SAISON 22/23

Formation

Introduction Deep Learning

Séquence 1

History and basic concepts

FIDLE

https://fidle.cnrs.fr

MIAI Grenoble Alpes · cnrs · UGA Université Grenoble Alpes · EFELIA - MIAI ECOLE FRANÇAISE DE L'INTELLIGENCE ARTIFICIELLE · DEVLOG · Resinfo · sari · ANITI · GRICAD · SIMaP

Cette session va être enregistrée.
Retrouvez-nous sur notre chaine YouTube :-)

This session will be recorded.
Find us on our YouTube channel :-)

https://fidle.cnrs.fr/youtube

SAISON 22/23

Formation

Introduction Deep Learning

Séquence 1

History and basic concepts

FIDLE

https://fidle.cnrs.fr

MIAI Grenoble Alpes

cnrs

UGA Université Grenoble Alpes

EFELIA - MIAI
ECOLE FRANÇAISE DE L'INTELLIGENCE ARTIFICIELLE

DEVLOG RÉSEAU DU DÉVELOPPEMENT LOGICIEL

Resinfo

sari

ANITI

GRICAD

SIMaP

Behind the scenes :

The FIDLE team

By : https://beta.dreamstudio.ai/dream

Alexis Dubos (CNRS/IDRIS)
Antoine Regnier (CNRS/IDRIS)
Bertrand Cabot (CNRS/IDRIS)
Bruno Tessier (CNRS/IDRIS)
Camille Parisel (CNRS/IDRIS)
Dominique Fournier (CNRS/CRIC)
Maldonado Eric (INRAE)
Genevieve Morvan (CNRS/IDRIS)
Hatim Bourfoune (CNRS/IDRIS)
Jean-Luc Parouty (CNRS/SIMaP)
Kamel Guerda (CNRS/IDRIS)
Laurent Risser (CNRS/IMT)
Léo Hunout (CNRS/IDRIS)
Maxime Song (CNRS/IDRIS)
Myriam Peyrounette (CNRS/IDRIS)
Nathan Cassereau (CNRS/IDRIS)
Pierre Cornette (CNRS/IDRIS)
Remy Dubois (CNRS/IDRIS)
Soraya Arias (INRIA)
Sylvie Thérond (CNRS/IDRIS)
Thibaut Very (CNRS/IDRIS)

« The image of a engineer with a computer, with the style of Leonardo da Vinci in color »

→Séq. 14 (16 mars 2023)

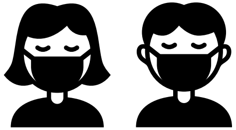
By : https://beta.dreamstudio.ai/dream

Behind the scenes :

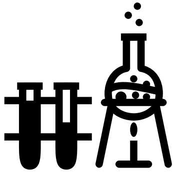
Those who made it possible

l'institut d'Intelligence Artificielle **MIAI** de Grenoble, via le projet **EFELIA**, le **CNRS** et l'Université Grenoble Alpes **UGA**, avec le soutien et la participation de **l'IDRIS**, de la Formation Permanente CNRS et de la Mission pour les Initiatives Transverses et Interdisciplinaires **MITI** du CNRS, via les réseaux **DevLOG**, **Resinfo** et **Calcul**, ainsi que du laboratoire **SIMaP**.

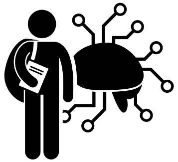
Before getting serious :



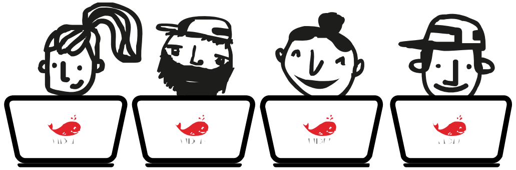
Face-to-face would be better, let's adapt!



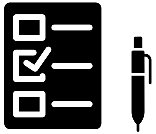
Be nice, we do our best ! ;-)



For **doctoral students**, we will provide proof of the following



New procedure !



We will make a small survey (anonymous) to know you better.

Objectives

Understand what Deep Learning is,
its concepts and basics,

Develop a **first experience** on
simple and ...understandable cases,

Learn **good practices**, how to use tools
and mutualized resources,

Promote **exchanges** and and reflections on deep
learning and its uses !

Resources

https://fidle.cnrs.fr

Powered by CNRS CRIC, and UGA DGDSI
of Grenoble, Thanks !

- Course materials (pdf)
- Practical work environment*
  - Datasets
  - Python venv
  - Notebooks
- Videos (YouTube)

(*) Procedure via Docket or pip
Remember to get the latest version !

Resources

You can also subscribe to :

FIDLE http://fidle.cnrs.fr/listeinfo

DEVLOG RÉSEAU DU DEVELOPPEMENT LOGICIEL Recherche et Enseignement Supérieur https://listes.services.cnrs.fr/wws/info/devlog[1]

GROUPECALCUL https://listes.math.cnrs.fr/wws/info/calcul[2]

(1) List of ESR* developers,
(2) List of ESR* « calcul » group
Where ESR is Enseignement Supérieur et Recherche, french universities and public academic research organizations

Resources

For all your questions :

http://fidle.cnrs.fr/q2a

Okay, but now let's talk about something more fun!

Program

20 Séquences
du 17 novembre
au 14 mai 2023

1. **History, Fundamental Concepts**
   - **Basic Regression** DNN
   - **Basic Classification** DNN
2. 3. **Hight Dimensional Data** CNN
4. Demystify **mathematics** for neural networks.
5. **Training strategies** Evaluation / **Sparse data (text)** Embedding
6. **Sequences data** RNN
7. PyTorch — A small detour with **PyTorch**.
8. **«Attention is All You Need»** Transformers
9. New ! **Graph Neural Network** GNN
10. **Autoencoder networks** AE
11. **Variational Antoencoder** VAE
12. **Project session** «My project in 180 s»
13. **Generative Adversarial Networks** GAN
14. New ! **Diffusion Model** Text to image
15. **AI, Law, Society** and **Ethics**
16. New ! **Model and training optimization** Resource efficiency
17. **Jean-Zay** GPU acceleration
18. New ! **Physics-Informed Neural Networks** PINNS
19. **Deep Reinforcement Learning** RL
20. **What will be tomorrow's AI** Review & perspectives !

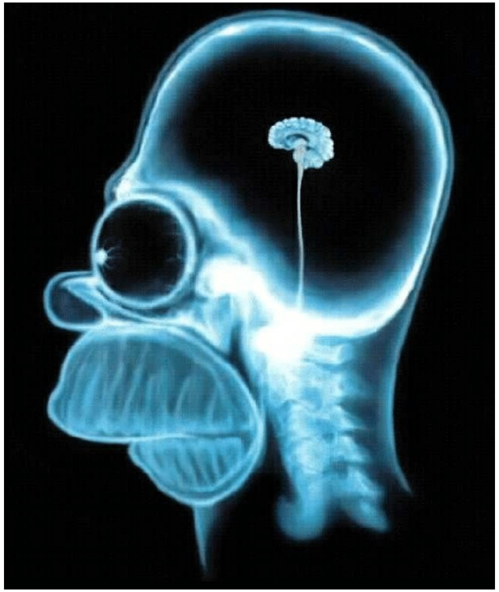
FIDLE

SAISON 22/23

History and basic concepts

Sequence 01

Questions of the day



Homer Simpson's brain seen with MRI/X ray.

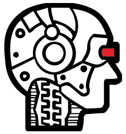
Why would we want to use artificial intelligence, when natural intelligence work so well?

➔ What is an artificial neuron ?

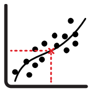
➔ How did they arrive ?

Roadmap

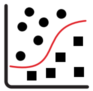
1



History, Fundamental Concepts



Basic Regression
DNN



Basic Classification
DNN

1.1 Introduction
Context, tools and ressources

1.2 From the linear regression to the first neuron

1.3 Neurons in controversy

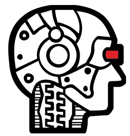
1.4 Data and neurons

Basic Regression

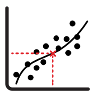
Basic Classification

Roadmap

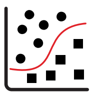
1



History, Fundamental Concepts



Basic Regression
DNN



Basic Classification
DNN

1.1 Introduction
Context, tools and ressources

1.2 **From the linear regression to the first neuron**

1.3 **Neurons in controversy**

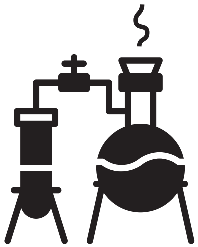
1.4 **Data and neurons**

Basic Regression

Basic Classification

Scientific paradigms

1st paradigm



Experimental science

2nd paradigm

$$i\hbar\frac{d}{dt}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle$$

$$\nabla \times H = J + \frac{\partial D}{\partial t}$$

$$F = G \cdot \frac{m_1 \cdot m_2}{r^2}$$

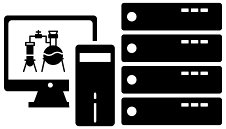
Theoretical science

3rd paradigm

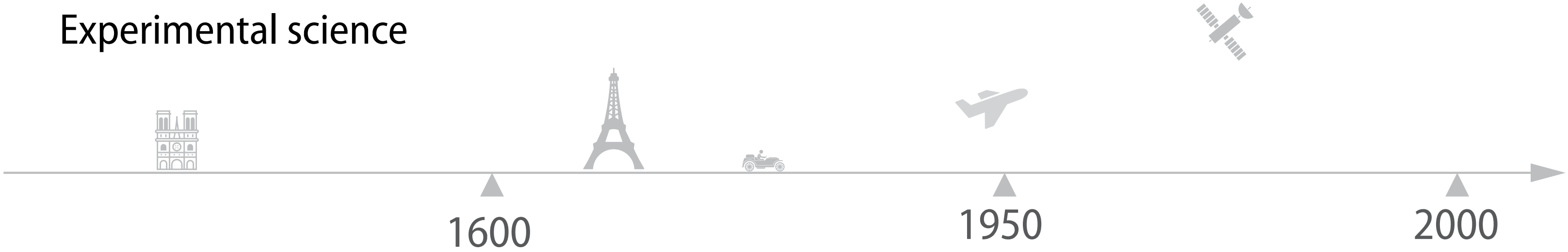
$$i\hbar\frac{d}{dt}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle$$

$$\nabla \times H = J + \frac{\partial D}{\partial t}$$

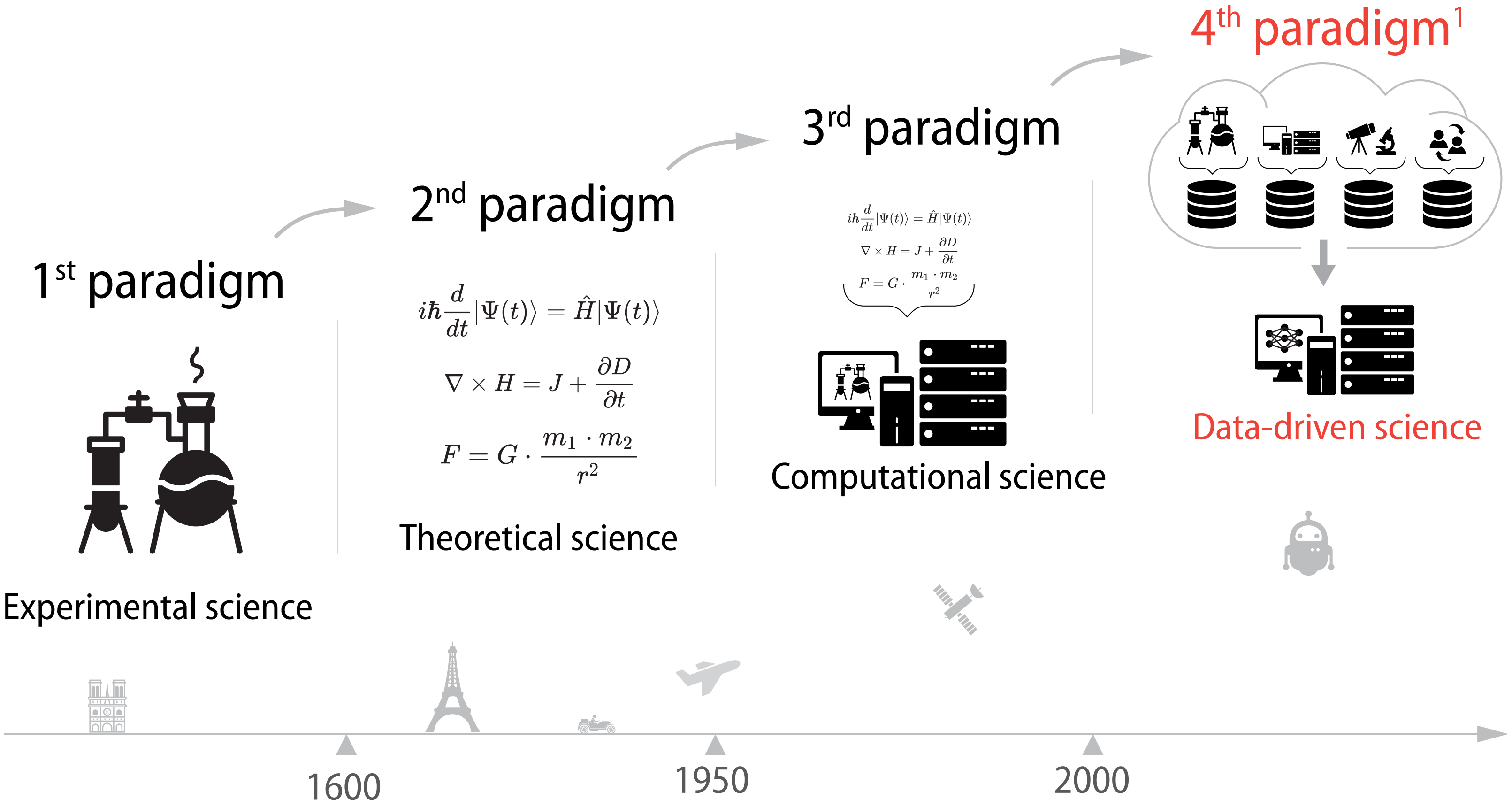
$$F = G \cdot \frac{m_1 \cdot m_2}{r^2}$$



