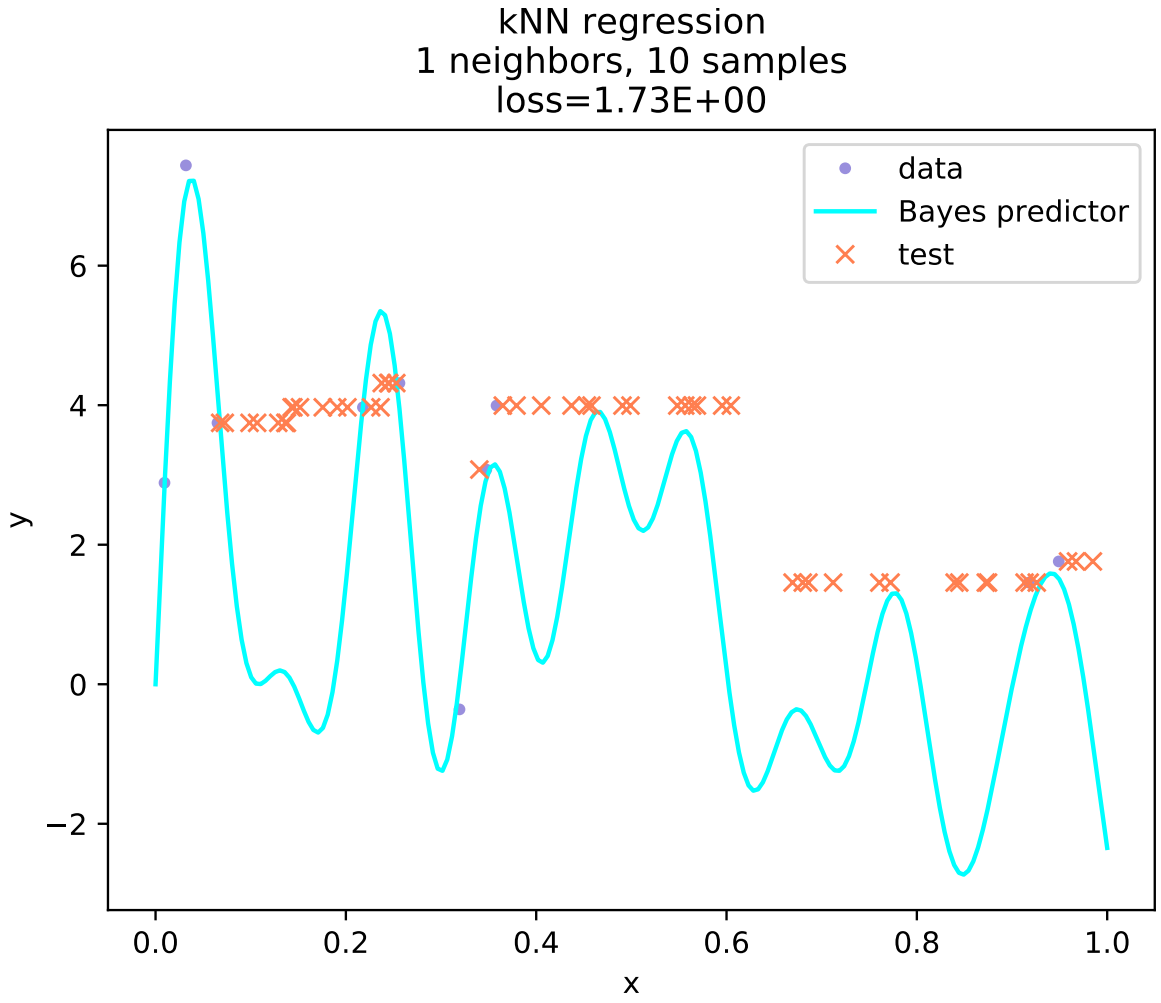


Figure – One Atari game

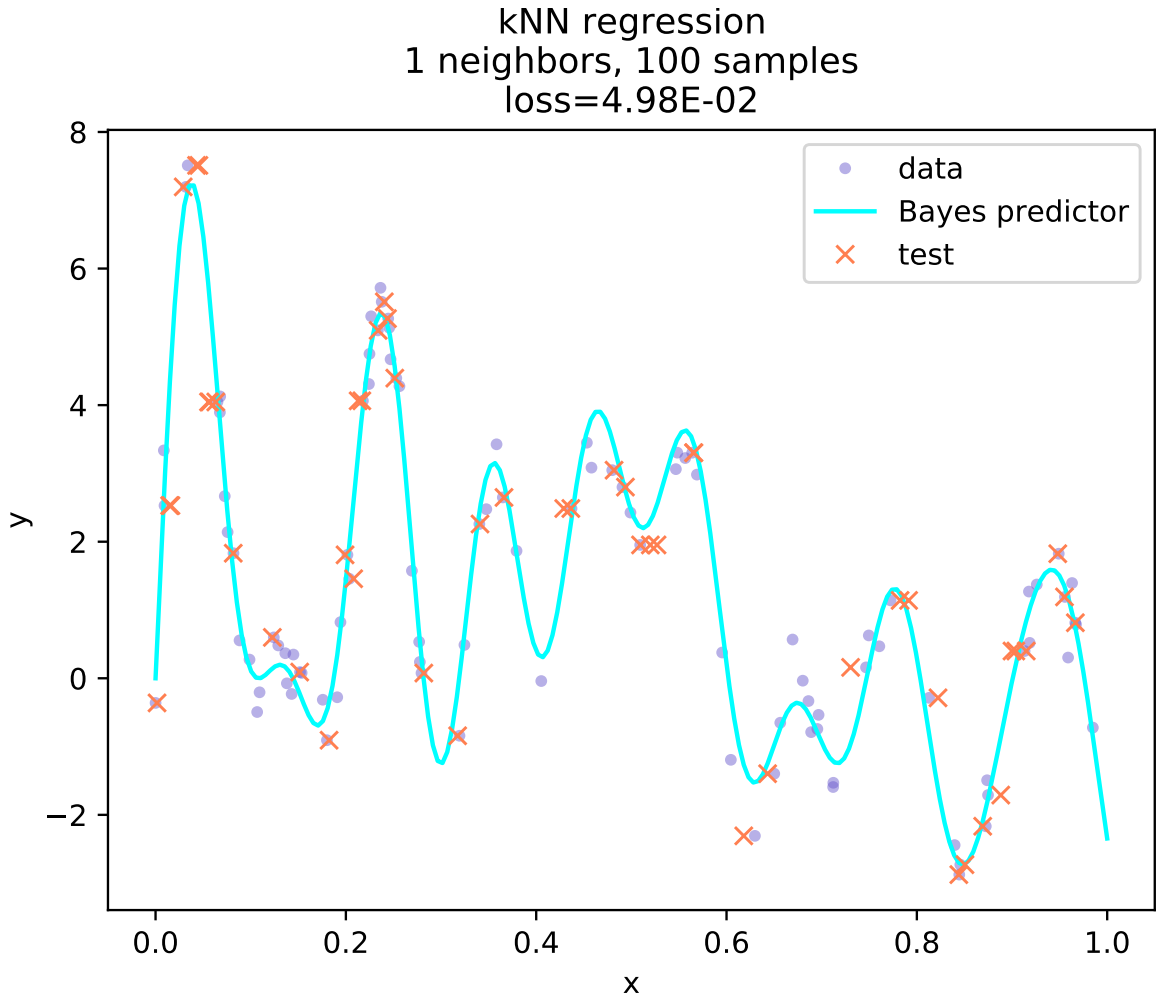
Local averaging

- The curse of dimensionality is the reason why **local averaging** does not work in large dimensions.
- It is also the reason why **dimensionality reduction** is sometimes necessary.

Local averaging



Local averaging



Optimization

The optimization problem of minimizing objective functions is often difficult.

- non convexity, non linearity (optimization problem)
- instability of estimators (statistical problem)

References I

Andrew, A. M. (1998).
Reinforcement Learning : An Introduction.

LeCun, Y. and Cortes, C. (2010).
{MNIST} handwritten digit database.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. (2016).
Mastering the Game of {Go} with Deep Neural Networks and Tree Search.
Nature, 529(7587) :484–489.

References II

Turing, A. M. (2009).
Computing machinery and intelligence.
In *Parsing the Turing Test : Philosophical and Methodological Issues in the Quest for the Thinking Computer*.