Introduction to AI

July 2, 2019

Introduction

- This morning we will start by discussing what AI is.

Introduction

- This morning we will start by discussing what AI is.
- We will see that it is not that easy to define.

Introduction

- This morning we will start by discussing what AI is.
- We will see that it is not that easy to define.
- Does the concept even make sense ?

Introduction

- This morning we will start by discussing what AI is.
- We will see that it is not that easy to define.
- Does the concept even make sense ?
- Let us first consider a few examples.

Introduction

- This morning we will start by discussing what AI is.
- We will see that it is not that easy to define.
- Does the concept even make sense ?
- Let us first consider a few examples.
- (Please feel free to ask questions !)

Introduction examples

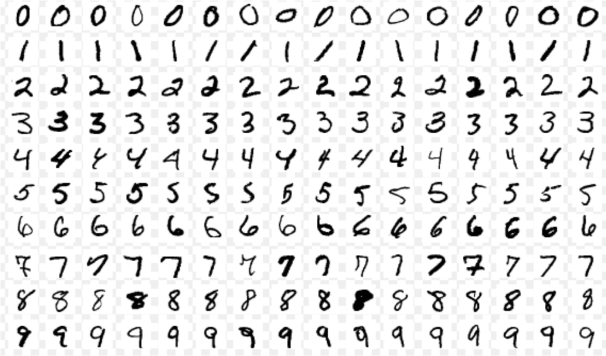


Figure: MNIST database [LeCun and Cortes, 2010]

Introduction examples

- Boston Dynamics robot (video)
- `https://www.youtube.com/watch?v=LikxFZZO2sk`
- `https://www.youtube.com/watch?v=gOTaYhjpOfo`

Introduction examples

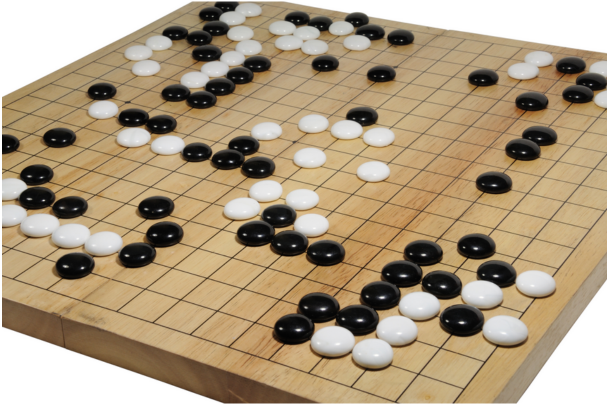


Figure: Go game, beaten by AlphaGo in 2017 [Silver et al., 2016]

Introduction examples

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
  - Coffee Machine
- All do different things but could be gathered under the term "AI".

Introduction

- It seems to be very varied, so let us try to see if we can define our topic.

Reflections and definitions on AI

AI paradigms

Some famous methods and use cases

Research and problems in AI

Conclusion : a problem that is hard to constrain

Definition

- Let us pick a definition of AI : "the theory and techniques aiming at emulating intelligence".

Definition

- Let us pick a definition of AI : "the theory and techniques aiming at emulating intelligence".
- Ok, but what is intelligence then ?

History

- 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).

Turing Test

- In 1950 and the article **"Computing machinery and intelligence"** [Turing, 2009], Turing introduces the **Turing test**.

Turing Test

- In 1950 and the article **"Computing machinery and intelligence"** [Turing, 2009], Turing introduces the **Turing test**.
- On of the forms of a Turing Test is a game in which a computer tries to behave like a human by answering questions.

Turing test

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**.
- Related to that, many 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.

Philosophy

- Further, some philosipher argue that the concept of "Artificial Intelligence" does not make sense.

Philosophy

- Some philisopher argue that the concept of "Artificial Intelligence" does not make sense.
- Or is even an **oxymoron**.

Moravec Paradox

- It is harder to emulate or simulate simple sensorimotor capacities than some higl 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, 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 at all.

History : AI Winter (1970's)

- In 1974 fundings for AI reseach started being cut by goverments because of unsuccessful projects.

History : Expert systems (1980's)

- In the 1980's and the 1990's, expert systems brought attention back to the field. They were used in several industries. Also the computing capacities were better allowing more applications.

History : Deep Blue (1997's)

- In the 1980's and the 1990's, expert systems brought attention back to the field. They were used in several industries. Also the computing capacities were better allowing more applications.
- In 1997, the victory of Deep Blue on Gary Kasparov (chess world champion) brought a lot more attention to AI.