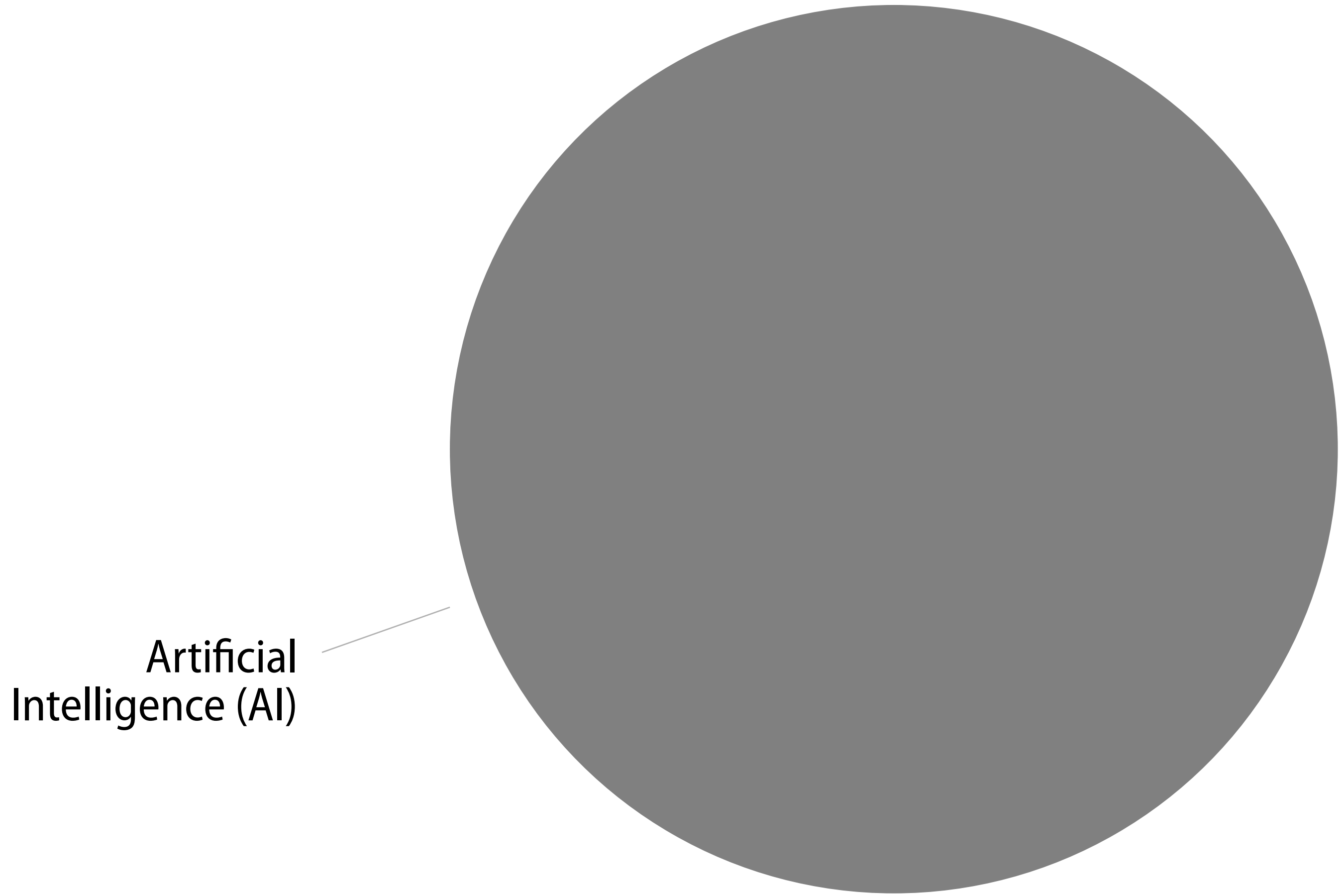
Computational science

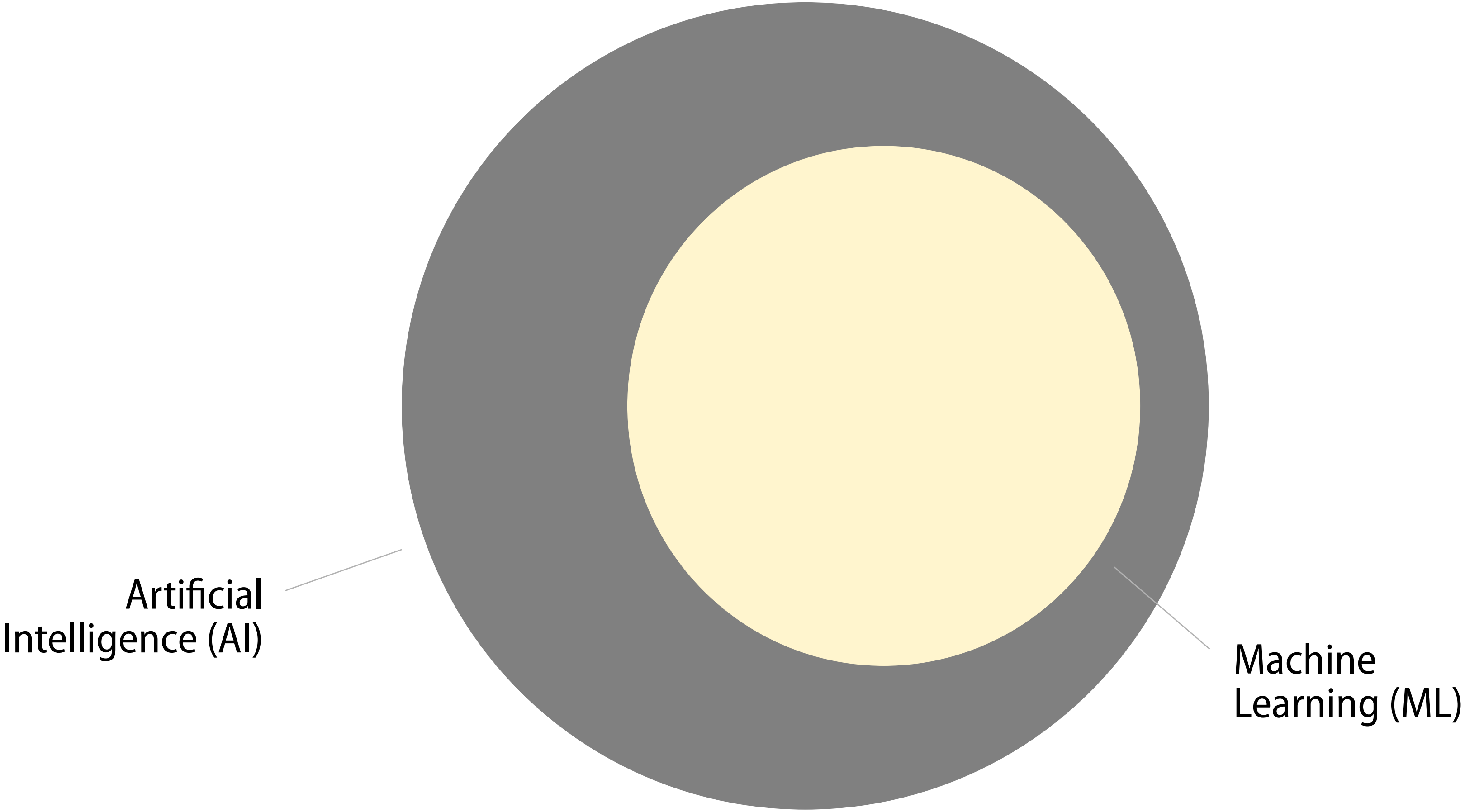


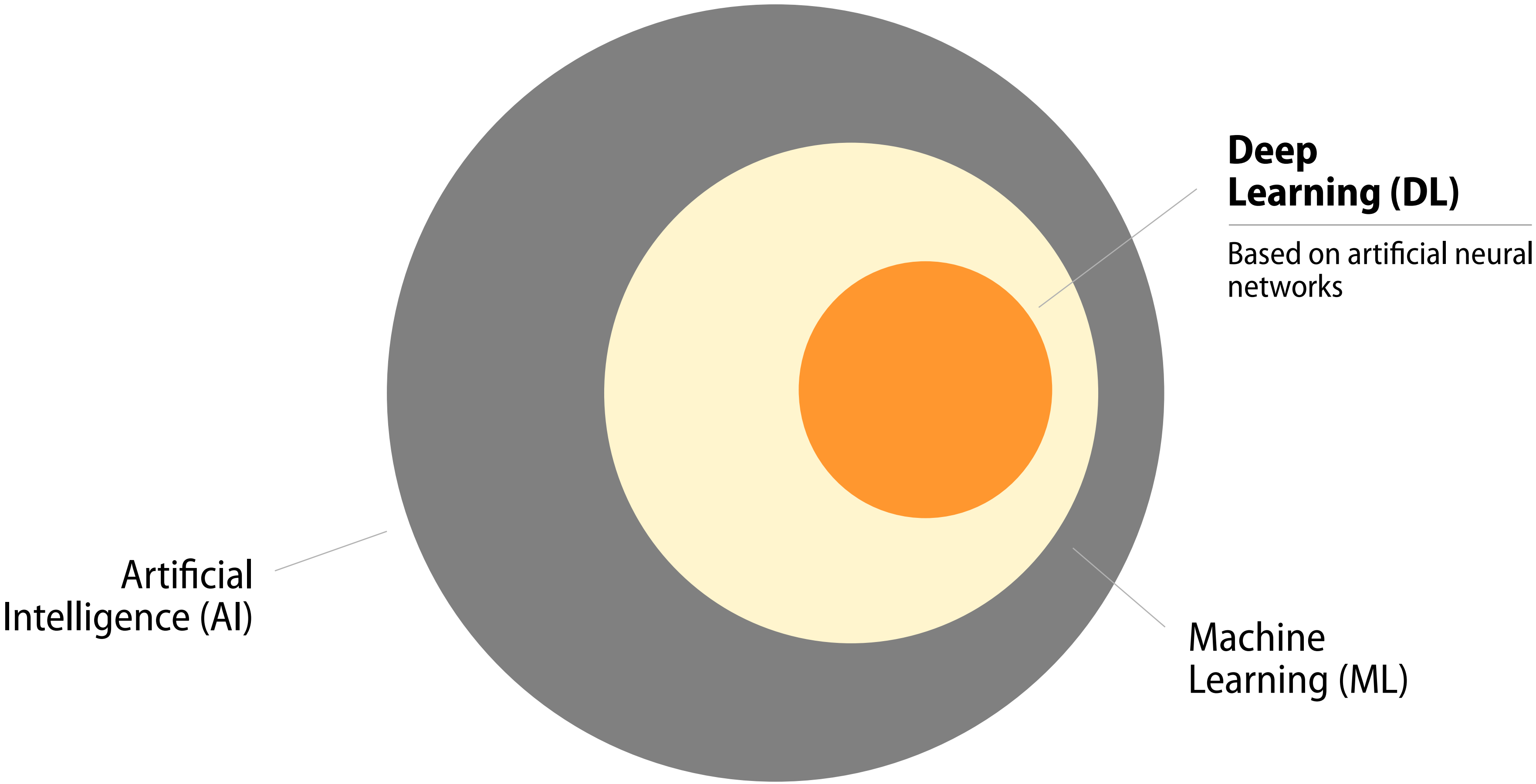
Scientific paradigms

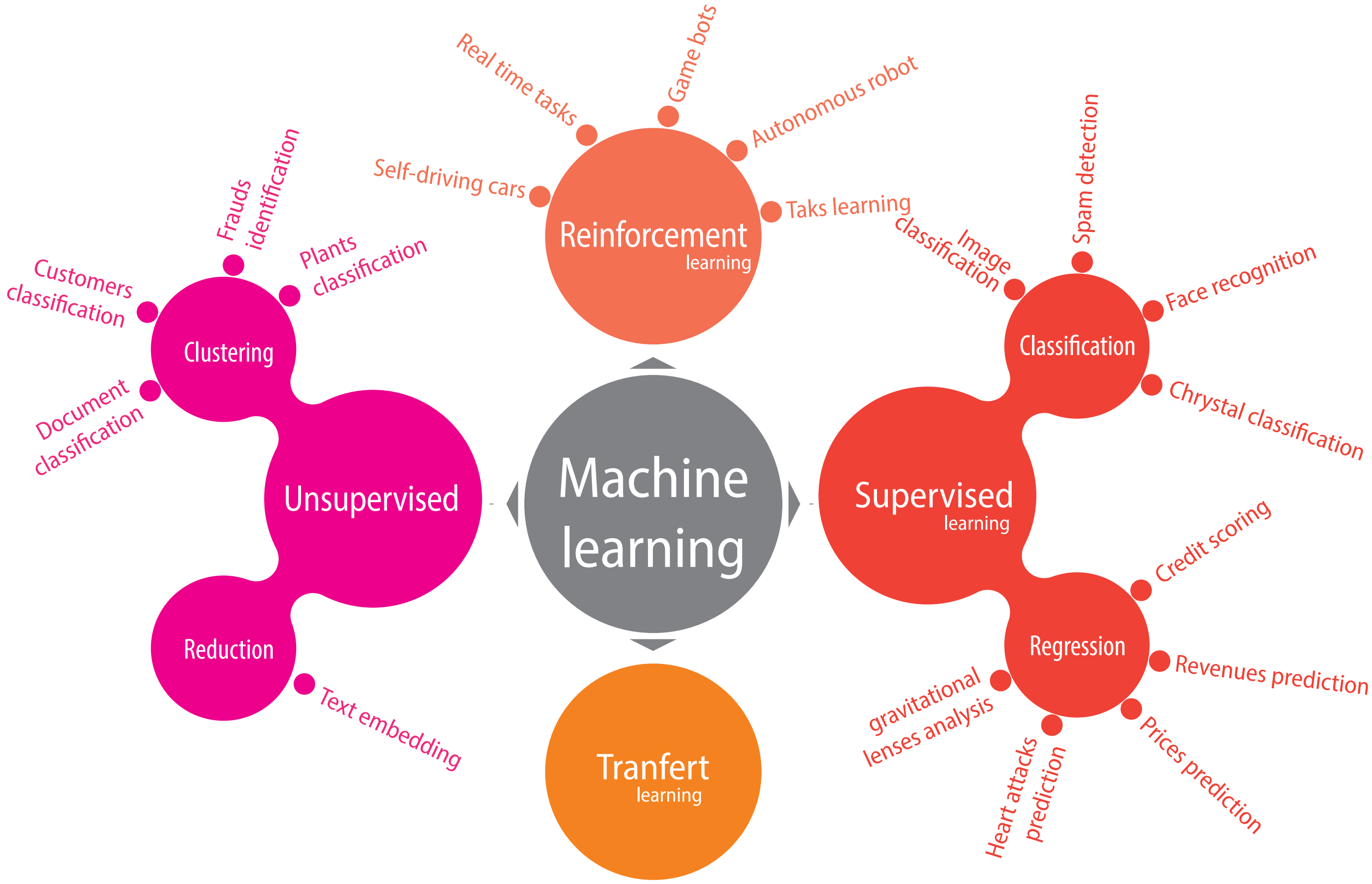


[1] Jim Gray, 2007 [GRAY]

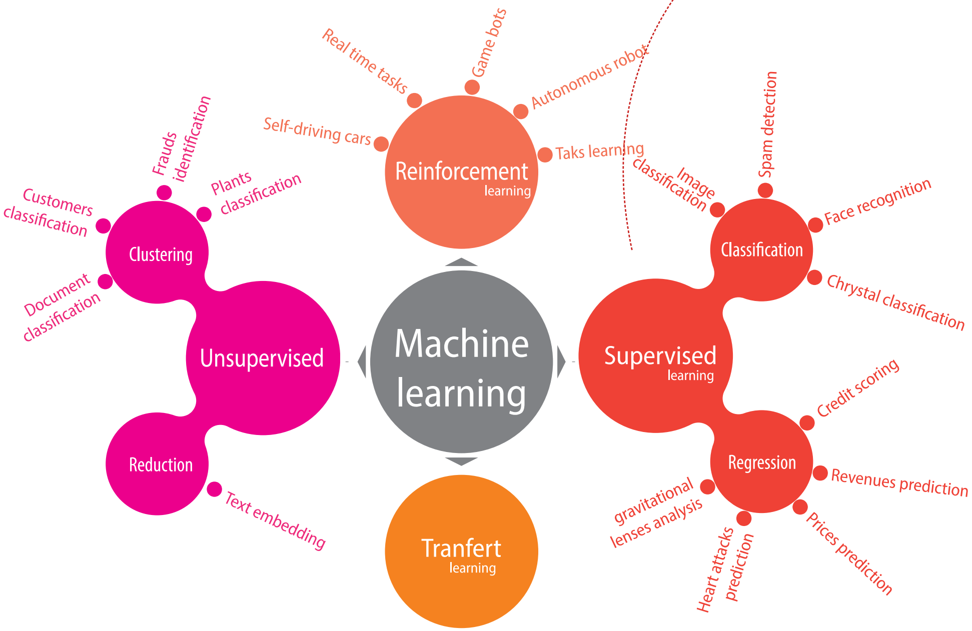

Artificial
Intelligence (AI)


Artificial Intelligence (AI)
Machine Learning (ML)


Deep
Learning (DL)
Based on artificial neural networks
Artificial
Intelligence (AI)
Machine
Learning (ML)

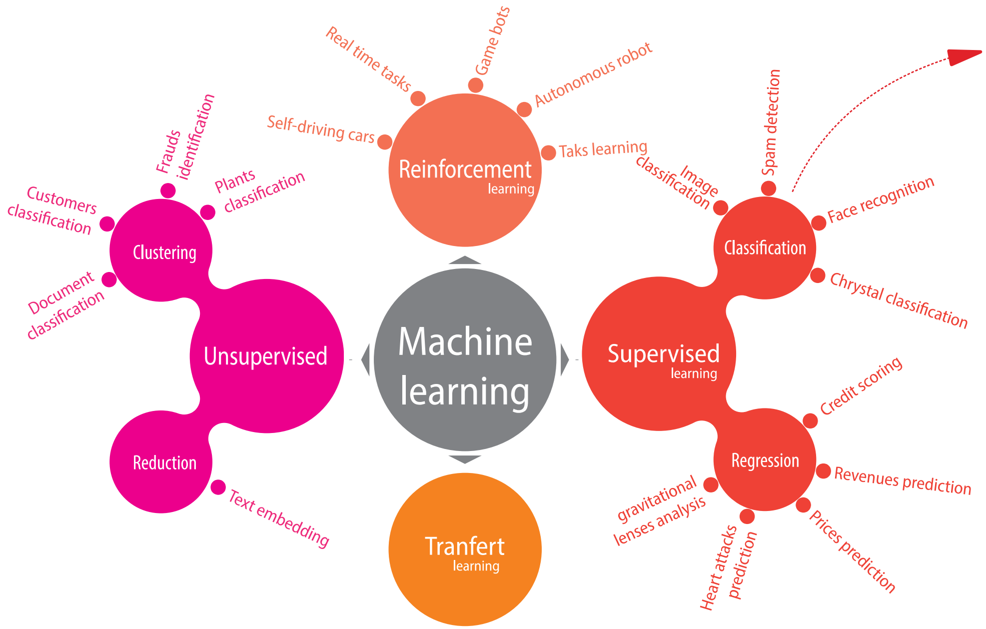

Machine learning
Reinforcement learning
Real time tasks
Game bots
Autonomous robot
Self-driving cars
Taks learning
Unsupervised
Clustering
Frauds identification
Plants classification
Customers classification
Document classification
Reduction
Text embedding
Supervised learning
Classification
Image classification
Spam detection
Face recognition
Chrystal classification
Regression
Credit scoring
Revenues prediction
gravitational lenses analysis
Heart attacks prediction
Prices prediction
Tranfert learning

Supervised learning



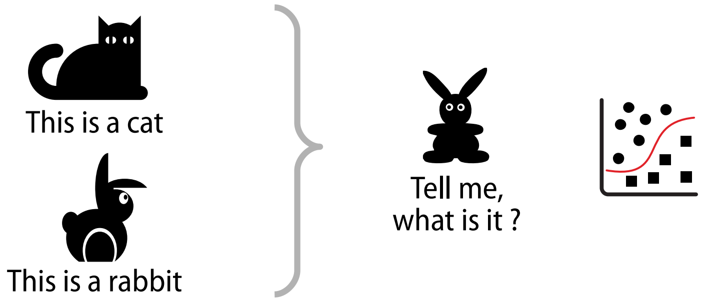
Learning
from **examples**

Supervised learning



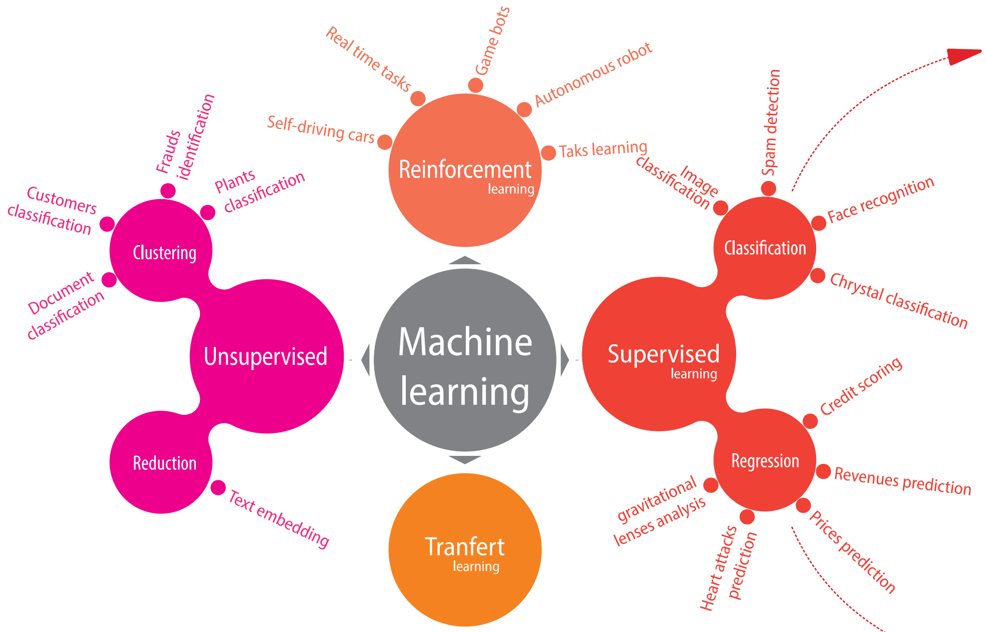
Classification :

Predict qualitative informations



Supervised learning

Learning from examples

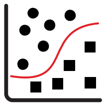


Classification :
Predict qualitative informations

This is a cat

This is a rabbit

Tell me, what is it ?

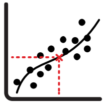


Régression :
Predict quantitative informations

150 K€ 400 K€

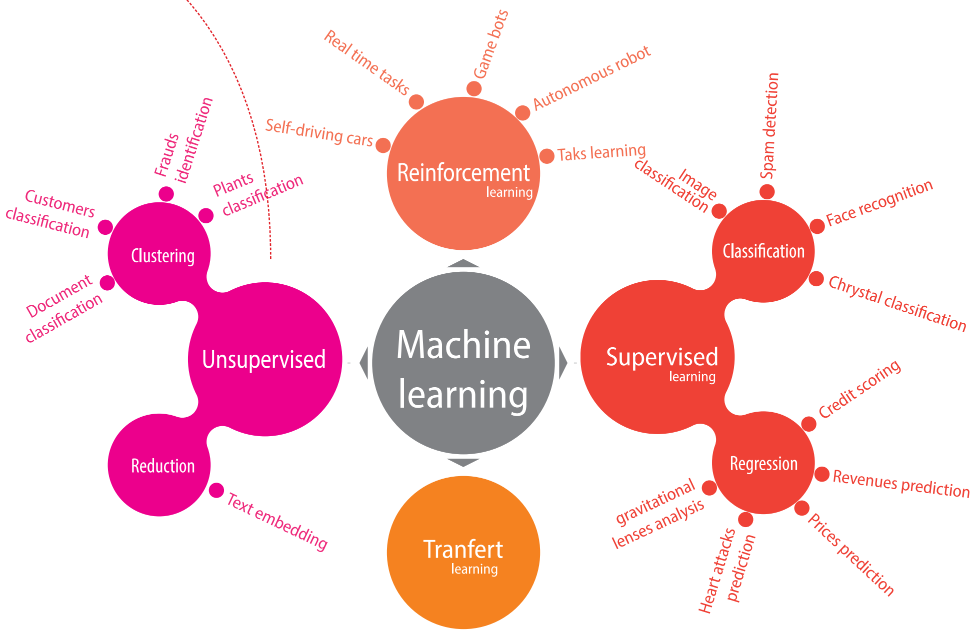
120 K€ 100 K€

Tell me, what's the price ?



Unsupervised learning

Learning
from data
alone

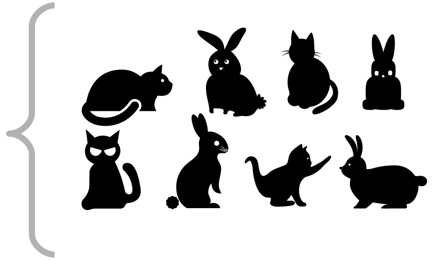


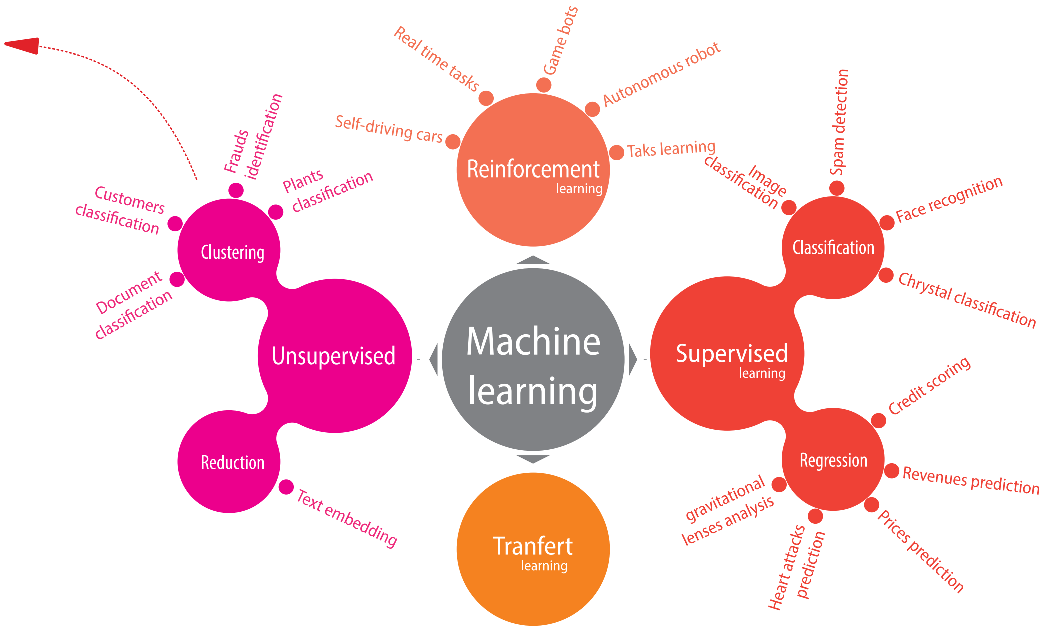
Unsupervised learning

Learning from data alone

Clustering :
Finding Common Relationships

What is the relationship between these data ?



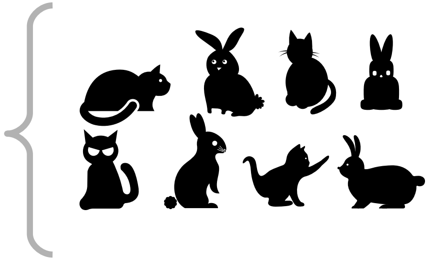


Unsupervised learning

Learning from data alone

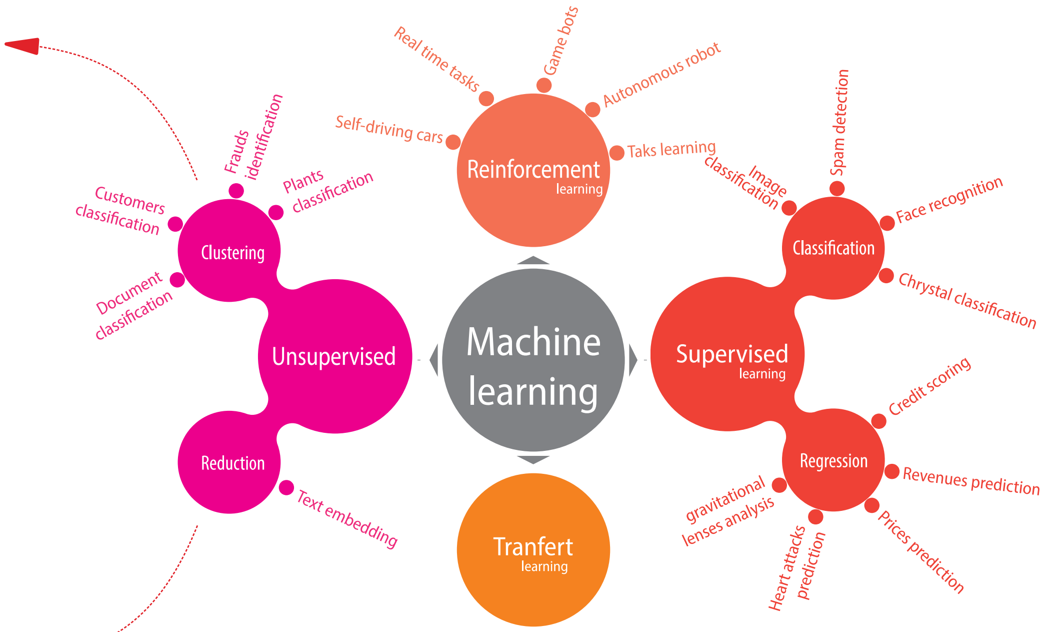
Clustering :
Finding Common Relationships

What is the relationship between these data ?