History : 2000's

- In the 2000's, the Computer Science boom, and the Internet boom happened.
- AI was more and more covered in Science Fiction

History : 2010's, Watson, Machine Learning

- In 2011, **Watson**, a question-answering system, won the game Jeopardy.
  https://www.youtube.com/watch?v=WFR3lOm_xhE

History : 2010's, Watson, Machine Learning

- In 2011, **Watson**, a question-answering system, won the game Jeopardy.
  `https://www.youtube.com/watch?v=WFR3lOm_xhE`
- However the deepest change is the rise of **Machine Learning** and **Data Science**.
- **Machine Learning** (ML) is a subdomain of AI. Is is a younger topic. We will discuss the distinction between ML and AI later.
- Machine Learning has had impressive results on some specific problems, especially in **computer vision** (for instance).

History : 2010

- The research in AI has received much more funding and focus : in companies and in universities.

AI and science

- People doing "AI" today can actually come from rather different scientific communities. Which ones according to you ?

AI and science

- People doing "AI" today can actually come from rather different scientific communities:
  - Statistics

AI and science

- People doing "AI" today can actually come from rather different scientific communities:
  - Statistics
  - Optimization

AI and science

- People doing "AI" today can actually come from rather different scientific communities:
  - Statistics
  - Optimization
  - Other mathematical fields :

AI and science

- People doing "AI" today can actually come from rather different scientific communities:
  - Statistics
  - Optimization
  - Other mathematical fields : graph theory, combinatorics, information theory

AI and science

- People doing "AI" today can actually come from rather different scientific communities:
  - Statistics
  - Optimization
  - Other mathematical fields : graph theory, combinatorics
  - Statistical physics

AI and science

- People doing "AI" today can actually come from rather different scientific communities:
  - Statistics
  - Optimization
  - Other mathematical fields : graph theory, combinatorics
  - Statistical physics
  - Robotics

AI and science

- People doing "AI" today can actually come from rather different scientific communities:
  - Statistics
  - Optimization
  - Other mathematical fields : graph theory, combinatorics
  - Statistical physics
  - Robotics
  - Cognitive sciences / neuroscience / psychology

AI and science

- People doing "AI" today can actually come from rather different scientific communities:
  - Statistics
  - 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

AI and Machine Learning

- *Machine Learning* is a term slightly more specific than "AI"
- In Machine Learning, some parameters are learned in an *automatic way* in order to solve a problem or to optimize a solution

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)

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)

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. However it is unlikely that ML technologies are implemented in their robots for now.

Practical applications of AI

- Let us review some applications of AI in our time and former times, firt without looking at the underlying technique.

Famous AI : Dendral (1960's)

- Assisstant for chemists in order to classify unknown molecules.

Famous AI : Deep Blue (1997)

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

Famous AI : Siri, Google Translate

- Belongs to the category of **Natural Language Processing**

Famous AI : Ross (2010's)

- "Artificial Attorney" : helps lawyers and legal assisstant to research the law.

Famous AI : AlphaGo

- AlphaGo beat the world Go champion in 2017. The underlying processes have significant differences with Deep Blue.

Division

- Let us discuss AI tools and try to classify them.

Division

- Let us discuss AI tools and try to classify them.
- There could be several possibilities to classify AI tools :

Division

- Let us discuss AI tools and try to classify them.
- There could be several possibilities to classify AI tools :
  - **Machine learning** vs **no Machine learning**

Division

- Let us discuss AI tools and try to classify them.
- There could be several possibilities to classify AI tools :
  - **Machine learning** vs **no Machine learning**
  - **Deterministic** vs **Scothastic (non-deterministic)**

Rule-based systems

- From a rule base and an **inference engine**, they can answer questions or solve a **specific problem**
- They can use logical rules, such as IF condition THEN

Expert systems

- Expert systems are an example of rule based system
- Example : medical assistant for a doctor, ROSS, Dendral.

Decision Trees

- When building a game AI, a famous method is to build a large tree of decisions, from a big knowledge base.
- The AI looks for optimal actions in this tree, for example with the minimax algorithm.
- DeepBlue belongs to this category.

Machine learning and Optimization

- Let us now study **Machine Learning paradigms**.

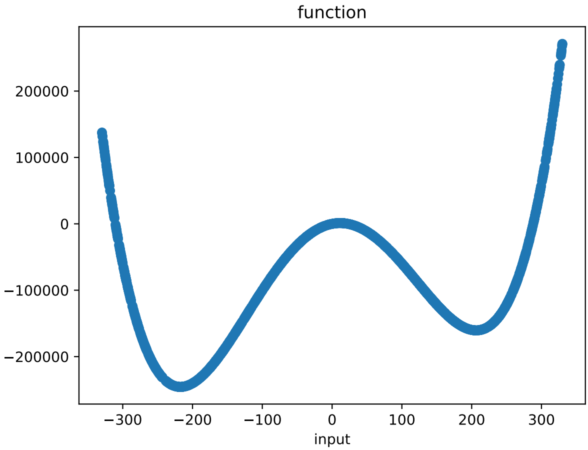
Machine learning and Optimization

- Let us now study **Machine Learning paradigms**.
- We said earlier that in Machine Learning, some parameters are learned in an automatic way to solve a problem.

Machine Learning and Optimization

- Let us now study **Machine Learning paradigms**.
- We said earlier that in Machine Learning, some parameters are learned in an automatic way to solve a problem.
- But how are the parameters learned ?

Notion of optimization



Supervised Learning : The problem

- For a certain input x, you want to predict an output y
- To do so, you learn from a number of **labeled examples** (x_i, y_i)
- In the case where what you want to predict is a **class**, it is a **classification problem**
- In the case where what you want to predict is a general function $y = f(x)$, it is a **regression problem**

Supervised learning

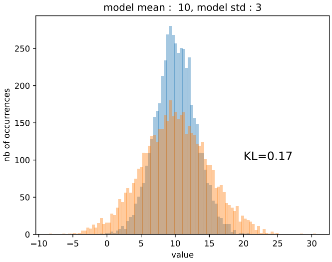
- For a certain input x, you want to predict an output y
- To do so, you learn from a number of **labeled examples** (x_i, y_i)
- In the case where what you want to predict is a **class**, it is a **classification problem**
- In the case where what you want to predict is a general function $y = f(x)$, it is a **regression problem**
- Example : MNIST (classification)
- Question : how do you choose and constrain your function f ?

Unsupervised Learning

- From a large number of samples x_i, you want to retrieve information on their **structure**
- For instance you want to learn a **distribution**, or a **clustering** of your data.
- Examples : social networks

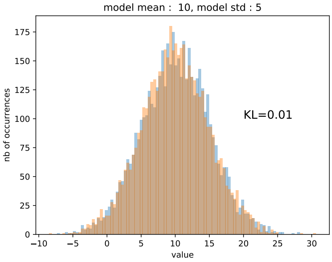
Unsupervised Learning

- From a large number of samples x_i, you want to retrieve information on their **structure**
- For instance you want to learn a **distribution**, or a **clustering** of your data.



Unsupervised Learning

- From a large number of samples x_i, you want to retrieve information on their **structure**
- For instance you want to learn a **distribution**, or a **clustering** of your data.



Unsupervised Learning

- From a large number of samples x_i, you want to retrieve information on their **structure**
- For instance you want to learn a **distribution**, or a **clustering** of your data.
- Question : how do you constrain your distribution to fit your data ?

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]

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

- A **more general paradigm** that describes an **agent** in a **world**.
- The standard formalization was the one proposed by Richard Sutton [Andrew, 1998]
- Example : a Chessplayer, AlphaGo, a game AI, automatic vacuum cleaner

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.

Example problem

- Typical Machine Learning situation : should I explore my environment more or exploit what I have learnt so far ?
- Concept of ϵ-greedy policy

Reinforcement learning

- We will study Reinforcement learning deeper **Week 2**.

Final remark

- These paradigms can be mixed

Final remark

- These paradigms wan be mixed
- Mosty, this means that
  - 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

Some famous methods

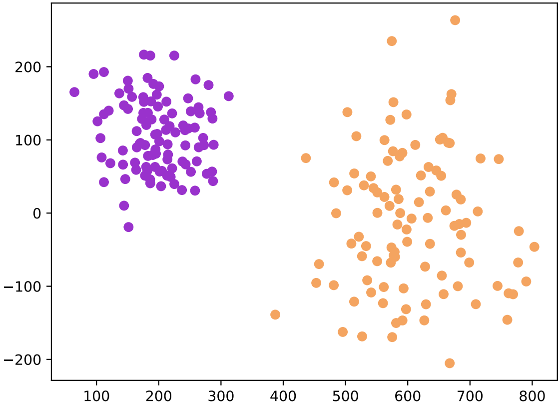
Let's look at some classical methods in ML

Linear separation

- We will start by studying a simple **supervised learning problem.**

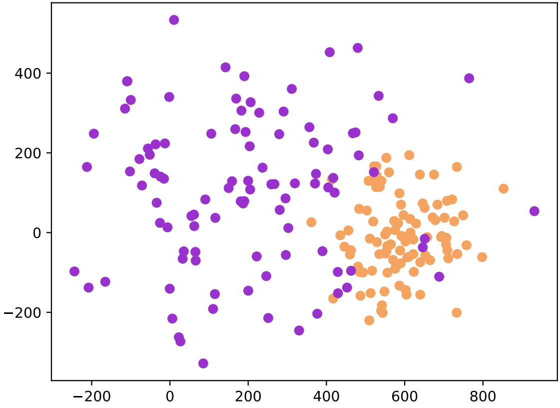
Linear separation

In some cases, the data are most easily separated.