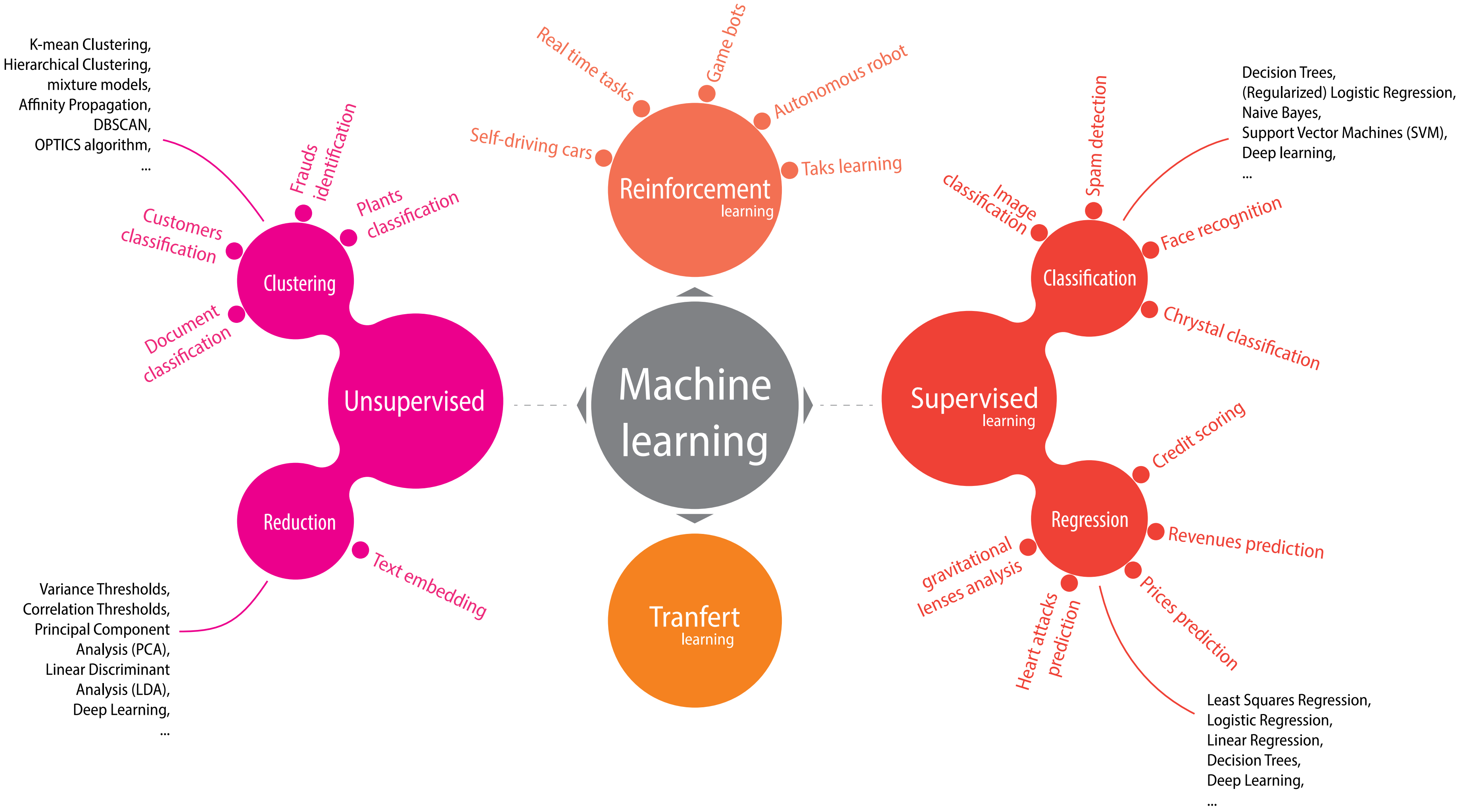


Reduction :
Reduce the number of dimensions

Simplify while keeping meaning



[*-learning]



Machine learning
Reinforcement learning
Real time tasks
Game bots
Autonomous robot
Self-driving cars
Taks learning
Tranfert learning
Unsupervised
Clustering
Frauds identification
Plants classification
Customers classification
Document classification
Reduction
Text embedding
K-mean Clustering, Hierarchical Clustering, mixture models, Affinity Propagation, DBSCAN, OPTICS algorithm, ...
Variance Thresholds, Correlation Thresholds, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Deep Learning, ...
Supervised learning
Classification
Image classification
Spam detection
Face recognition
Chrystal classification
Decision Trees, (Regularized) Logistic Regression, Naive Bayes, Support Vector Machines (SVM), Deep learning, ...
Regression
Credit scoring
Revenues prediction
Prices prediction
Heart attacks prediction
gravitational lenses analysis
Least Squares Regression, Logistic Regression, Linear Regression, Decision Trees, Deep Learning, ...

A Python centered world

« *Deep learning for humans* »

Widely used in the implementation of practical solutions

Widely used in the field of AI research

Keras

By François Cholet (Google)
Hight level API
Part on TensorFlow since 2017
MIT licence

TensorFlow

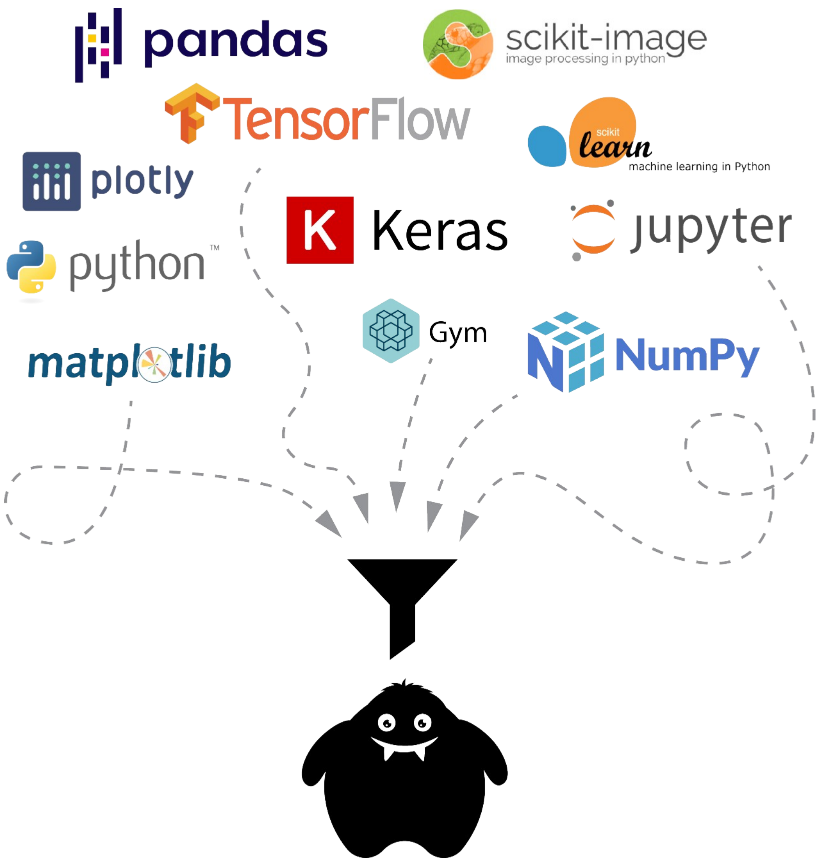
Most used DL framework
Supported by Google
Low level API – an hard way
Apache licence

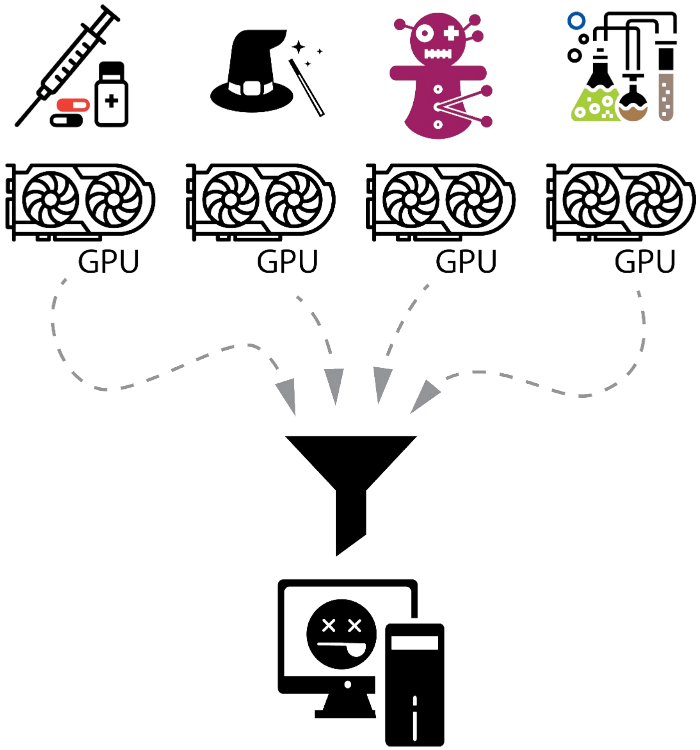
PYTORCH

From Torch library
Supported by Facebook
BSD licence

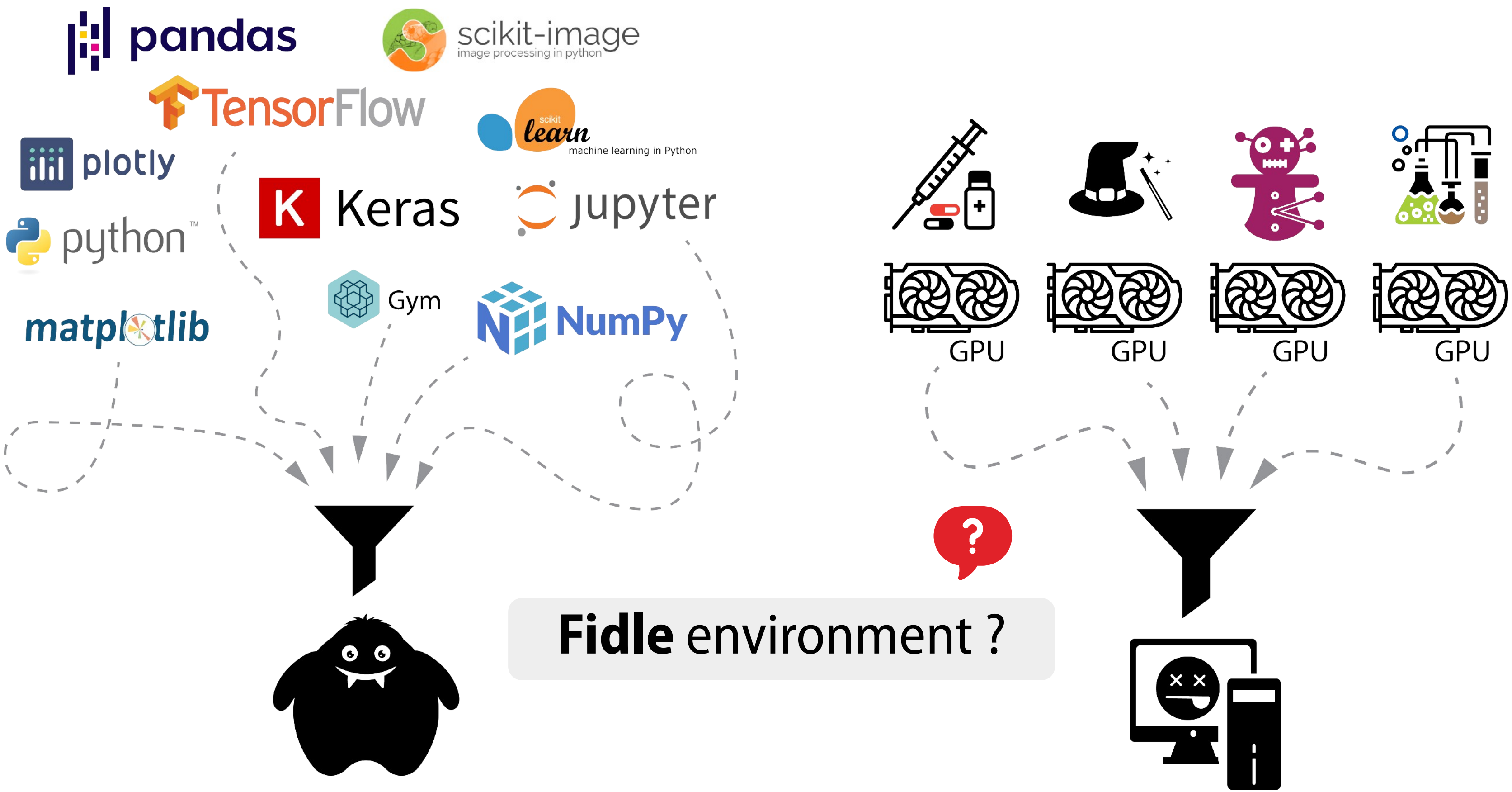
python™ NumPy jupyter

A certain complexity...

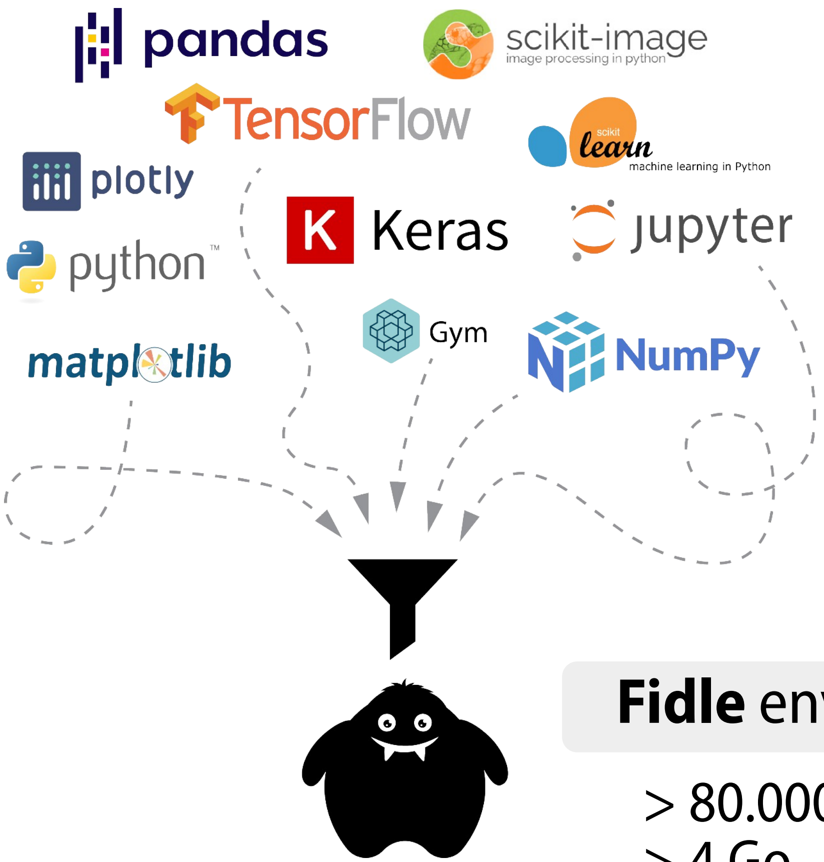


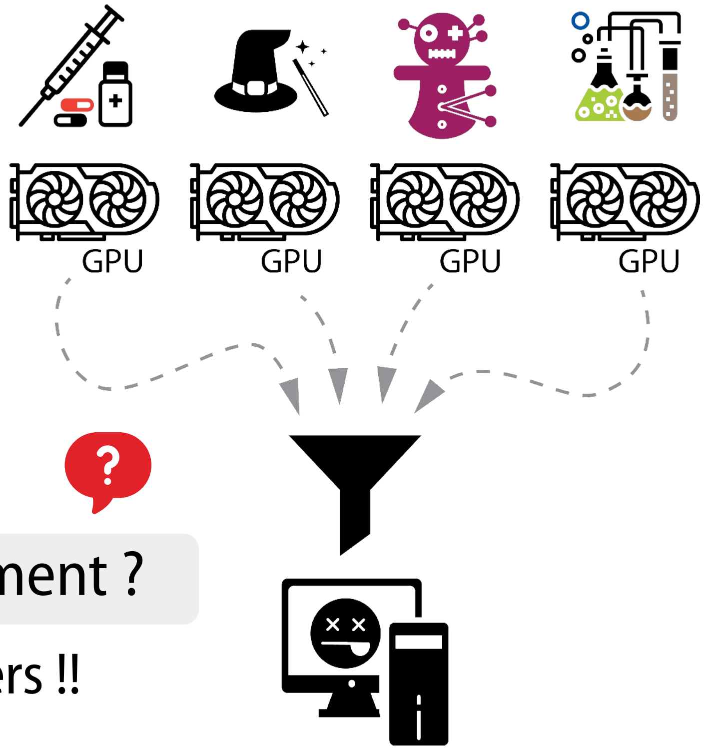


A certain complexity...



A certain complexity...

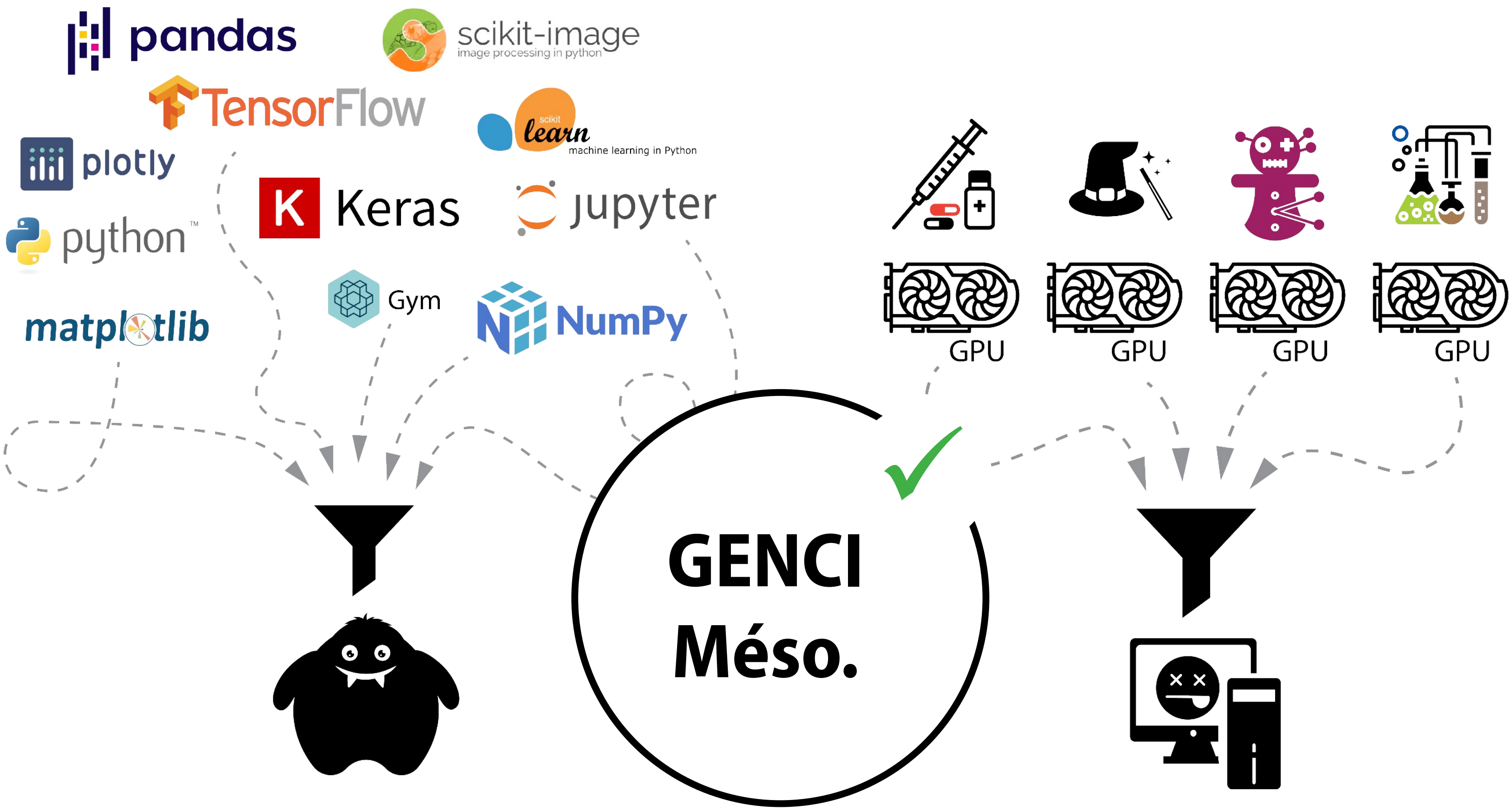


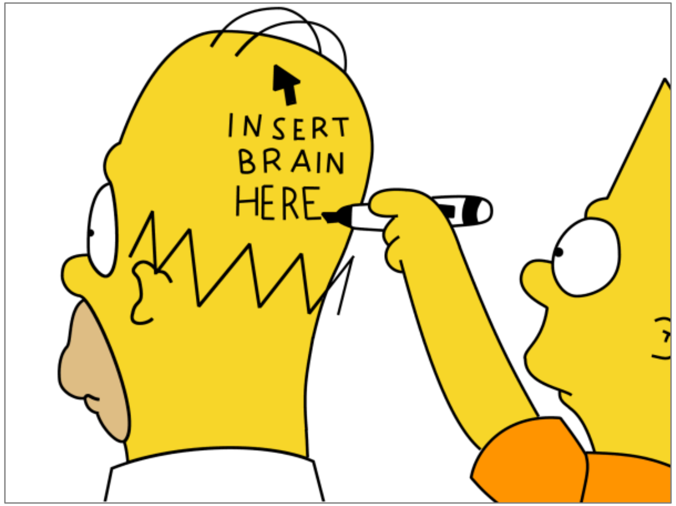


Fidle environment ?

> 80.000 fichiers !!
> 4 Go

A certain complexity...

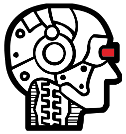




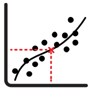
Fine, but Deep Learning What's that?

Roadmap

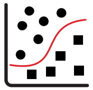
1



History, Fundamental Concepts



Basic Regression
DNN



Basic Classification
DNN

1.1 Introduction
Context, tools and ressources

1.2 **From the liner regression to the first neuron**

1.3 **Neurons in controversy**

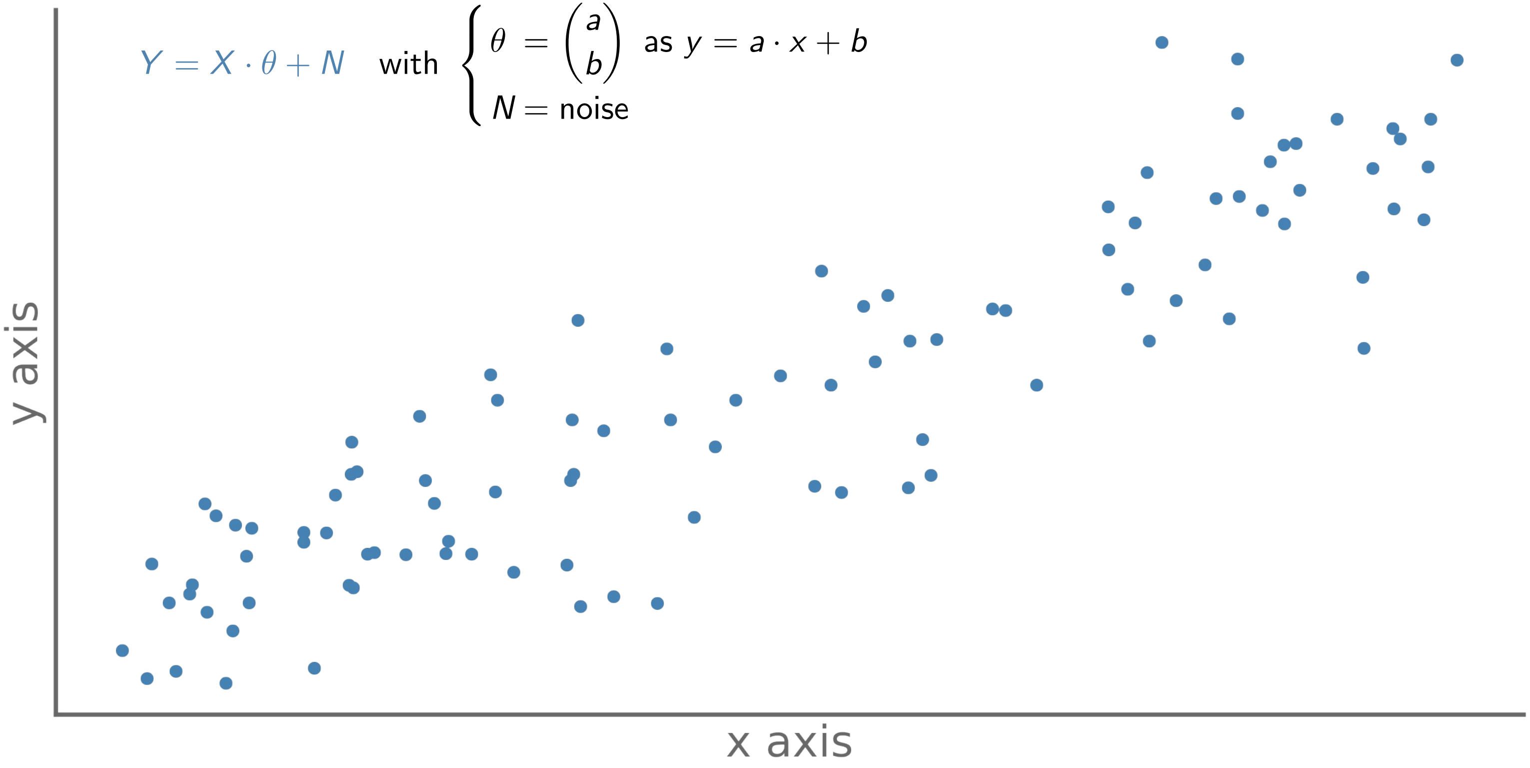
1.4 **Data and neurons**

- Basic Regression
- Basic Classification

Linear regression

We have a phenomenon, for which we have observations

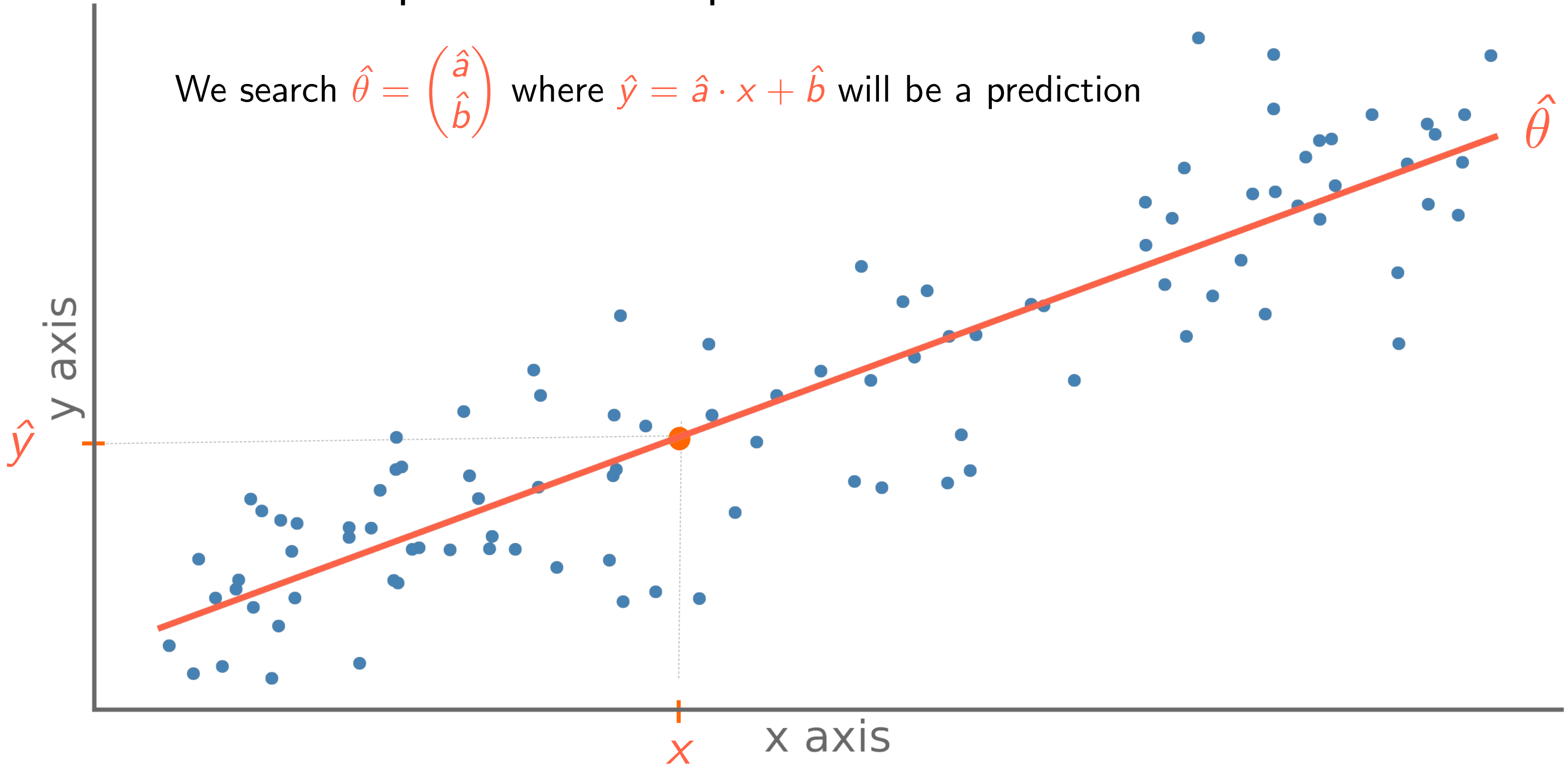
$$Y = X \cdot \theta + N \quad \text{with} \quad \begin{cases} \theta = \begin{pmatrix} a \\ b \end{pmatrix} \text{ as } y = a \cdot x + b \\ N = \text{noise} \end{cases}$$



Linear regression

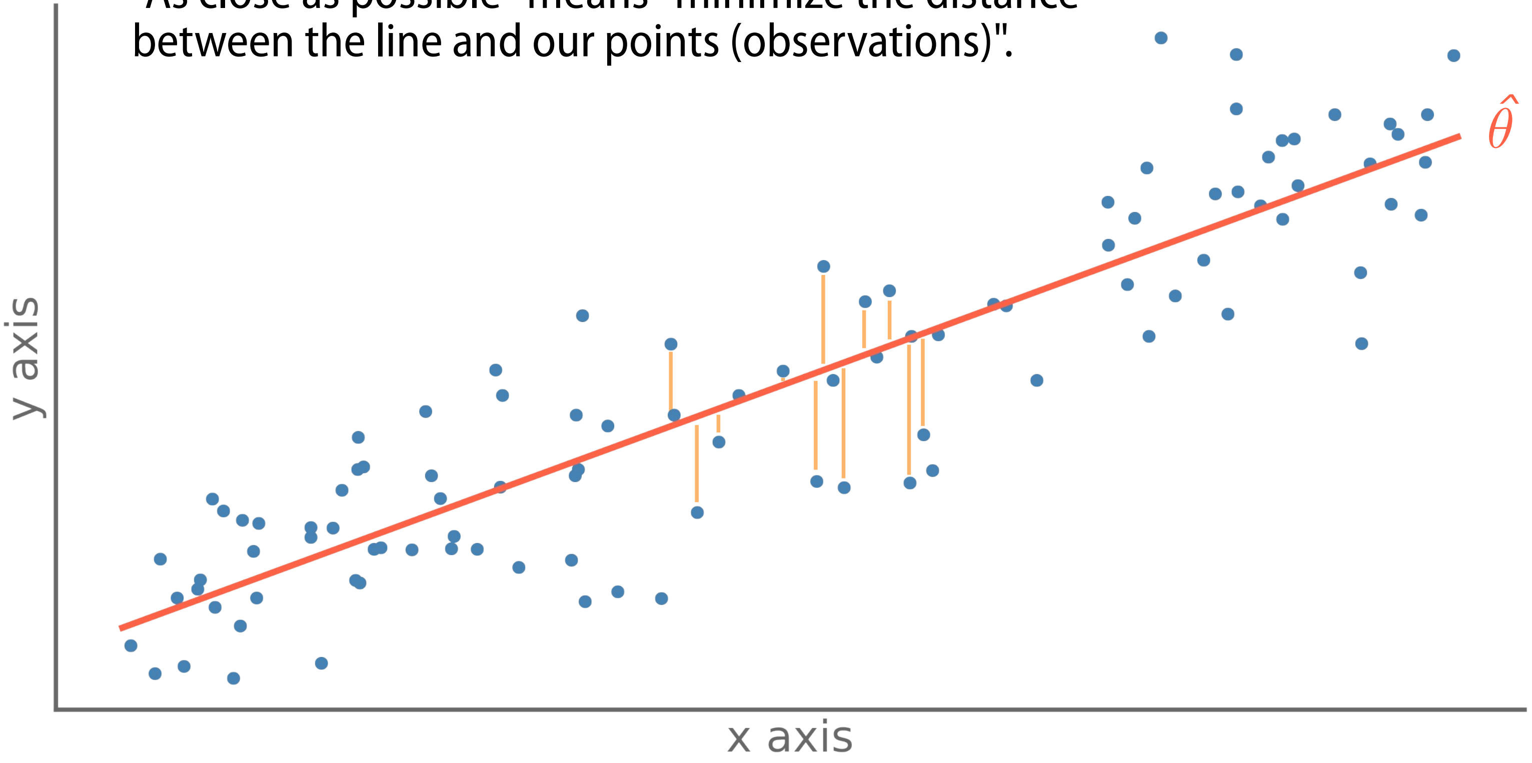
We are looking for a straight line that passes
« as close as possible » to our points.

We search $\hat{\theta} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$ where $\hat{y} = \hat{a} \cdot x + \hat{b}$ will be a prediction



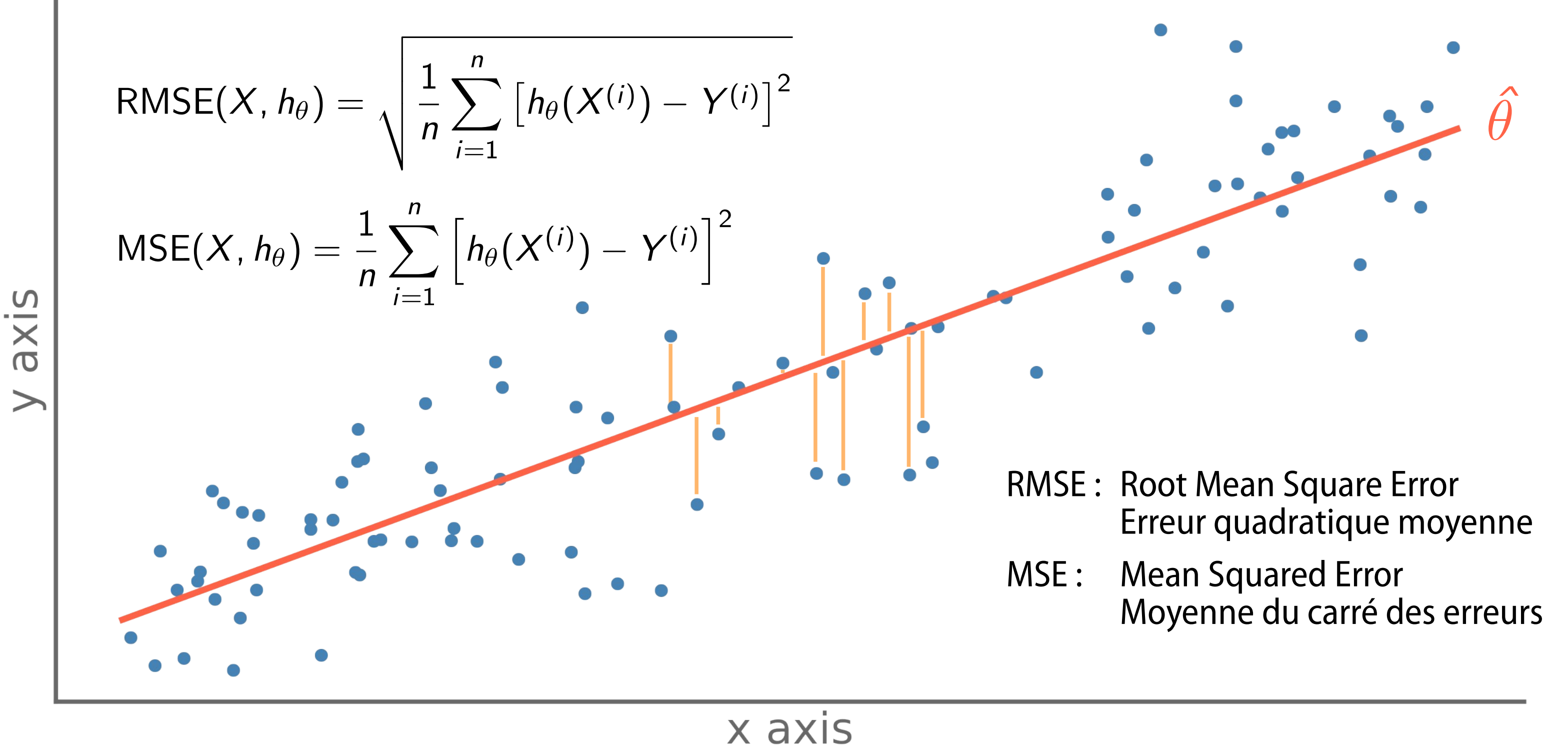
Linear regression

"As close as possible" means "minimize the distance between the line and our points (observations)".



Linear regression

For this, we will use an «loss function », which we will try to minimize.



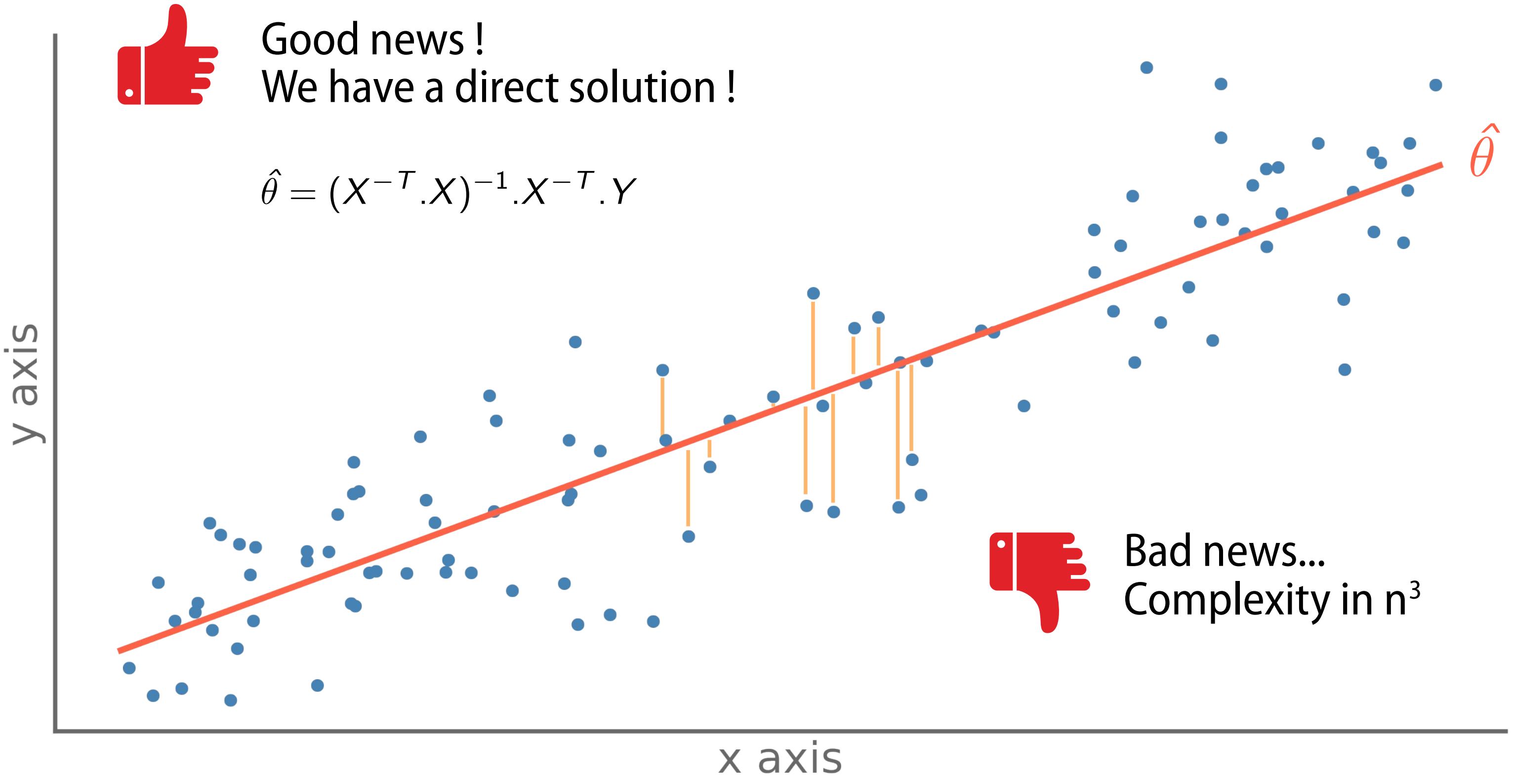
$$\mathrm{RMSE}(X, h_\theta) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[h_\theta(X^{(i)}) - Y^{(i)}\right]^2}$$

$$\mathrm{MSE}(X, h_\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[h_\theta(X^{(i)}) - Y^{(i)}\right]^2$$

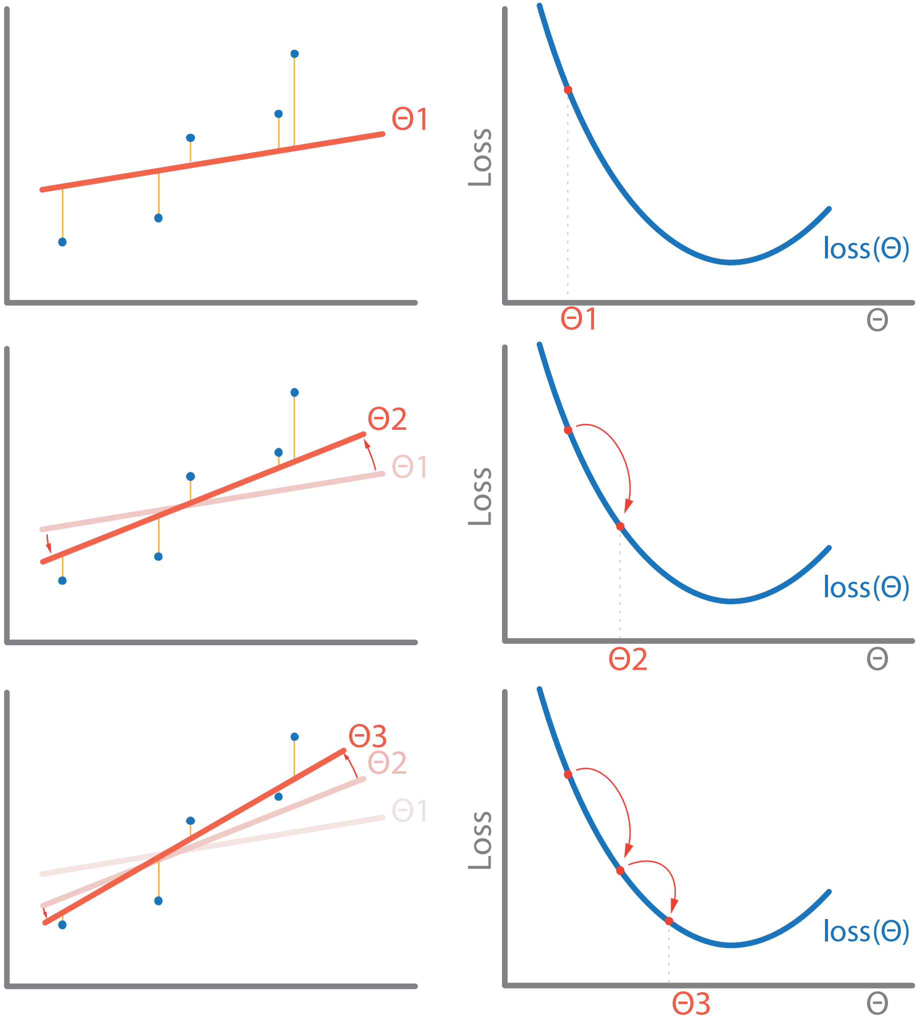
RMSE : Root Mean Square Error
Erreur quadratique moyenne

MSE : Mean Squared Error
Moyenne du carré des erreurs

Linear regression



Gradient descent



We will iteratively look for the best position of our line, by varying its parameters (Θ).

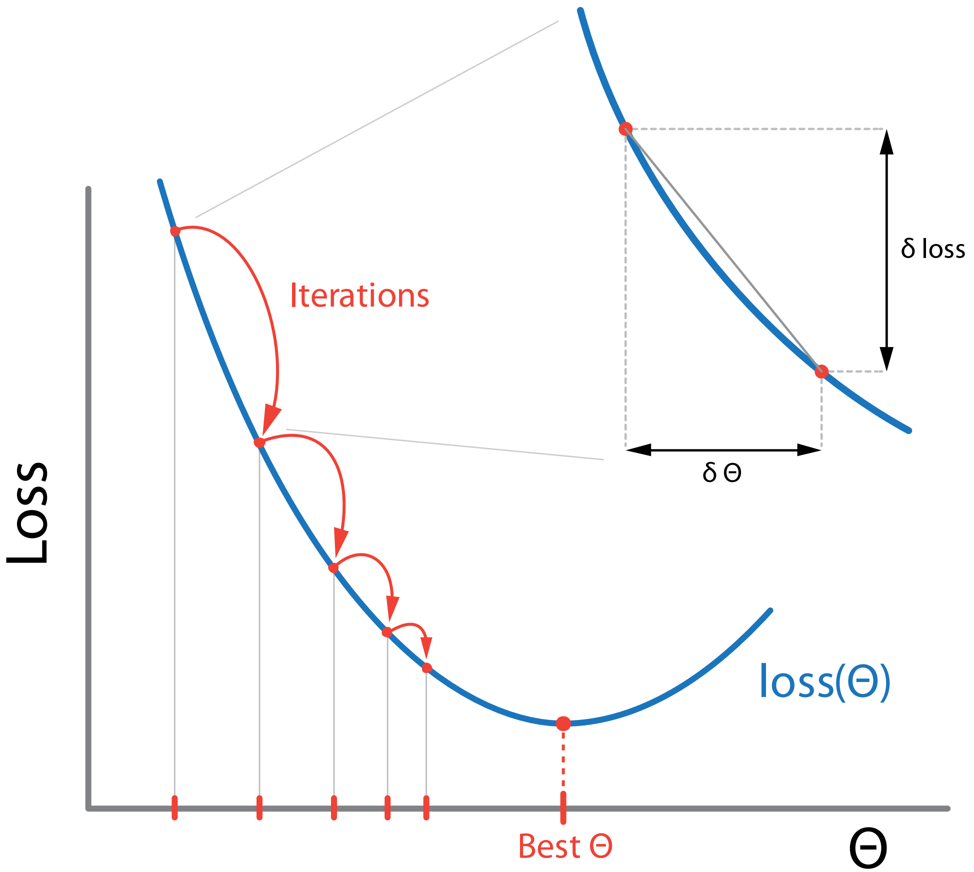
But how can we efficiently vary our parameters (Θ)?