Linear separation

What is the difference with this situation ?



Hyperplans

- In two dimensions, a linear separator will be a straight line

$$y = ax + b, a \in \mathbb{R}, b \in \mathbb{R} \tag{1}$$

- And in a problem with more dimensions ?

Hyperplans

- In two dimensions, a linear separator will be a straight line

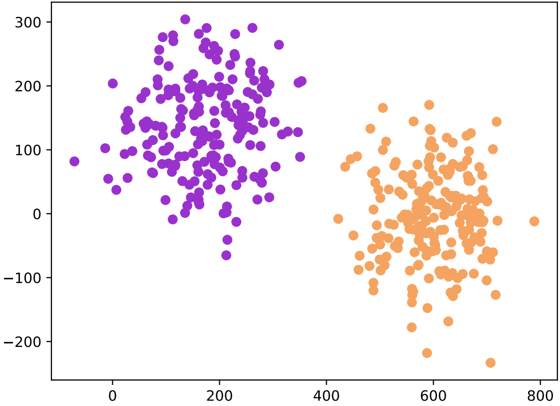
$$y = ax + b, a \in \mathbb{R}, b \in \mathbb{R} \tag{2}$$

- And in a problem with more dimensions ?

$$w \cdot x = b \tag{3}$$

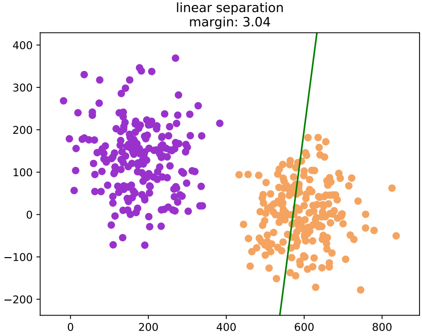
Linear separation

We will choose a linear separator for these data. Whate is the **best** linear separator ?



Exercise 1 : maximum margin

cd margin and use the file **linear_separator** in order to manually find the best linear separator for this dataset.



Linear separatation

- Unfortunately, most problems are *not linearly separable*

K means clustering

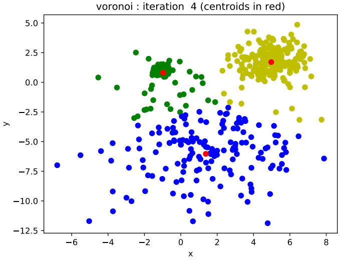
- A famous unsupervised clustering method
- 

Figure: K means clustering

Kmeans

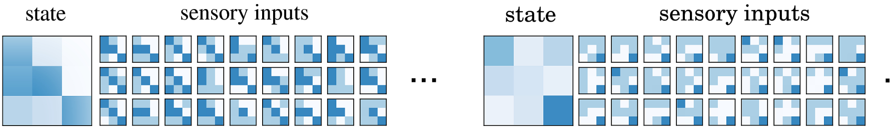


Figure: Other example of kmeans clustering, this time in 9 dimensions [Le Hir et al., 2018]

Kmeans : Expectation Maximisation algorithm

- Classical Machine Learning algorithm (EM)
- Blackboard
- What could be the drawbacks of this algorithm ?

Exercise 2 : Kmeans clustering

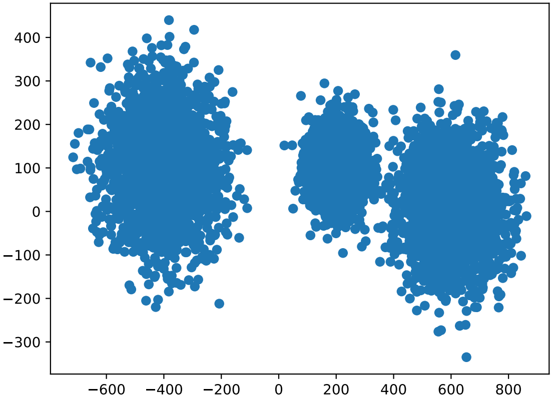


Figure: Data we want to cluster

Kmeans clustering

cd kmeans

- Modify the **k_means.py** file so that it performs the kmeans algorithm.
- There are **two mistake series** :
  - line 64
  - around line 84

  you will need to fix them.