Note : Loss functions could be :

$$\mathrm{RMSE}(X, h_\theta) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[h_\theta(X^{(i)}) - Y^{(i)}\right]^2}$$

$$\mathrm{MSE}(X, h_\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[h_\theta(X^{(i)}) - Y^{(i)}\right]^2$$

Gradient descent



By changing Θ from δΘ
We improve loss(Θ) of δloss

The gradient is the slope we will follow to minimize our loss function.

$$\text{gradient} = \frac{\delta \text{loss}}{\delta \theta}$$

One iterative solution is : $\theta \leftarrow \theta - \eta \cdot \frac{\delta \text{loss}}{\delta \theta}$

where η is the learning rate

This process is called gradient descent and the function used to optimize the descent, optimization function

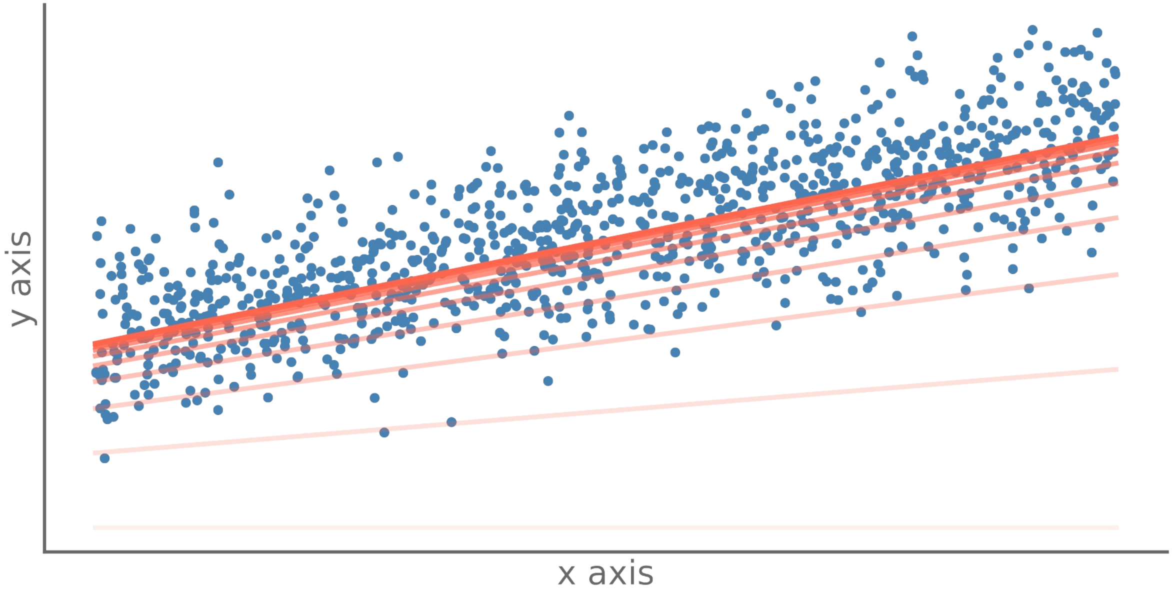
Gradient descent

$$MSE(X, h_\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[h_\theta(X^{(i)}) - Y^{(i)}\right]^2$$

$$\nabla_\theta MSE(\Theta) = \begin{bmatrix} \frac{\partial}{\partial\theta_0} MSE(\Theta) \\ \frac{\partial}{\partial\theta_1} MSE(\Theta) \\ \vdots \\ \frac{\partial}{\partial\theta_m} MSE(\Theta) \end{bmatrix} = \frac{2}{n} X^T \cdot (X \cdot \Theta - Y)$$

Iterative solution is : $\Theta \leftarrow \Theta - \eta \cdot \nabla_\theta MSE(\Theta)$
where η is the learning rate

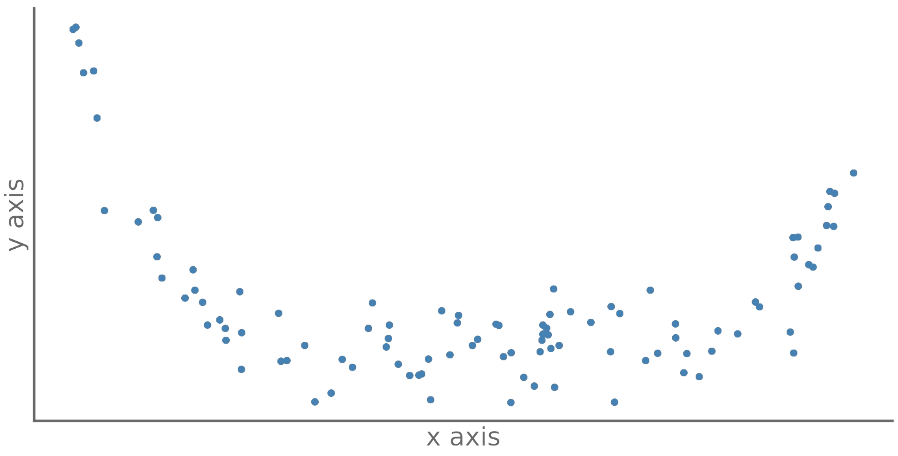
n : number of observations
m : number of characteristics

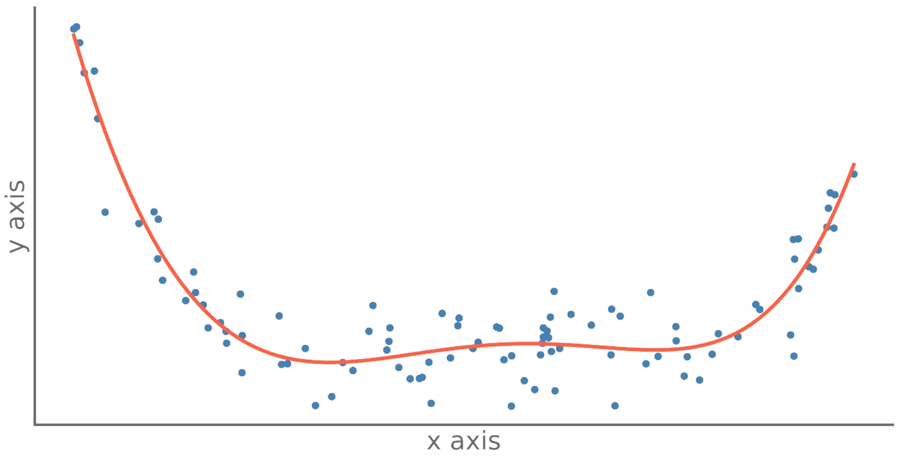


#i	Loss	Gradient		Theta	
0	+12.481	-6.777	-1.732	-3.388	+0.000
20	+4.653	-4.066	-1.039	-2.033	+0.346
40	+1.835	-2.440	-0.624	-1.220	+0.554
60	+0.821	-1.464	-0.374	-0.732	+0.679
80	+0.455	-0.878	-0.224	-0.439	+0.754
100	+0.324	-0.527	-0.135	-0.263	+0.799
120	+0.277	-0.316	-0.081	-0.158	+0.826
140	+0.260	-0.190	-0.048	-0.095	+0.842
160	+0.253	-0.114	-0.029	-0.057	+0.851
180	+0.251	-0.068	-0.017	-0.034	+0.857
200	+0.250	-0.041	-0.010	-0.020	+0.861

Notebook [GRAD1]

Polynomial regression

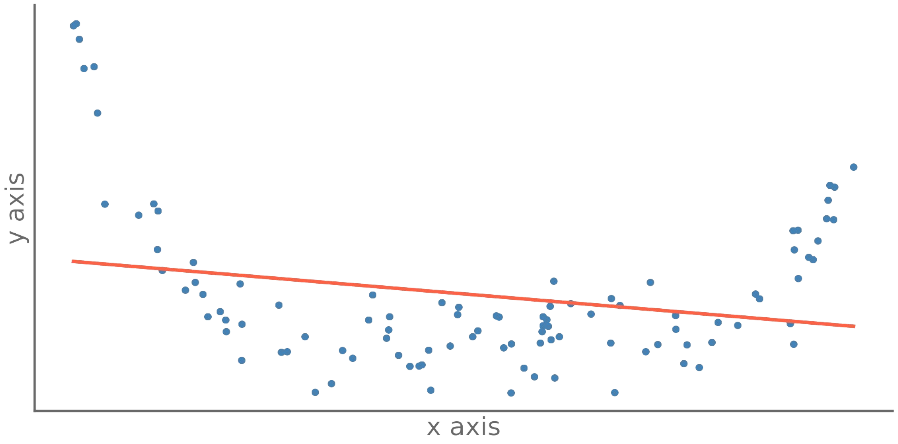




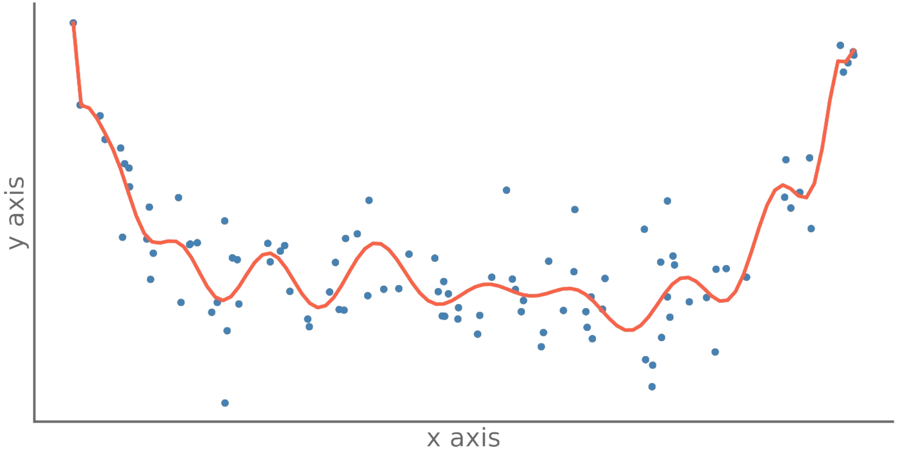
$$P_n(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \cdots + a_n \cdot x^n = \sum_{i=0}^{n} a_i \cdot x^n$$

Notebook [POLR1]

Polynomial regression



Underfiting

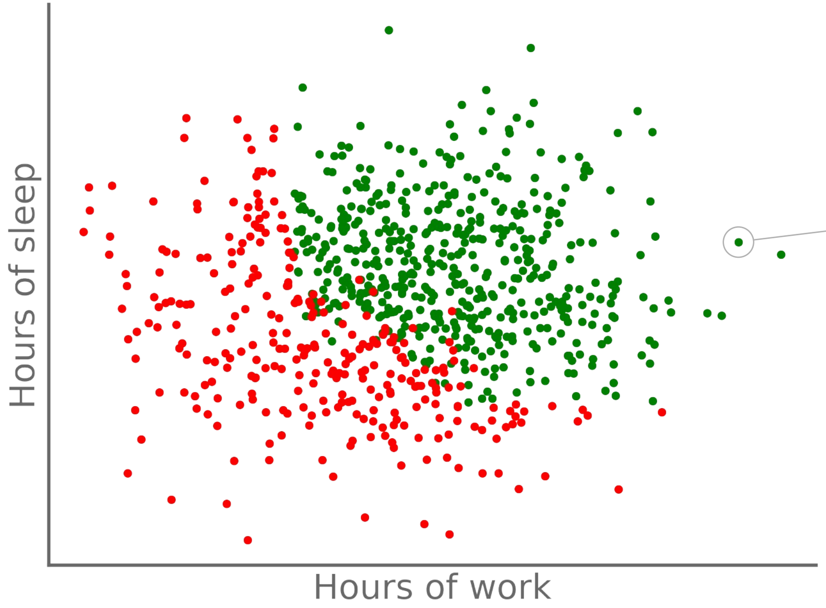


Overfiting

Notebook [POLR1]

A logistic regression is intended to provide a probability of belonging to a class.

Dataset : X Observations
y Classe



$$(X_i, y_i) \begin{cases} X_i = \begin{pmatrix} x_{i1} = \text{Hours of work} \\ x_{i2} = \text{Hours of sleep} \end{pmatrix} \\ y_i = \begin{cases} 1 & \text{belong to the class} \\ 0 & \text{don't belong} \end{cases} \end{cases}$$

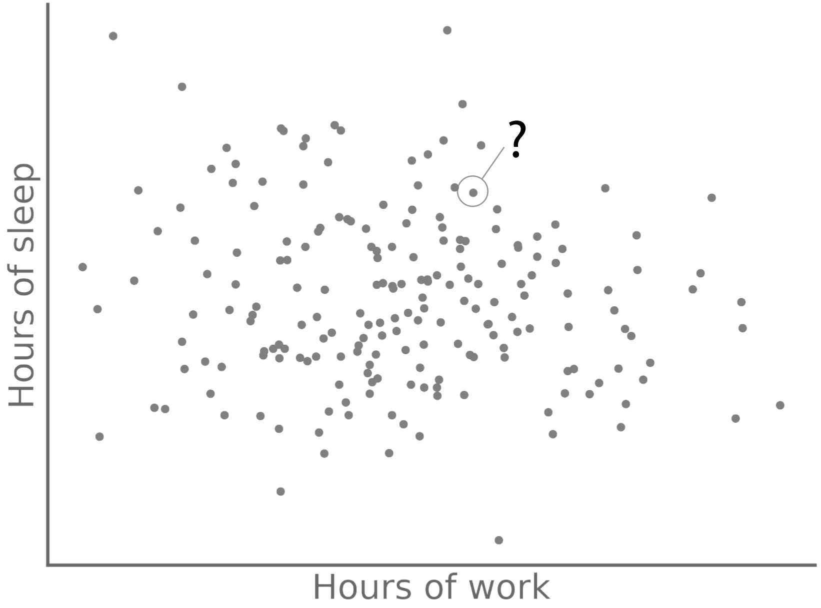
Notebook [LOGR1]

Logistic regression

A logistic regression is intended to provide a probability of belonging to a class.

Dataset : X Observations
y Classe

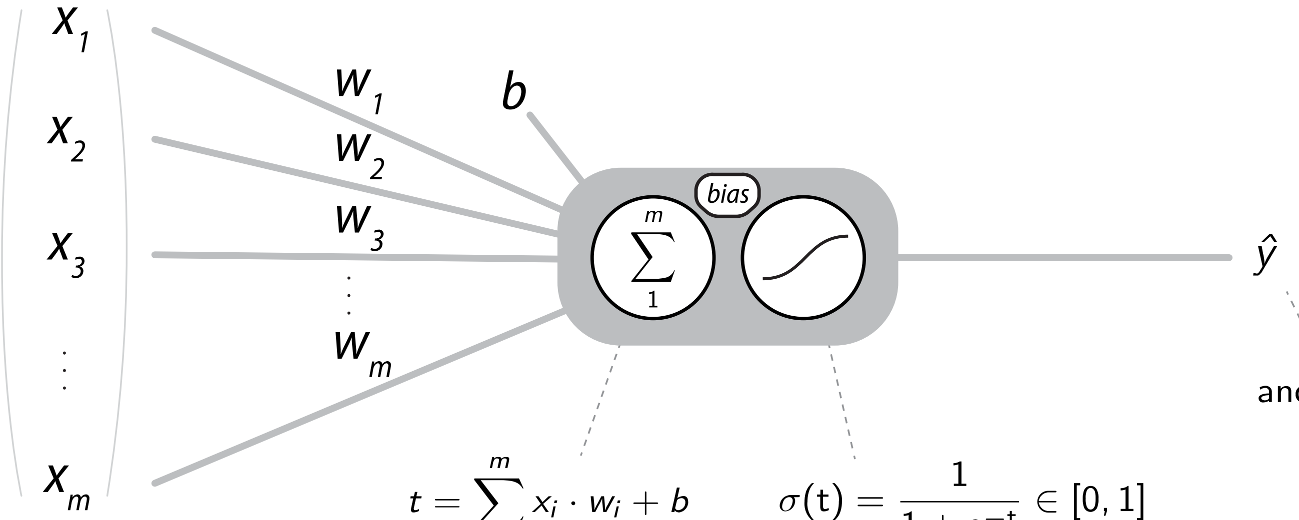
Objective : Predict the class
x given, we want to predict y
$\mathbf{y_{pred} = f(x)}$
where f is a linear function



Notebook [LOGR1]

Logistic regression

$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$



$$t = \sum_1^m x_i \cdot w_i + b$$

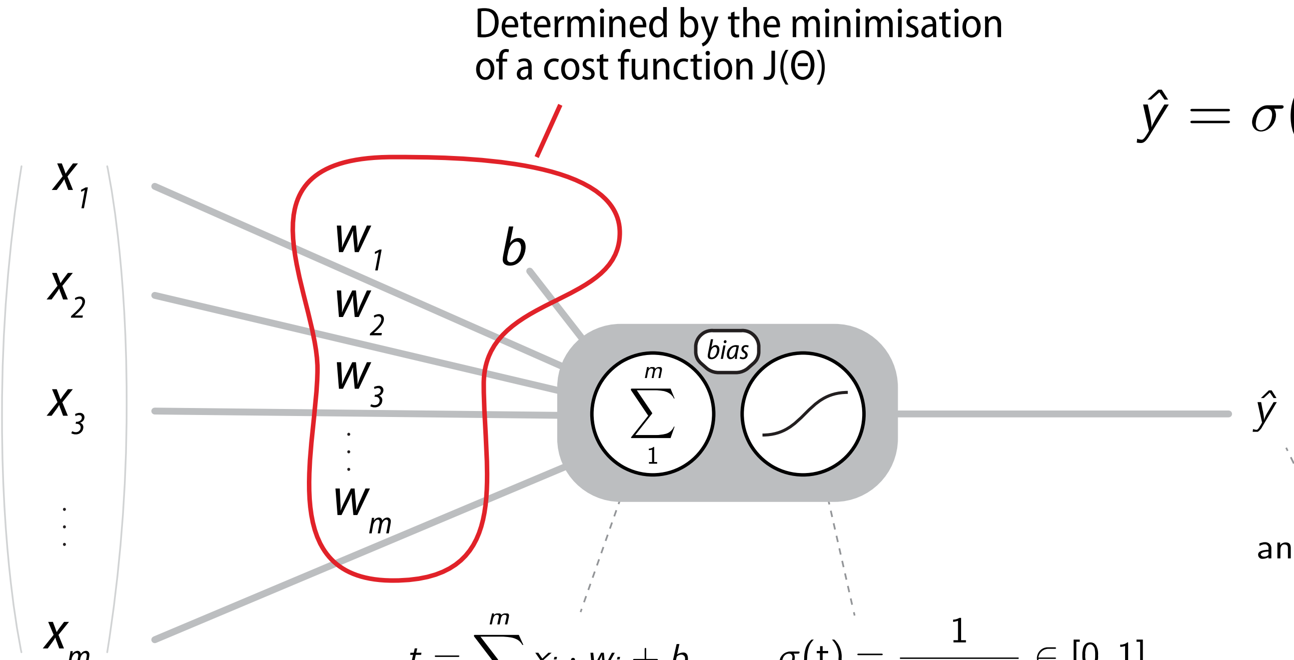
$$\sigma(t) = \frac{1}{1 + e^{-t}} \in [0, 1]$$

$$\text{and } \bar{y} = \begin{cases} 0 & \text{if } \hat{y} < 0.5 \\ 1 & \text{if } \hat{y} \geq 0.5 \end{cases}$$

Input	**Bias / Weight**	**Activation function**	**Output**
X	Θ	$\sigma(t)$	$\hat{y}$

Logistic regression

$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$



$$t = \sum_{1}^{m} x_i \cdot w_i + b$$

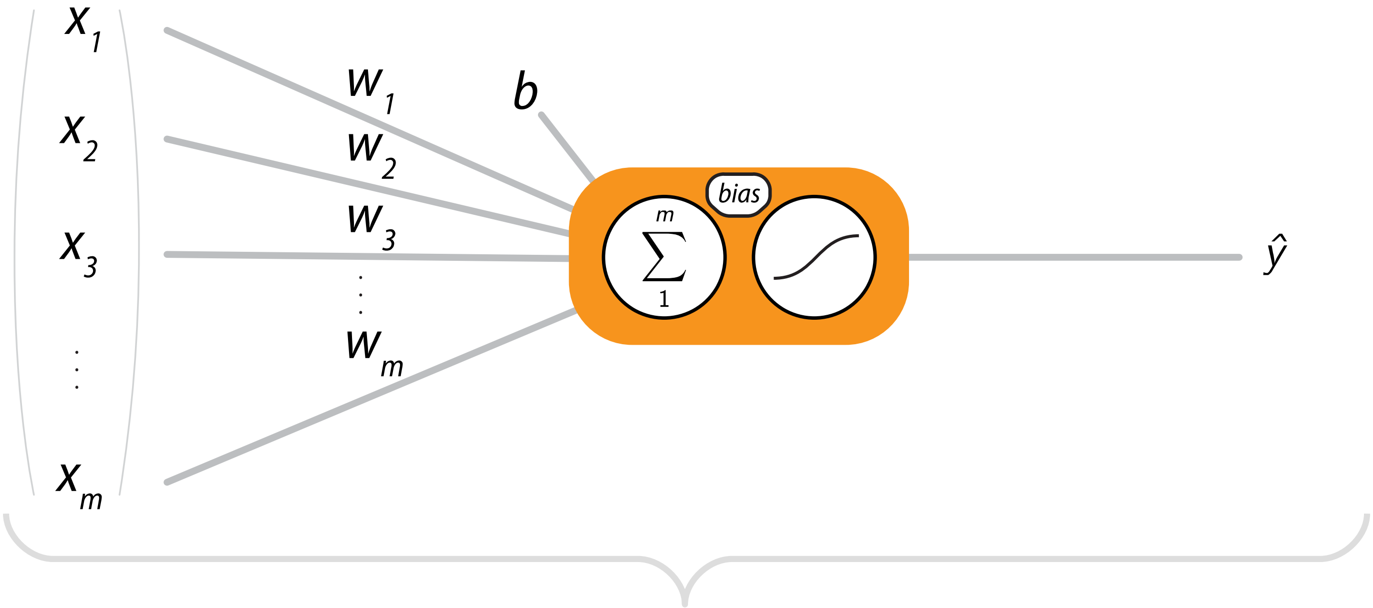
$$\sigma(\mathrm{t}) = \frac{1}{1 + \mathrm{e}^{-\mathrm{t}}} \in [0, 1]$$

and $\bar{y} = \begin{cases} 0 & \text{if } \hat{y} < 0.5 \\ 1 & \text{if } \hat{y} \geq 0.5 \end{cases}$

Input	Bias / Weight	Activation function	Output
X	Θ	$\sigma(t)$	$\hat{y}$

Logistic regression

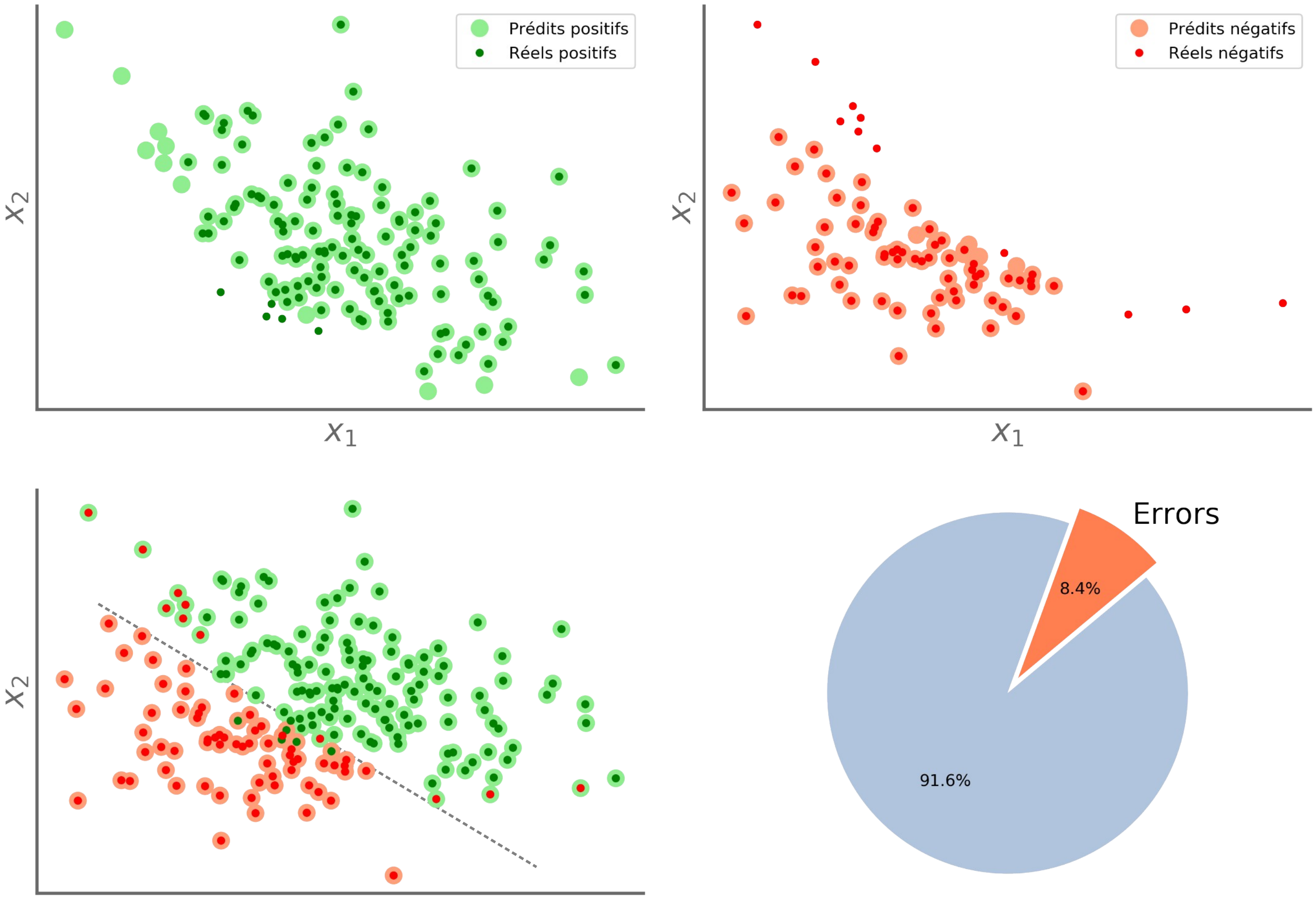
$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$



That's an « artificial neuron » !