You should obtain something like this:

Kmeans

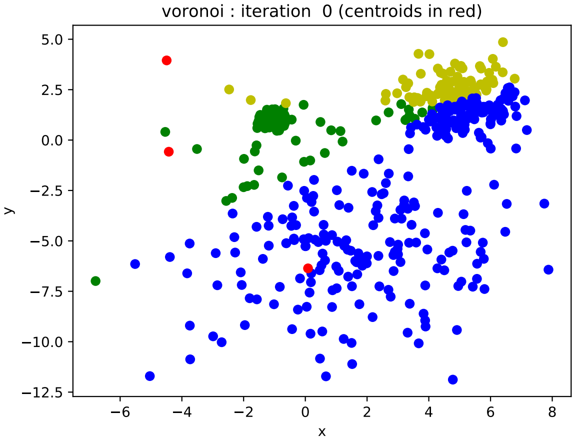


Figure: Voronoi 0th iteration

Kmeans

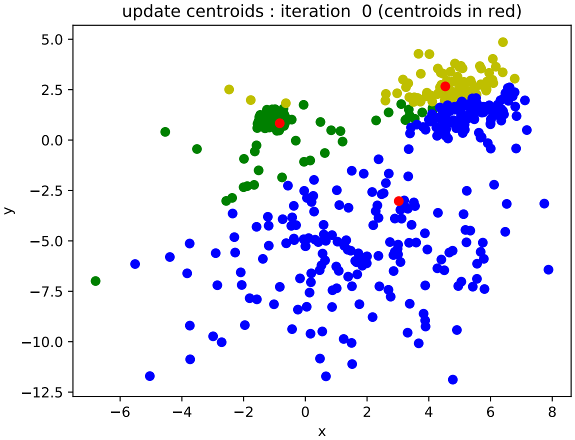


Figure: Centroids 0th iteration

Kmeans

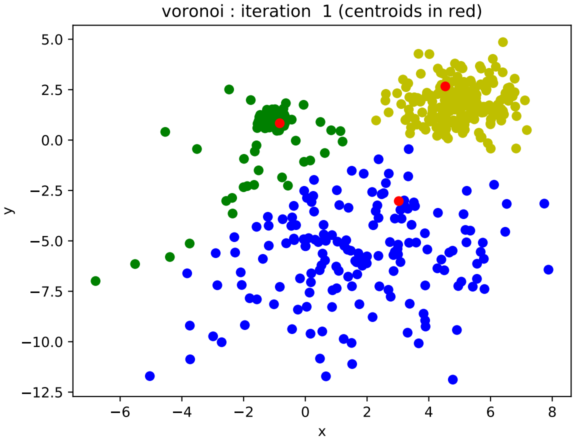


Figure: Voronoi 1st iteration

Kmeans

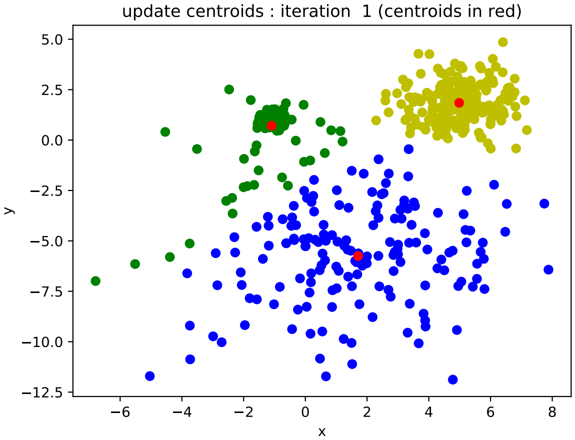


Figure: Centroids 1st iteration

Kmeans : Expectation Maximisation algorithm

- What would you do if the algorithm falls in a local optimum ?

Neural network

- A **neuron** is a simple elementary function
- A neural network a more complex function built with several neurons

Neural networks

- A **neuron** is a simple elementary function
- A neural network a more complex function built with several neurons
- A **Deep Neural Network** is a big neural networks ont more than two stacked layers of neurons

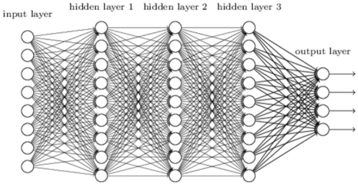


Figure: A deep neural network : source `https://datawarrior.wordpress.com/2017/10/31/interpretability-of-neural-networks/`

AlexNet

- **AlexNet** [Krizhevsky et al., 2012] is an example of Deep Neural Network : famous for a good performance at the ImageNet recognition challenge. Is is a **Convolutional Neural Network**

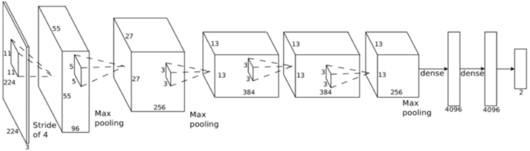
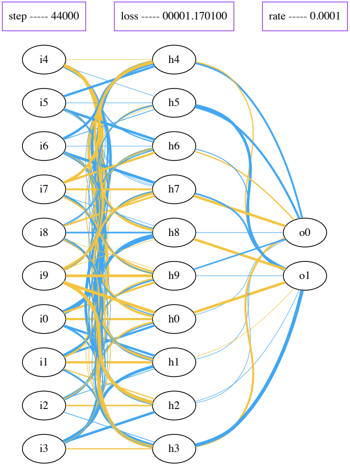


Figure: AlexNet

Neural Networks

- On week 3 we wil study neural networks
- We will go into technical details
- And apply them to MNIST (Supervised Learning canonical example)

Neural Networks



Spectral Clustering

- How can you cluster data if you do not have a **distance** between them ?