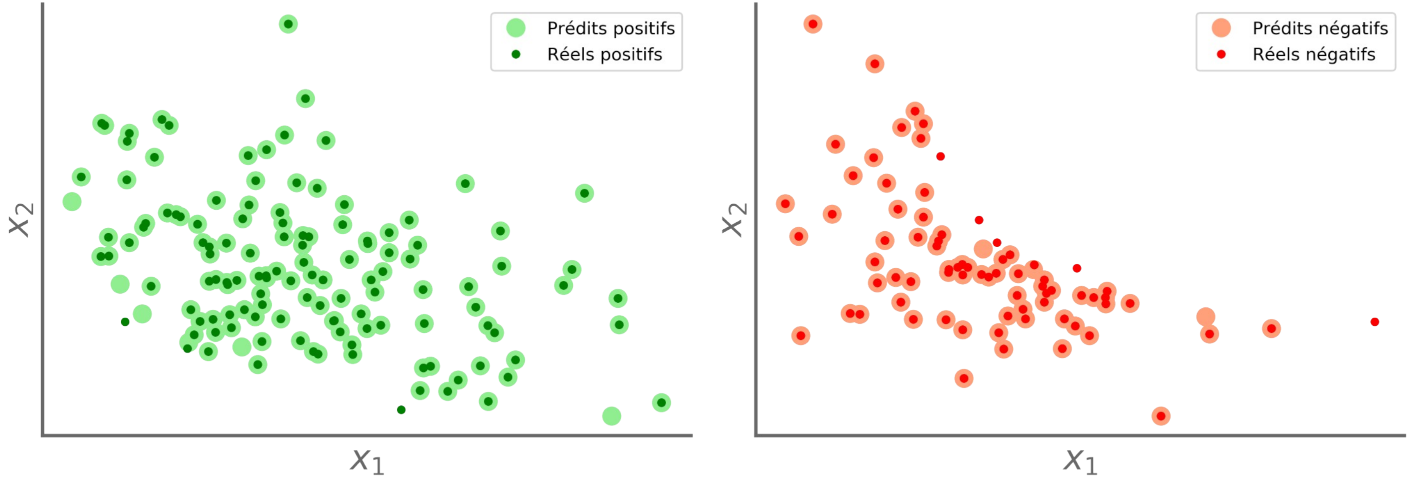
So, we have a neural network of… 1 neuron !

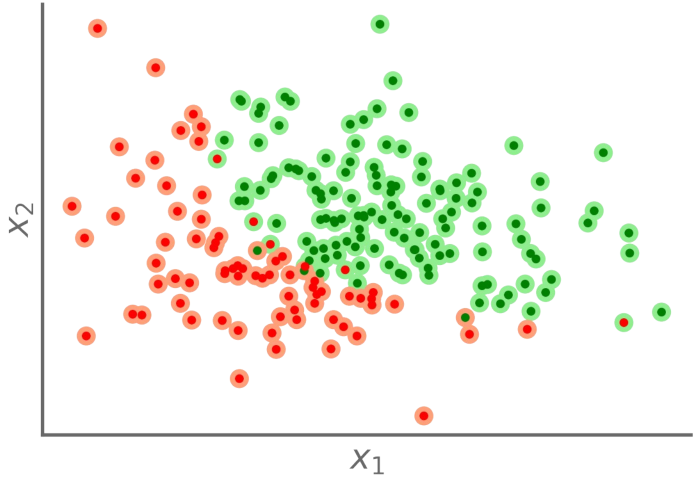
Logistic regression

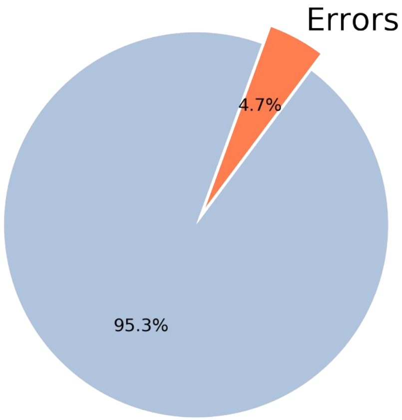


Notebook [LOGR1]

Logistic regression







Linear => Non linear

$\forall i \in [0, m]$, we add : $x_{i1}^2, x_{i2}^2, x_{i1}^3, x_{i2}^3$ to X_i

so, for :

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ \vdots & \dots & \\ 1 & x_{m1} & x_{m2} \end{bmatrix}$$

we have :

$$\mathring{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{12}^2 & x_{11}^3 & x_{12}^3 \\ \vdots & & & \dots & & & \\ 1 & x_{m1} & x_{m2} & x_{m1}^2 & x_{m2}^2 & x_{m1}^3 & x_{m2}^3 \end{bmatrix}$$

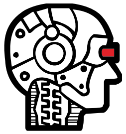
Notebook [LOGR1]

Théodore Géricault (1791 – 1824) - Le Radeau de La Méduse

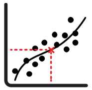
Help !
No more maths, please… ;-)

Roadmap

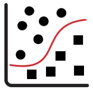
1



History, Fundamental Concepts



Basic Regression
DNN



Basic Classification
DNN

1.1 Introduction
Context, tools and ressources

1.2 **From the liner regression to the first neuron**

1.3 **Neurons in controversy**

1.4 **Data and neurons**

Basic Regression

Basic Classification

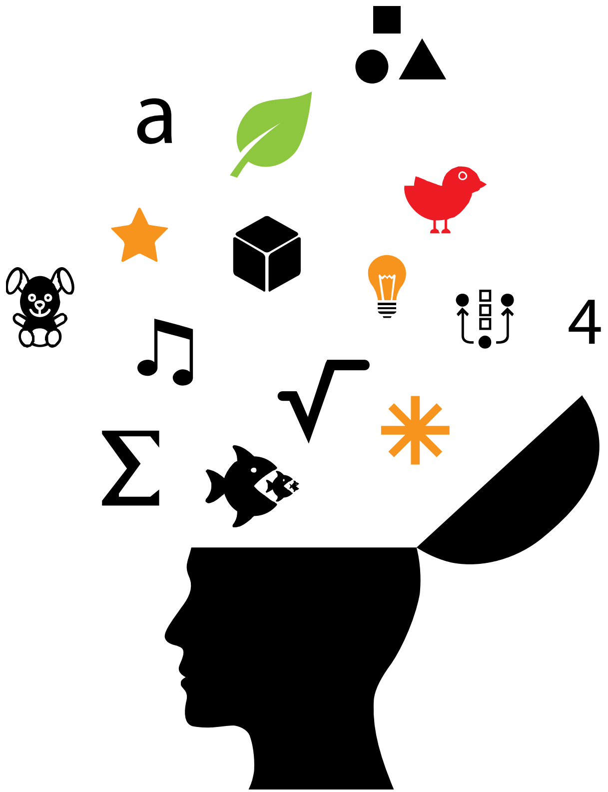
[intelligence]

[intelligence]

« Capacité de percevoir ou d'inférer l'information, et de la conserver comme une connaissance à appliquer à des comportements adaptatifs dans un environnement ou un contexte donné »

*« Ability to perceive or infer information, and to retain it as knowledge to be appli towards adaptive behaviors within an environment or context »**

* Wikipedia - Illustration : [POTATO]



[intelligence]

« Ensemble des fonctions mentales ayant pour objet la connaissance conceptuelle et rationnelle »*

« Set of mental functions aimed at conceptual and rational knowledge »

* https://www.larousse.fr

The big Controversy

Modelling the brain :

« Penser s'apparente
à un calcul massivement parallèle de
fonctions élémentaires.

L'information est un **signal** avant
d'être un code »[1]

Making a mind :

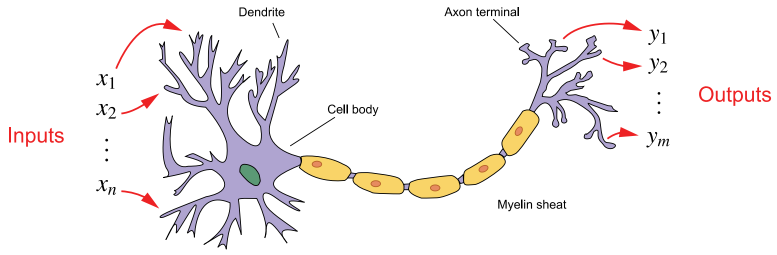
« Penser, c'est calculer des **symboles** qui
ont à la fois une réalité matérielle et une
valeur sémantique de représentation »[1]

L'information est une donnée
symbolique de **haut niveau**.

Connectionnism vs Symbolic

Modelling the brain
Modéliser le cerveau

Making a mind
Forger une opinion



Tout [homme] est [mortel]
[Socrate] est un [homme]

Donc [Socrate] est [mortel]

[1] Dominique Cardon, Jean-Philippe Cointet, Antoine Mazieres (2018) [LRDN]

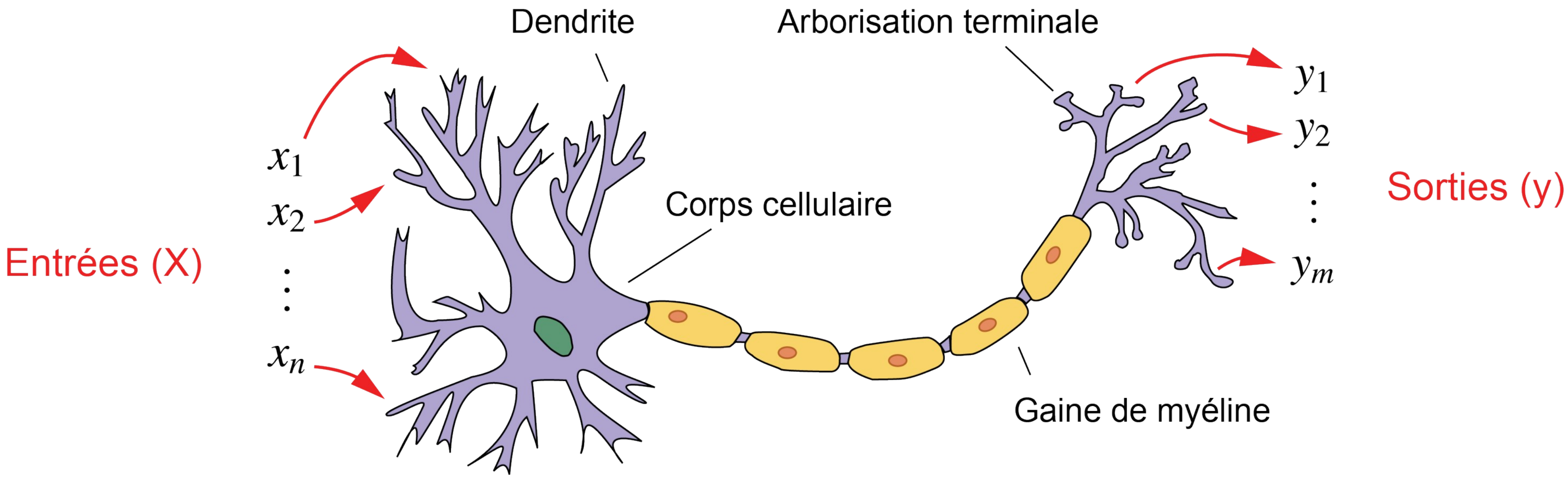
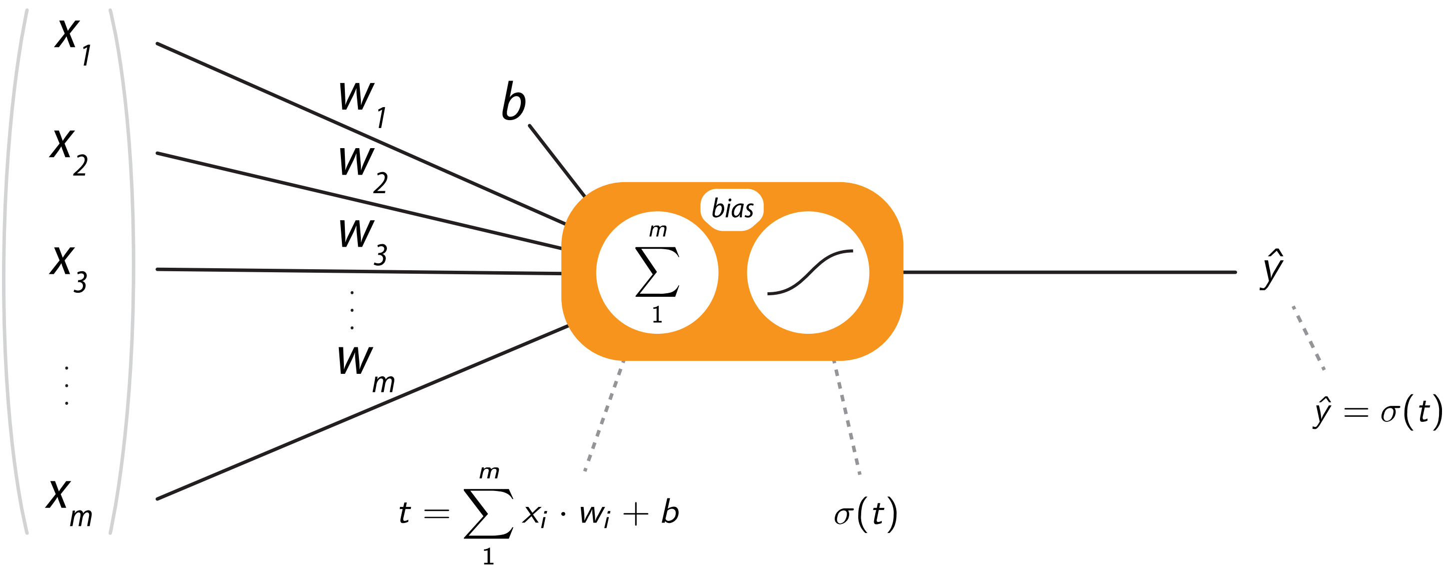


Illustration : [NEURON]

$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$



Input	Bias / Weight	Activation function	Output
X	Θ, b	$\sigma(t)$	$\hat{y}$

The big Controversy

Modelling the brain :

« Penser s'apparente
à un calcul massivement parallèle de
fonctions élémentaires.

L'information est un **signal** avant
d'être un code »[1]

Making a mind :

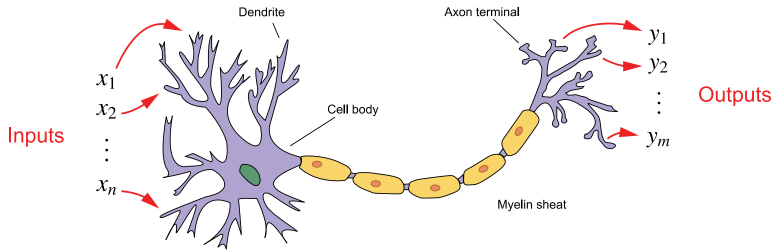
« Penser, c'est calculer des **symboles** qui
ont à la fois une réalité matérielle et une
valeur sémantique de représentation »[1]

L'information est une donnée
symbolique de **haut niveau**.

Connectionnism vs Symbolic

Modelling the brain
Modéliser le cerveau

Making a mind
Forger une opinion



Tout [homme] est [mortel]
[Socrate] est un [homme]

Donc [Socrate] est [mortel]

[1] D Cardon, JP Cointet, A Mazieres, 2018 [LRDN]

The big Controversy

Inductive approach

Input → Program (model)

Output →

Deductive approach

Input → Output

Program →

Connectionnism vs Symbolic

Facts ▶ Rules and laws

Model

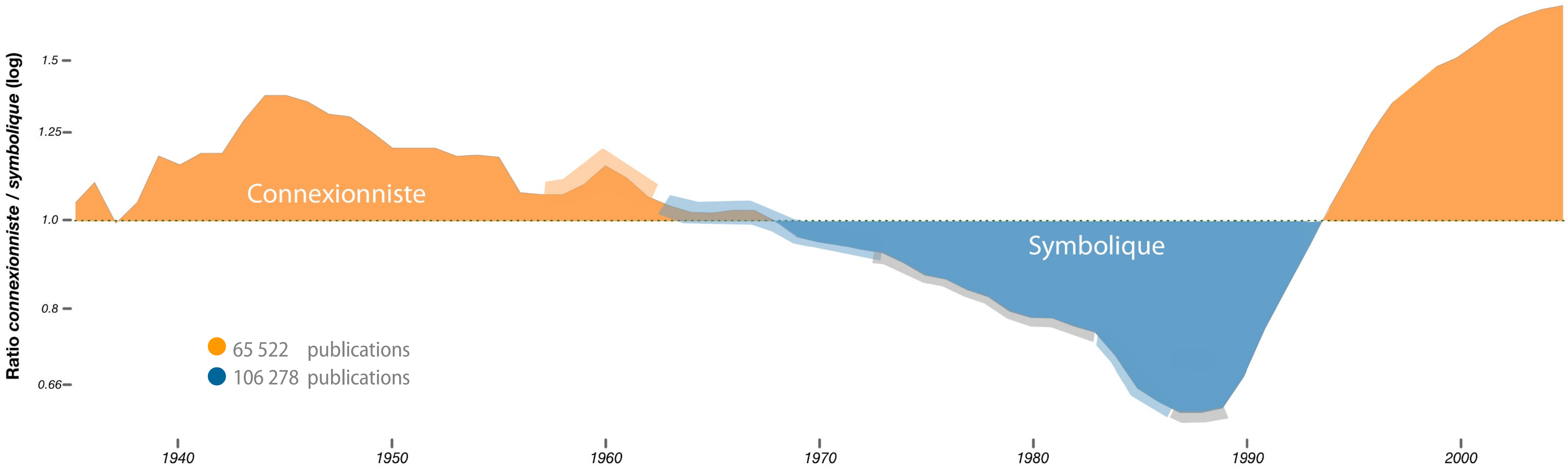
Expert

Rules and laws ▶ Special case

The big Controversy

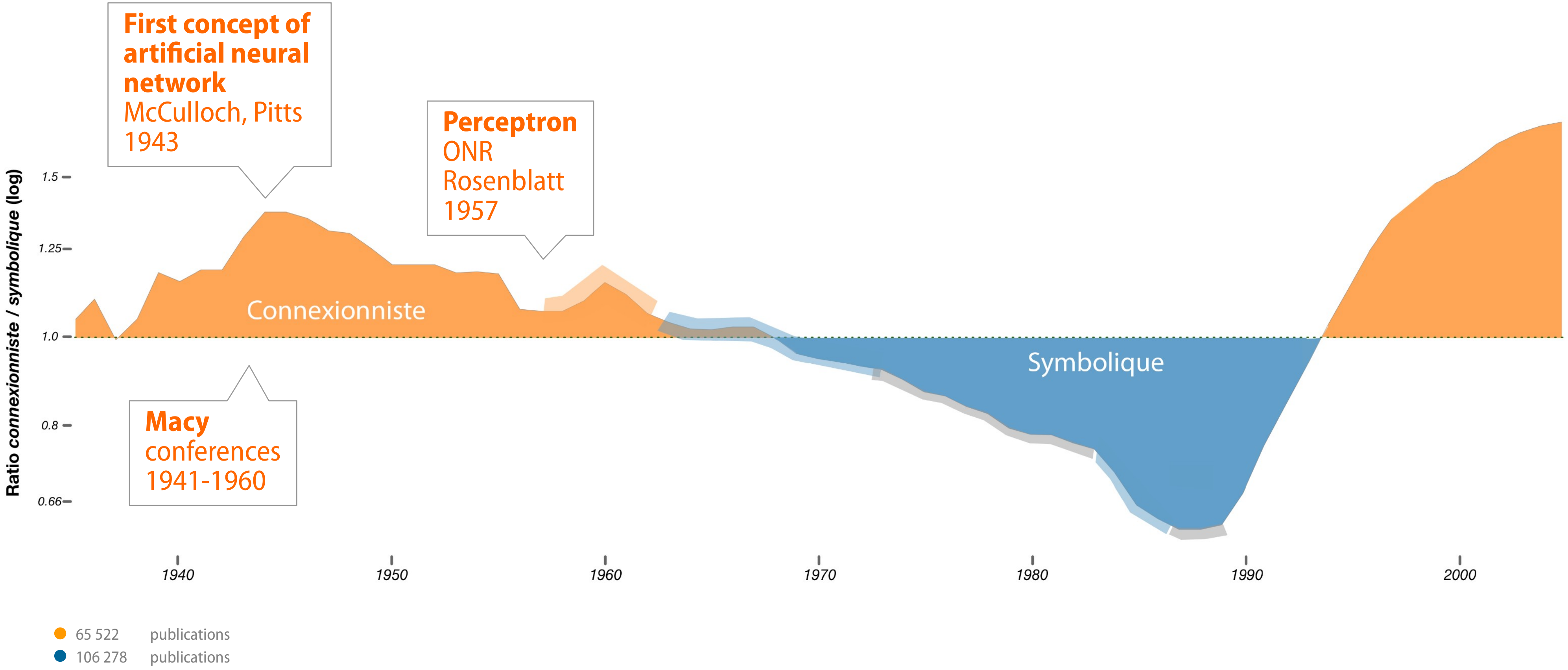
Evolution of the academic influence of connexionist and symbolic approaches[1]

Ration of publications between connexionists and symbolists



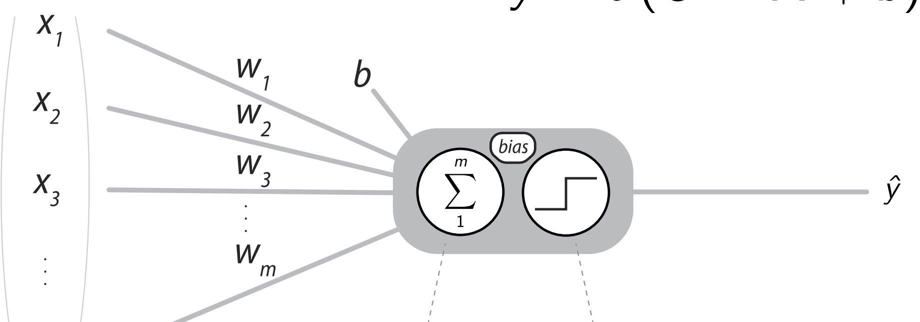
[1] D Cardon, JP Cointet, A Mazieres, 2018 [LRDN]

Evolution of the academic influence of connexionist and symbolic approaches[1]

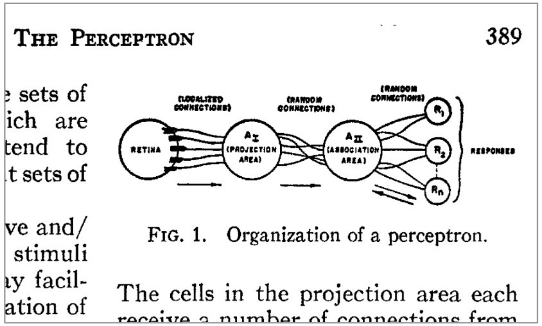


[1] D Cardon, JP Cointet, A Mazieres, 2018 [LRDN]

$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$



$$t = \sum_{1}^{m} x_i \cdot w_i + b \qquad \sigma(t) = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{if } t \leq 0 \end{cases}$$


THE PERCEPTRON 389

e sets of
ich are
tend to
t sets of

ve and/
stimuli
ay facil-
ation of

RETINA — (LOCALIZED CONNECTIONS) — A_I (PROJECTION AREA) — (RANDOM CONNECTIONS) — A_{II} (ASSOCIATION AREA) — (RANDOM CONNECTIONS) — R_1, R_2, R_n — RESPONSES

FIG. 1. Organization of a perceptron.