Spectral Clustering

- How can you cluster data if you do not have a **distance** between them ?
- Actually, what is a **distance** ?

Spectral Clustering

- How can you cluster data if you do not have a **distance** between them ?
- A **Similarity** is a more general notion that allows you to compare data
- It can be used in unsupervised learning contexts [Le Hir et al., 2018]

Research and problems in AI

What makes AI a hard problem ?

Dimensionality

- Was is the dimensionality of a problem ?

Curse of dimensionality

- First aspect : algorithmic complexity
- The objects considered are in high dimensional spaces, and in high number
- Even "simple" situations grow very complex (Atari games)

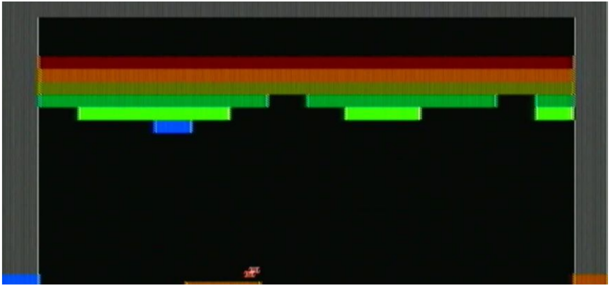


Figure: One Atari game

Curse of dimensionality

- Even "simple" situations grow very complex (Atari games)

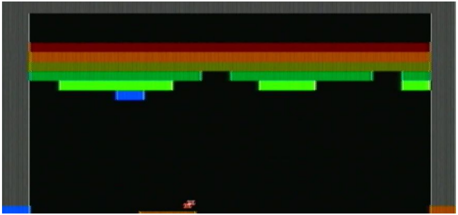


Figure: One Atari game

- Especially true for reinforcement learning
- If it is algorithmically hard to solve an Atari game, how hard would a real world problem be ?

Curse of dimensionality

- The problem of dimensionality was one of the reasons of AI winter in the 1970's.

Linear functions

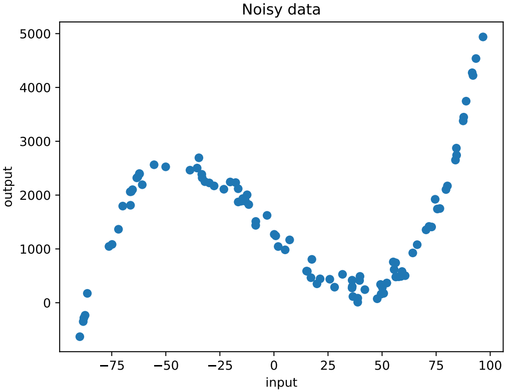
- Notion of linearity.

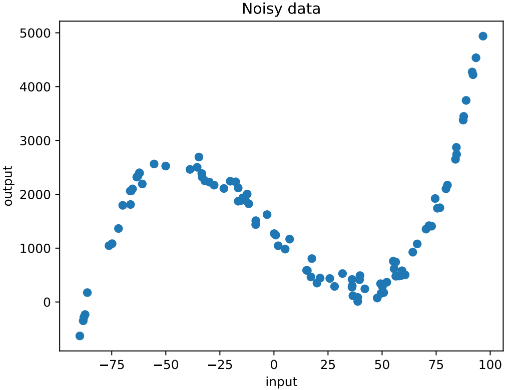
Non convexity, non linearity

- We try to optimize crazy functions : in extremely high dimensional spaces, with crazy shapes. (blackboard)
- So the power of the mathematical tools si limited and **experimentation** is needed.
- So there lacks **grounding** to the results

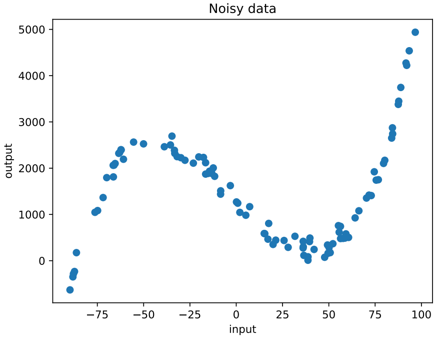
Overfitting

We will learn a **model** of the following data, in a **supervised learning** context.