The cells in the projection area each
receive a number of connections from

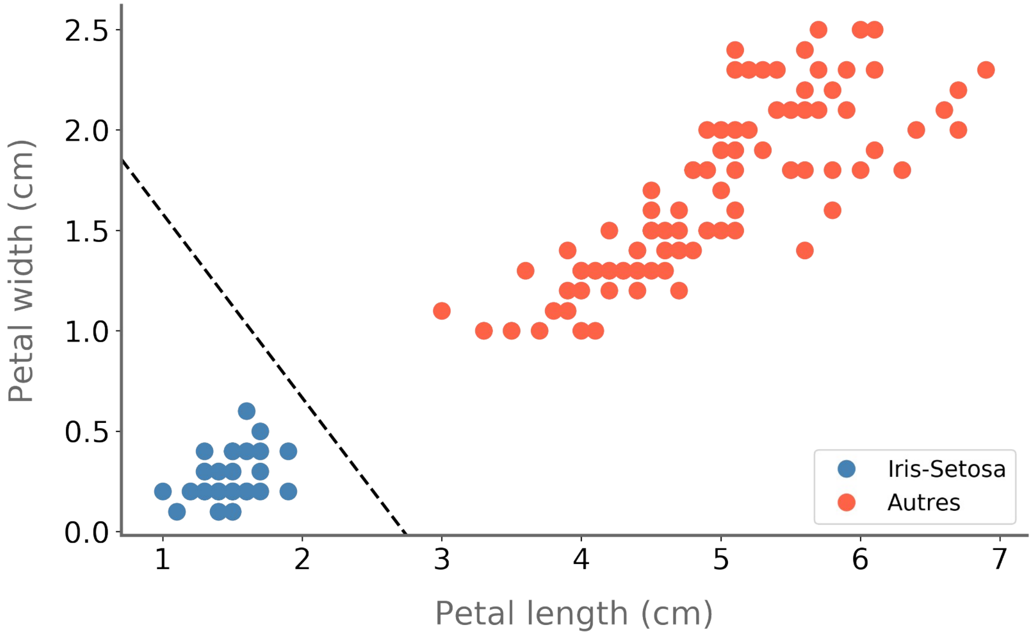
Perceptron
Frank Rosenblatt
1958

Linear and binary classifier

F. Rosenblatt, 1958 [FROS]

Iris plants dataset

Dataset from : Fisher, R.A. "The use of multiple measurements in taxonomic problems" Annual Eugenics, 7, Part II, 179-188 (1936)

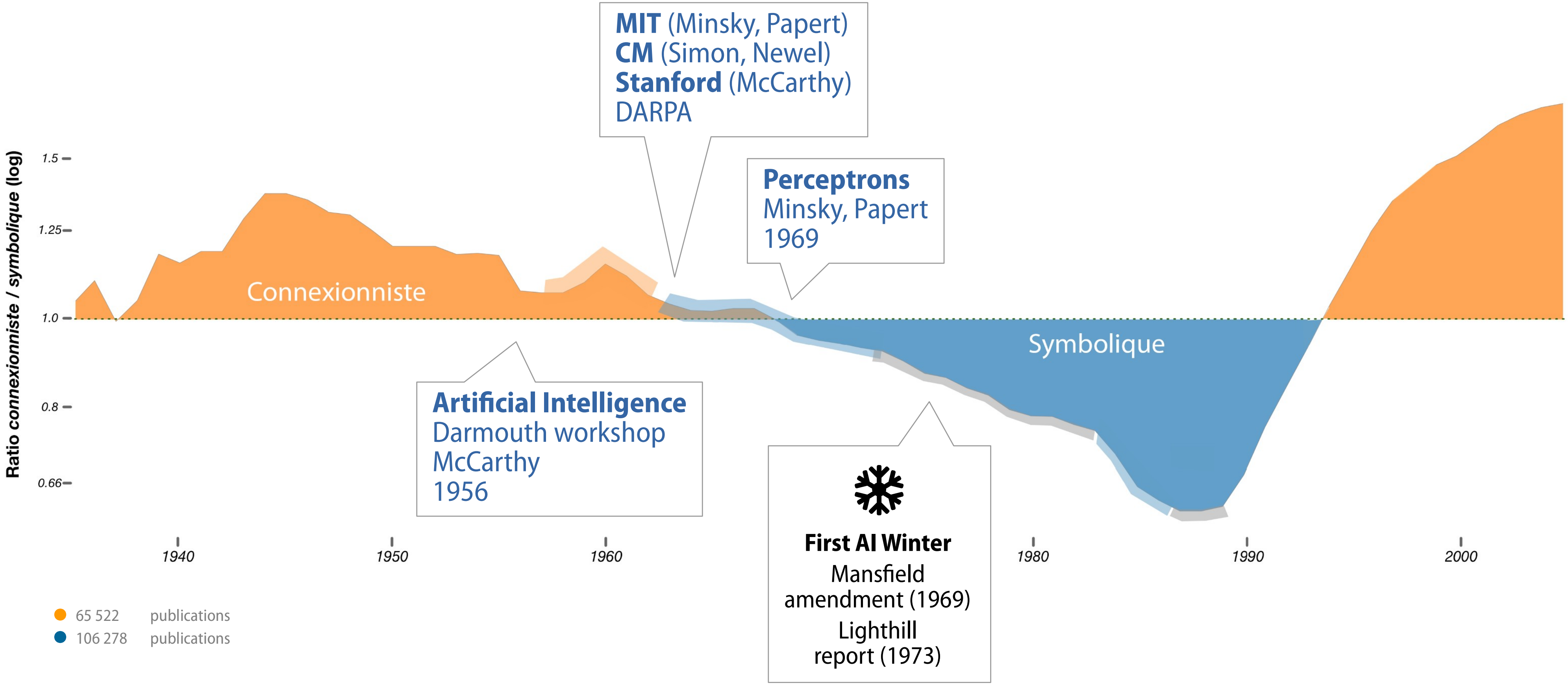


Length	Width	Iris Setosa (0/1)
x_1	x_2	y
1.4	1.4	1
1.6	1.6	1
1.4	1.4	1
1.5	1.5	1
1.4	1.4	1
4.7	4.7	0
4.5	4.5	0
4.9	4.9	0
4.0	4.0	0
4.6	4.6	0
(...)		

Notebook [PER57]

The big Controversy

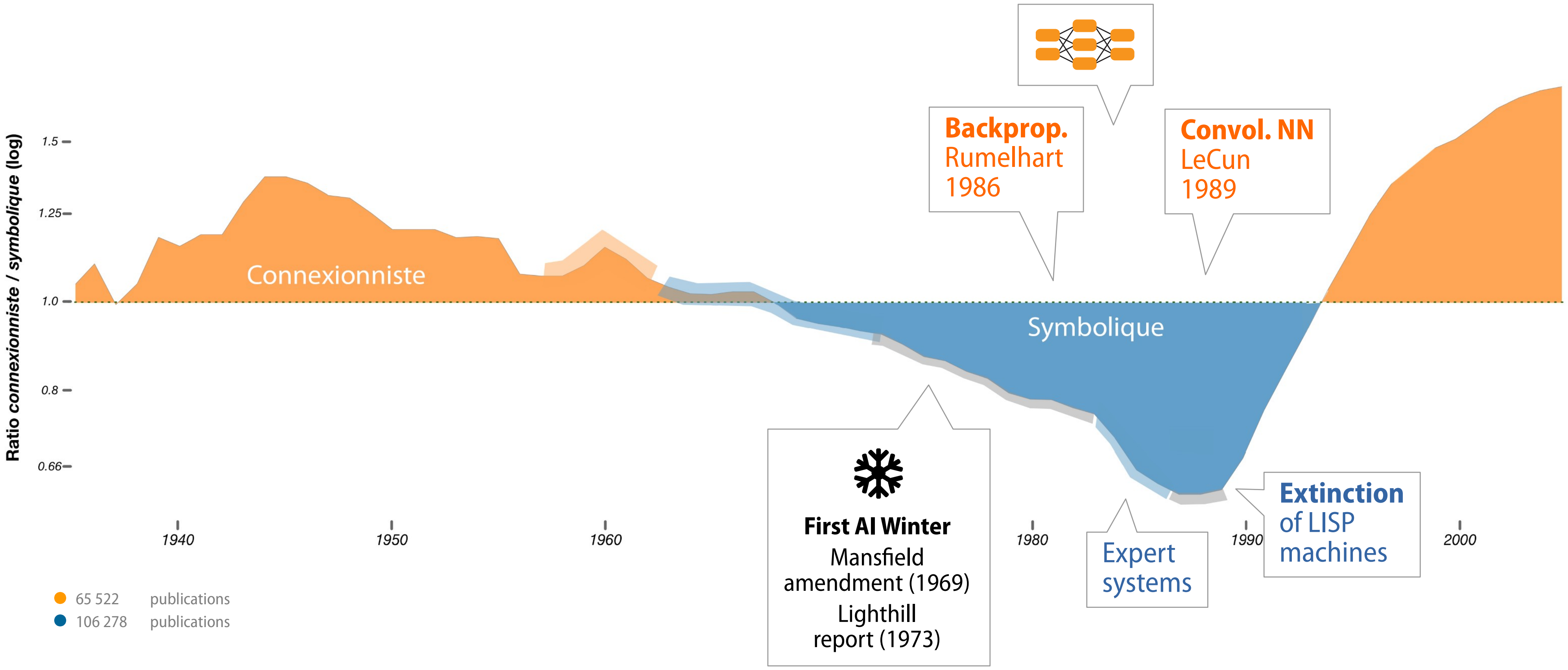
Evolution of the academic influence of connexionist and symbolic approaches[1]



[1] D Cardon, JP Cointet, A Mazieres, 2018 [LRDN]

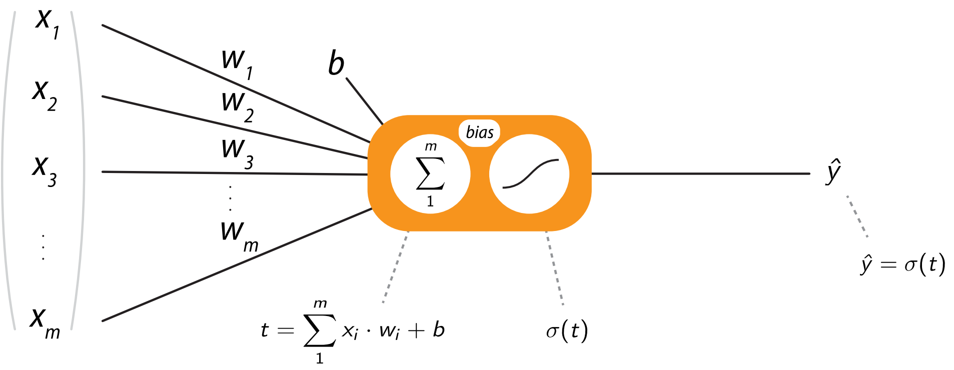
The big Controversy

Evolution of the academic influence of connexionist and symbolic approaches[1]

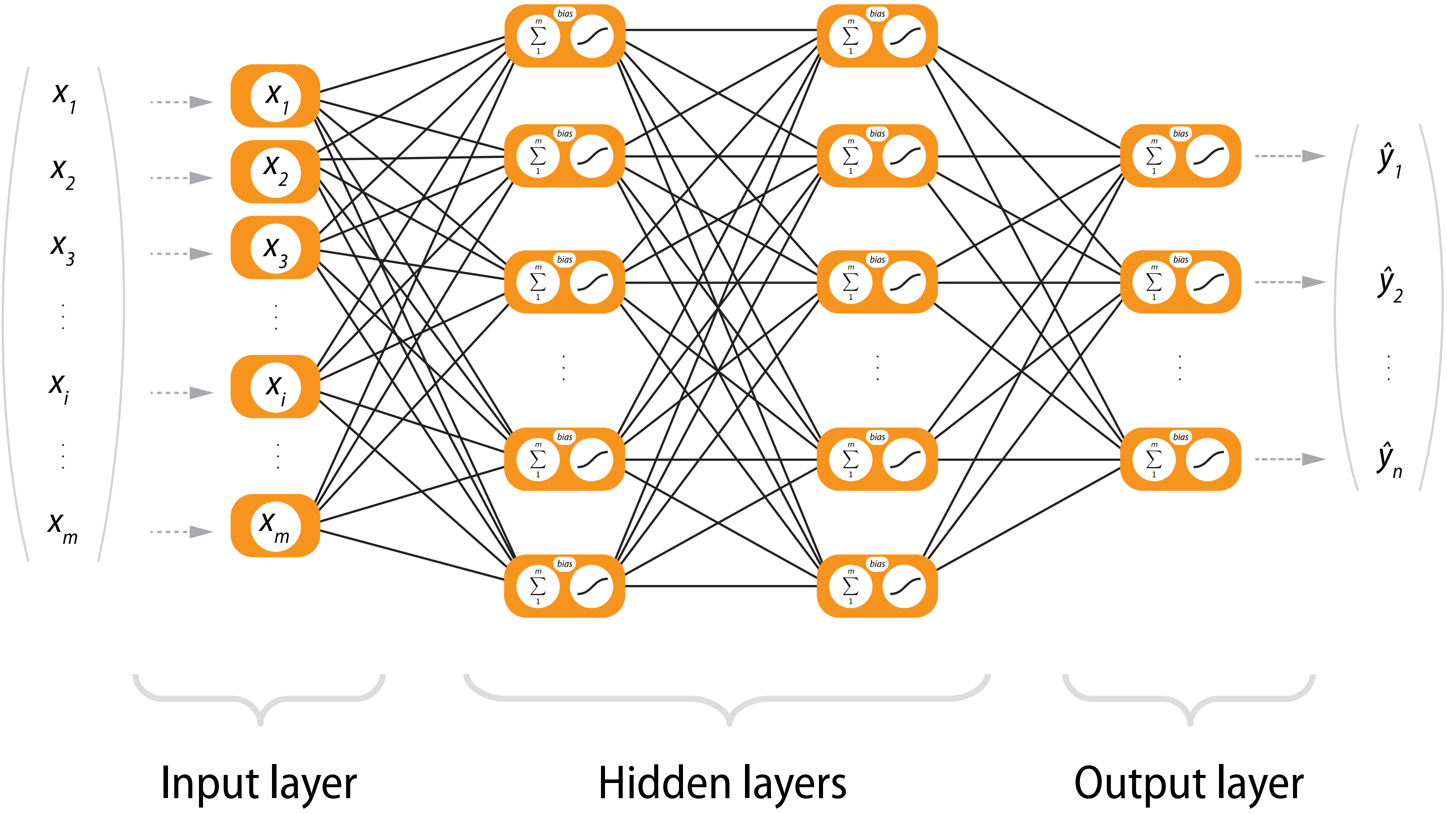


[1] D Cardon, JP Cointet, A Mazieres, 2018 [LRDN]

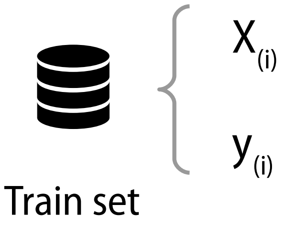
Deep Neural Networks



Deep Neural Networks

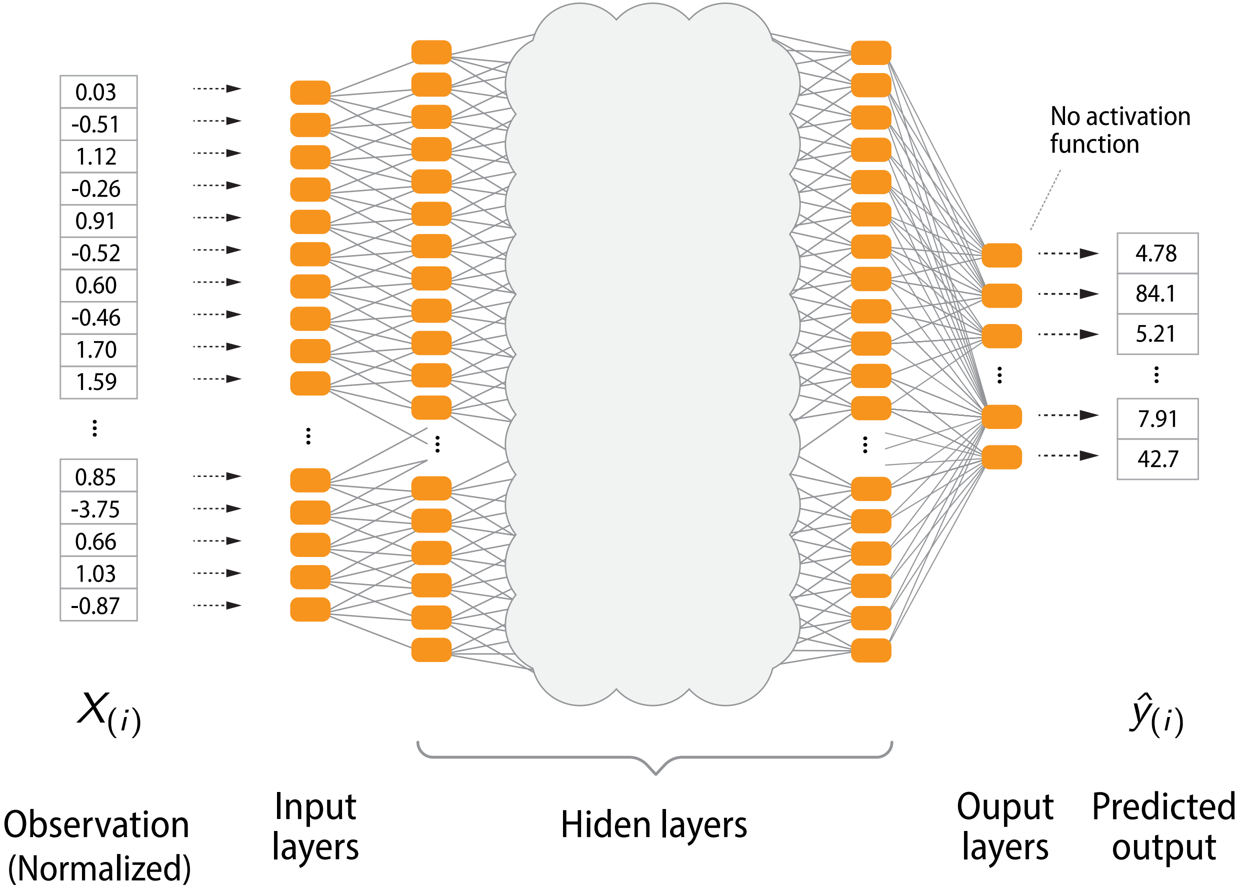


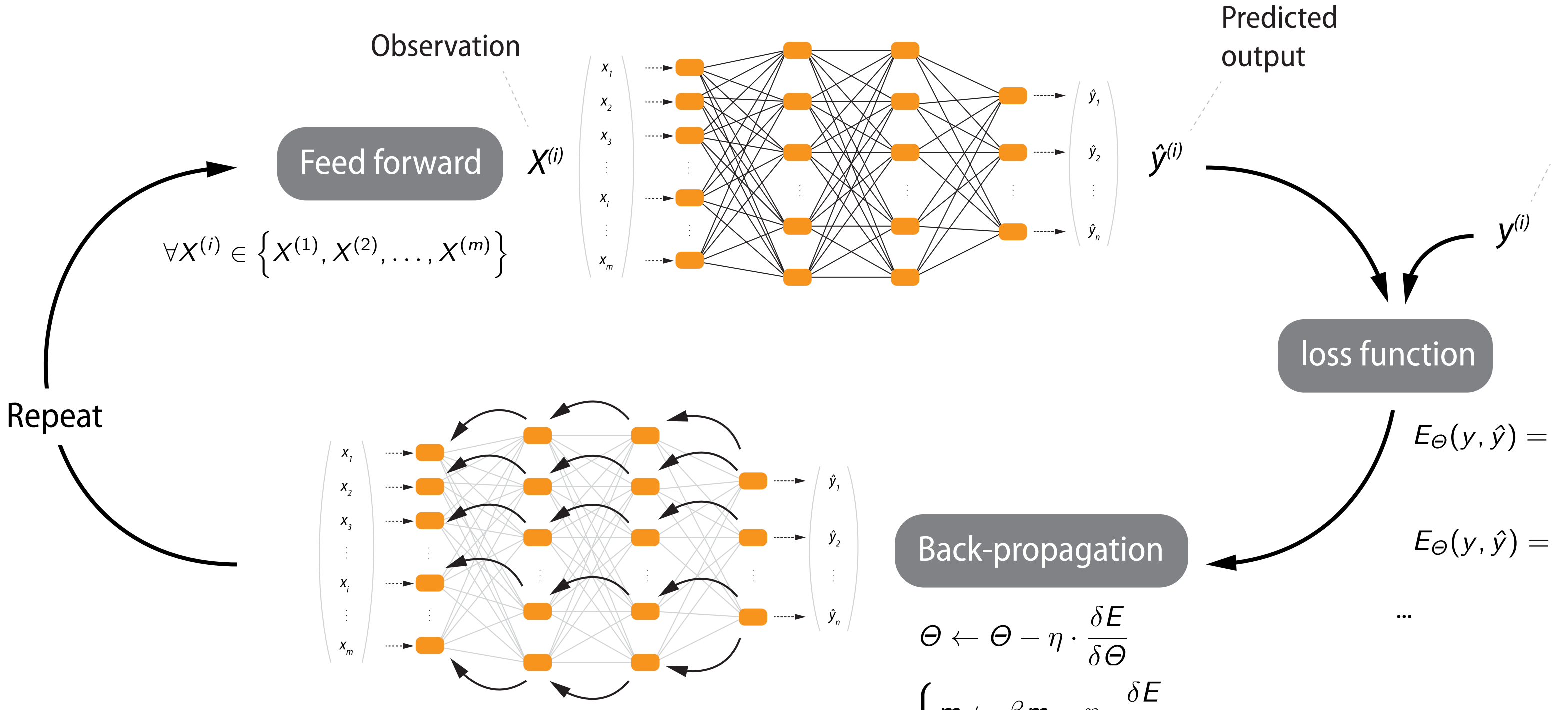
Deep Neural Networks



$X_{(i)}$: Observations

$y_{(i)}$: Expected output





$$E_\Theta(y,\hat{y}) = \frac{1}{2m}\sum_{i=1}^{m}\left\|y^{(i)} - \hat{y}^{(i)}\right\|^2$$

$$E_\Theta(y,\hat{y}) = -\sum_{i=1}^{m}\sum_{j=1}^{n} y_j^{(i)} \log(\hat{y}_j^{(i)})$$

...

$$\Theta \leftarrow \Theta - \eta \cdot \frac{\delta E}{\delta \Theta}$$

$$\begin{cases} m \leftarrow \beta m - \eta \cdot \dfrac{\delta E}{\delta \Theta} \\ \Theta \leftarrow \Theta + m \end{cases}$$

...