Our **model** should allow us to predict the **output** for new **inputs**.
For instance what should be predicted fot an input of -48 ?



We need to choose:

- A **class** of model.
- A relevant **complexity** once the class is chosen.

Overfitting

- What could be the drawbacks of using a very simple model (very few parameters) ?

Overfitting

- What could be the drawbacks of using a very simple model (very few parameters)
  - Weak expressive power

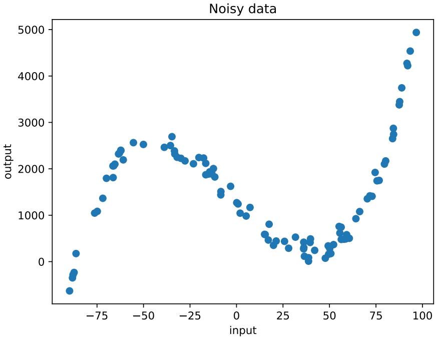
Overfitting

- What could be the drawbacks of using a very simple model (very few parameters)
  - Weak expressive power
- What could be the drawbacks of having a very complex model (that contains a very large number of parameters, e.g. millions as in a very deep neural network) ie a very high expressive power ?

Overfitting

- What could be the drawbacks of using a very simple model (very few parameters)
  - Weak expressive power
- What could be the drawbacks of having a very complex model (that contains a very large number of parameters, e.g. millions as in a very deep neural network) ie a very high expressive power ?
  - Harder to optimize
  - Harder to interpret
  - Can **overfit**

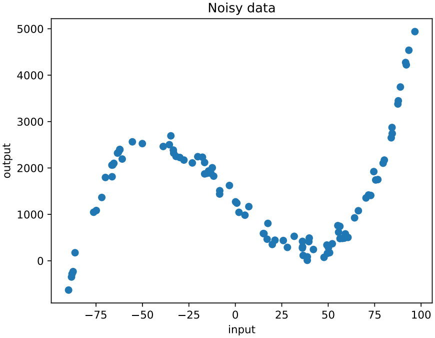
Exercise 3 : fitting



We want to perform supervised learning in order to be able to predict the output y for a new sample x.

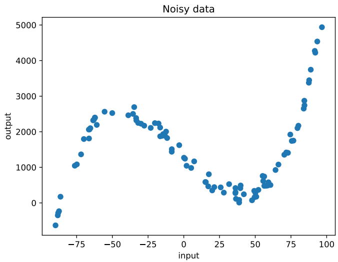
Exercise 3 : fitting

- We want to perform supervised learning in order to be able to predict the output y for a new sample x.



- To illustrate the problem of overfitting, we will use **polynoms** as models.

Exercise 3 : fitting



We will divide the dataset into two subsets :

- a **training set** : used to learn the most relevant polynom once the degree is chosen
- a **test set** : used to evaluate overfitting

Exercise 3 : fitting

- **cd overfitting**. Use the dataset contained in **linear_noisy_data.csv**, load it from **fit_data.py** in order to assess the impact of the **degree** of the polynom on overfitting.
- You need to edit the loop at the end of the file.

Exercise 3 : fitting

- The higher the degree of the polynom, the more parameters it has and the better it can fit the training points :

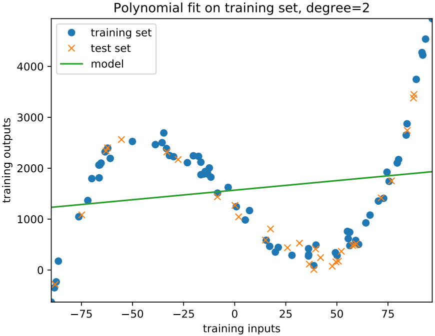


Figure: degree 2

Exercise 3 : fitting

- The higher the degree of the polynom, the more parameters it has and the better it can fit the training points :

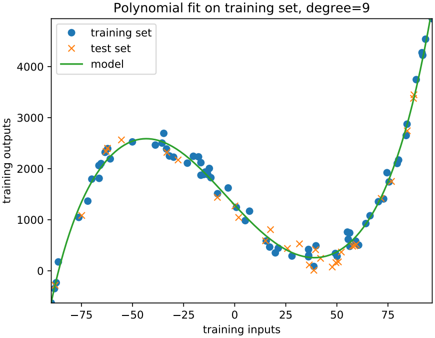


Figure: degree 9

Exercise 3 : fitting

- The higher the degree of the polynom, the more parameters it has and the better it can fit the training points :

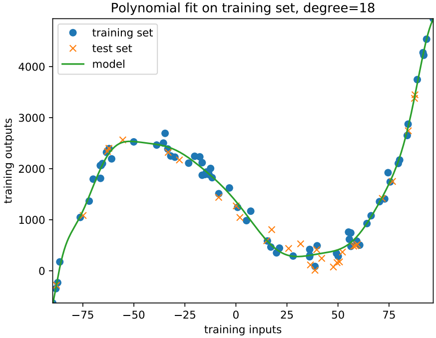


Figure: degree 19

Exercise 3 : fitting

- However, the error on the test set increases and the model looses **signification**

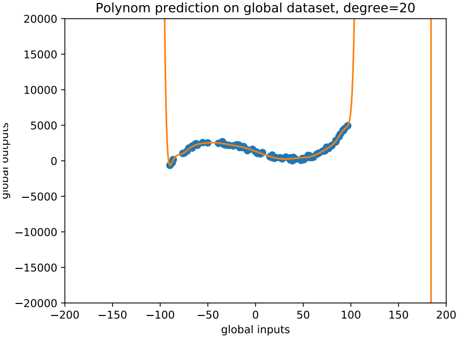


Figure: Useless solution

Exercise 3 : fitting

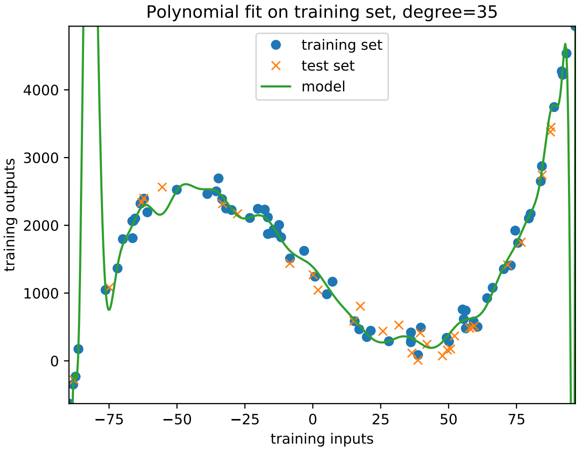


Figure: When the degree is too high.

Exercise 3 : fitting

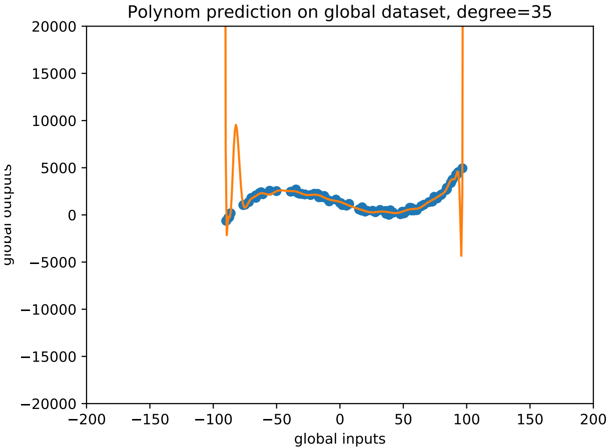
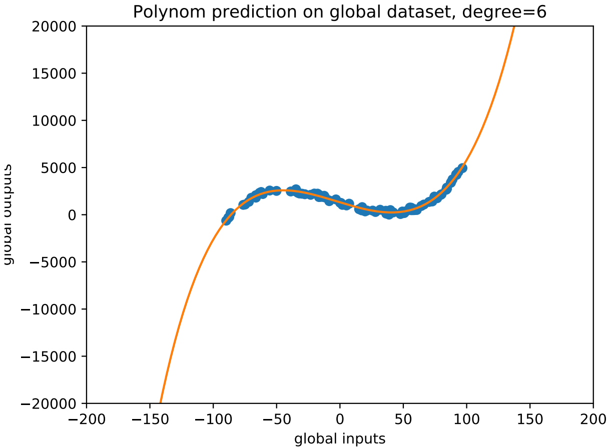


Figure: When the degree is too high.

Exercise 3 : fitting

In that situation, what degree should we use ?



Trying to prevent overfit

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

Regularization methods

- Penalize the magnitude of the weight in a neural network
- Remove neurons in a neural network (pruning)
- use smooth functions (continuous)

Deep learning

- Deep learning is powerful for some situations but is subject to the above shortcomings
- Some researchers try to have a better understanding of their behavior. Some famous ones are Yoshua Bengio (Montral), Geoffrey Hinton (Toronto), Stphane Mallat (Paris)

The End

References I

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

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012).
ImageNet Classification with Deep Convolutional Neural Networks.

Le Hir, N., Sigaud, O., and Laflaquière, A. (2018).
Identification of Invariant Sensorimotor Structures as a Prerequisite for the Discovery of Objects.
Frontiers in Robotics and AI, 5(June):1–14.

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

References II

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.

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