η : Learning rate
β : Momentum

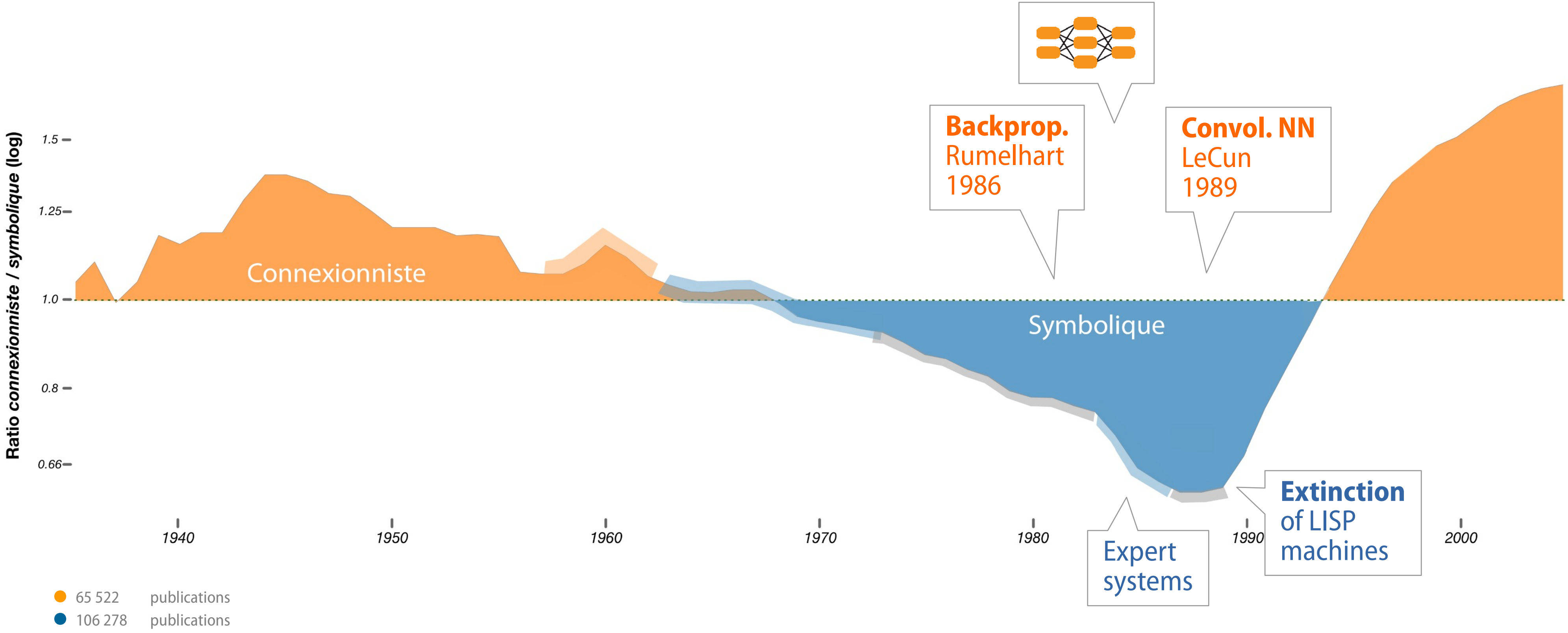
To define :
- Optimization
- Activation
- Loss
- Metrics
- ...

Back-propagation

Learning process

The big Controversy

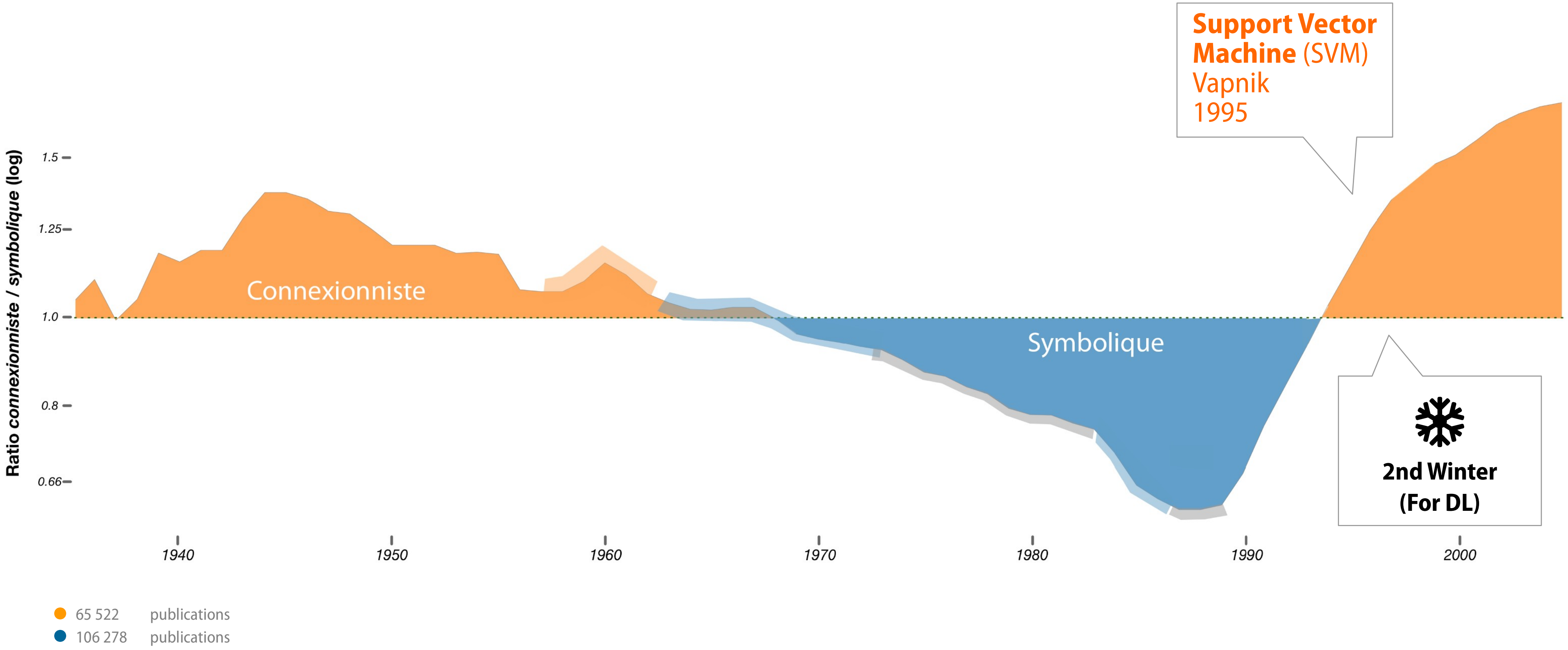
Evolution of the academic influence of connexionist and symbolic approaches[1]



[1] D Cardon, JP Cointet, A Mazieres, 2018 [LRDN]

The big Controversy

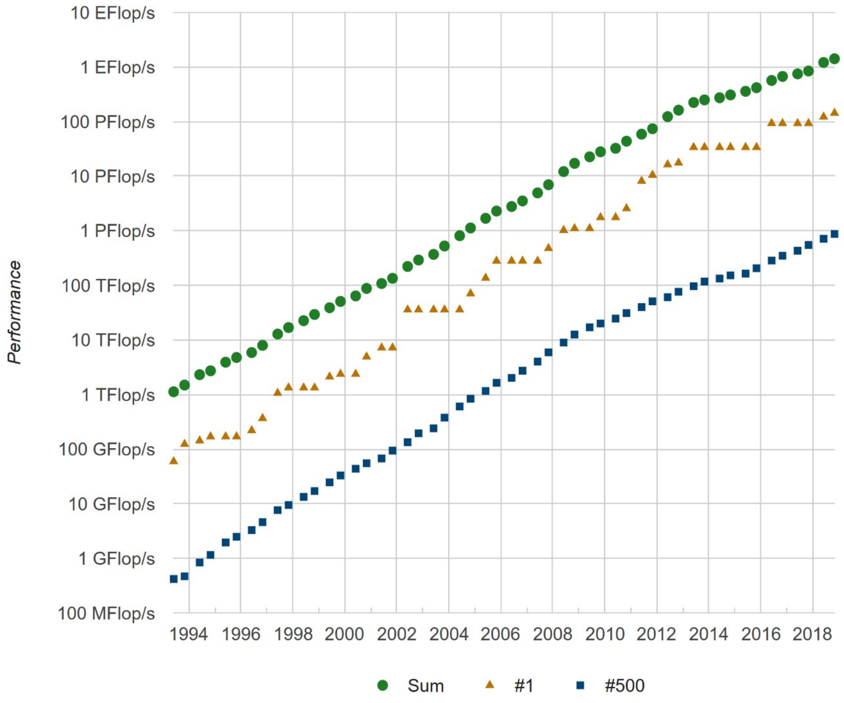
Evolution of the academic influence of connexionist and symbolic approaches[1]



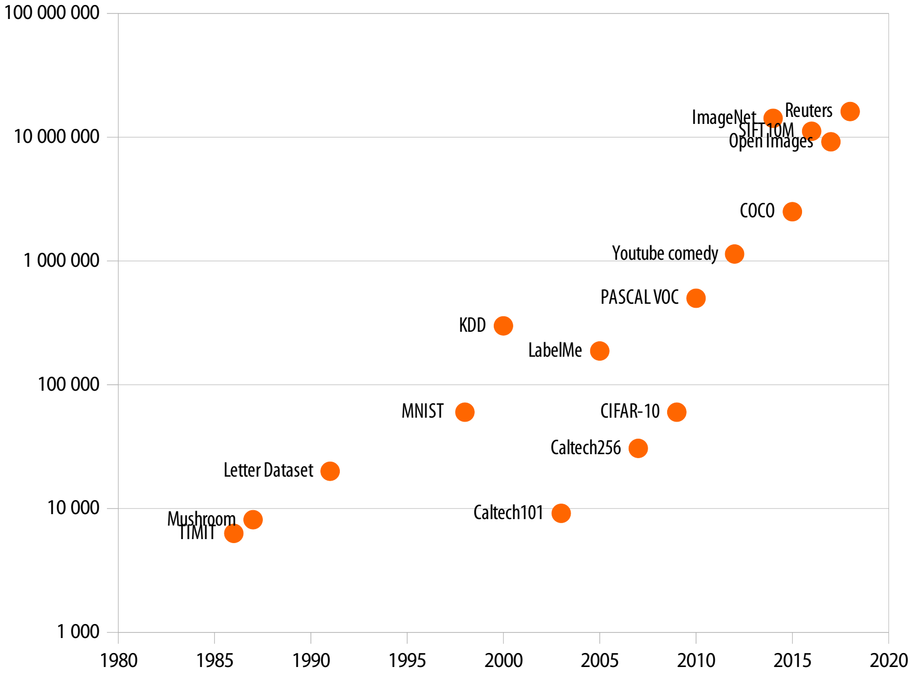
[1] D Cardon, JP Cointet, A Mazieres, 2018 [LRDN]

The big Controversy

Performance Development[1]



Datasets for machine-learning[2]



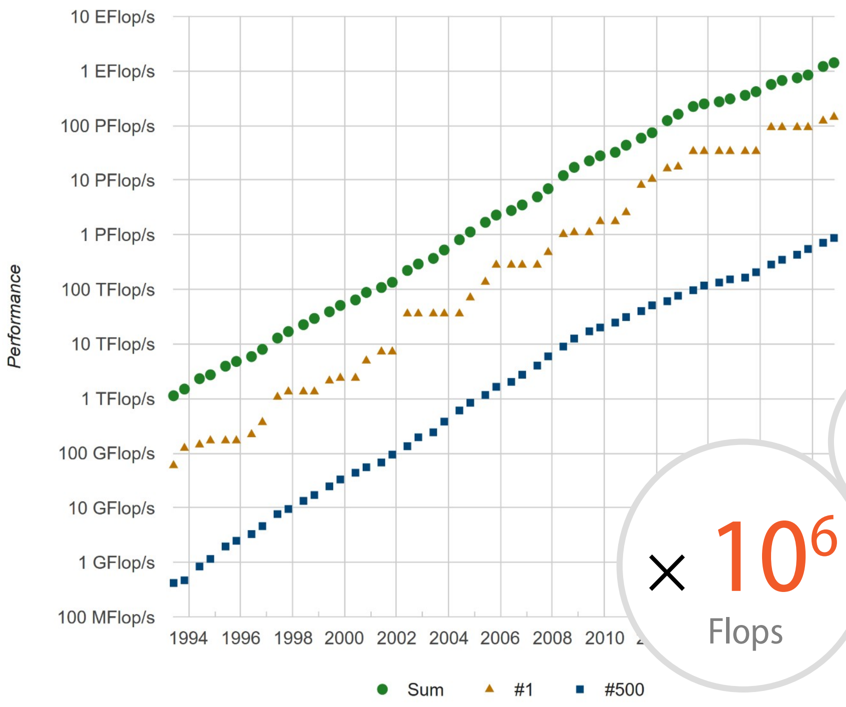
Laboratoire
Cas particulier

Monde réel

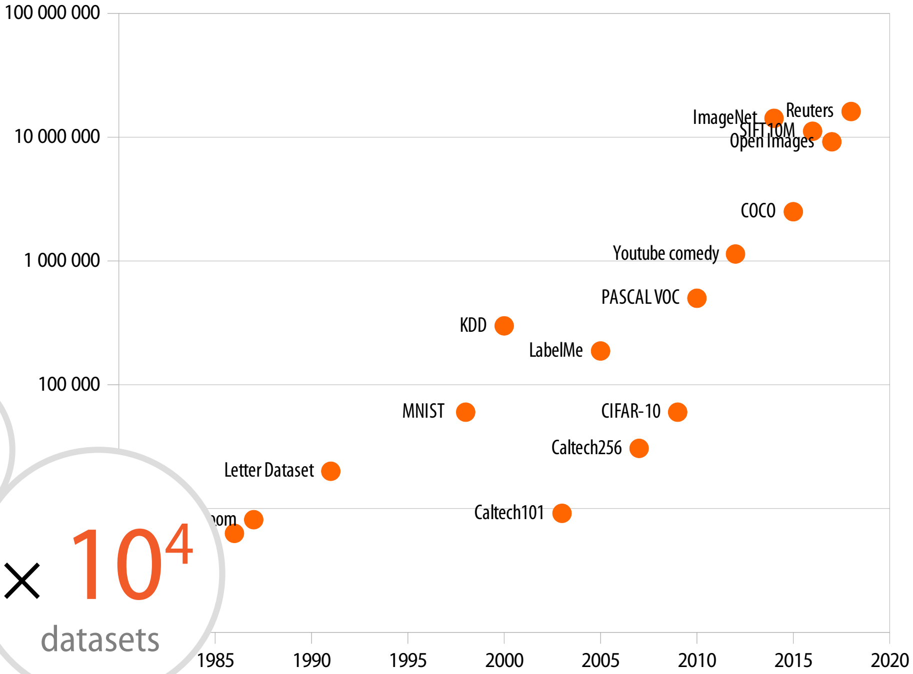
[1] TOP500 List [TOP500]

[2] Wikipedia [WKP1]

Performance Development[1]



Datasets for machine-learning[2]



25 ans

$\times\ 10^6$ Flops

$\times\ 10^4$ datasets

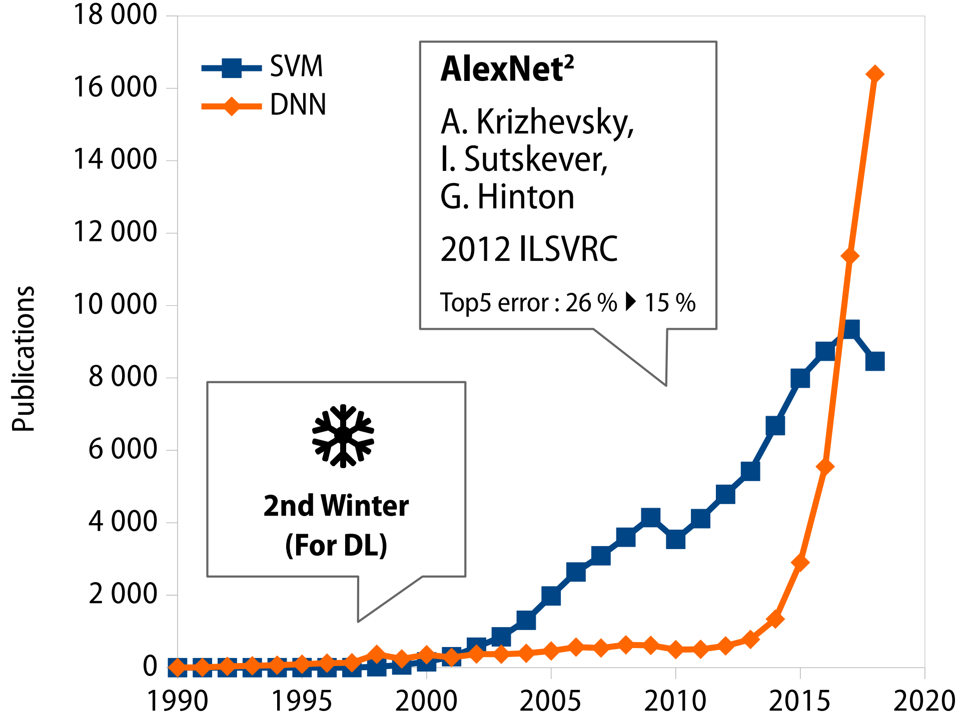
Laboratoire
Cas particulier ⇏ Monde réel

[1] TOP500 List [TOP500]

[2] Wikipedia [WKP1]

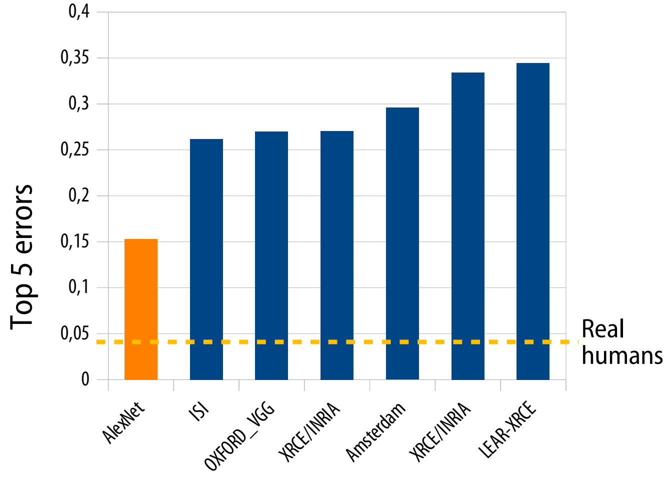
The big Controversy

Publications SVM vs DNN[1]



Images classification
Top 5 error at ILSVRC 2012[3,4]

DNN



Without mathematical guarantee, DNN have proven to be more effective in the face of the complexity of the real world !

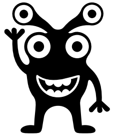
[1] Web of Science [WOS1][WOS2]
[2] AlexNet [ALEX]
[3] ImageNet Large Scale Visual Recognition [ILSVRC]

[4] Similar evolution in Natural language processing, translation, board games, etc.
See : DeepL.com, AlphaGo, AlphaZero, ...

"Physarum polycephalum", commonly called blob
Audrey DUSSUTOUR / CRCA / CNRS Photothèque

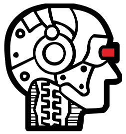
Zero brain
Zero neurons
720 possible sexes !

Call me the blob !

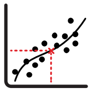


Roadmap

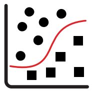
1



History, Fundamental Concepts



Basic Regression
DNN



Basic Classification
DNN

1.1 Introduction
Context, tools and ressources

1.2 **From the liner regression to the first neuron**

1.3 **Neurons in controversy**

1.4 **Data and neurons**

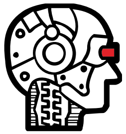
Basic Regression
Basic Classification

Short break for certificates of attendance

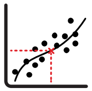
FIDLE

Roadmap

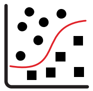
1



History, Fundamental Concepts



Basic Regression
DNN



Basic Classification
DNN

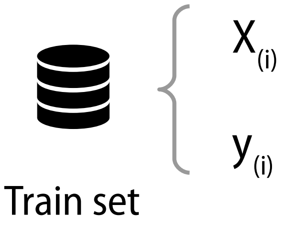
1.1 Introduction
Context, tools and ressources

1.2 **From the liner regression to the first neuron**

1.3 **Neurons in controversy**

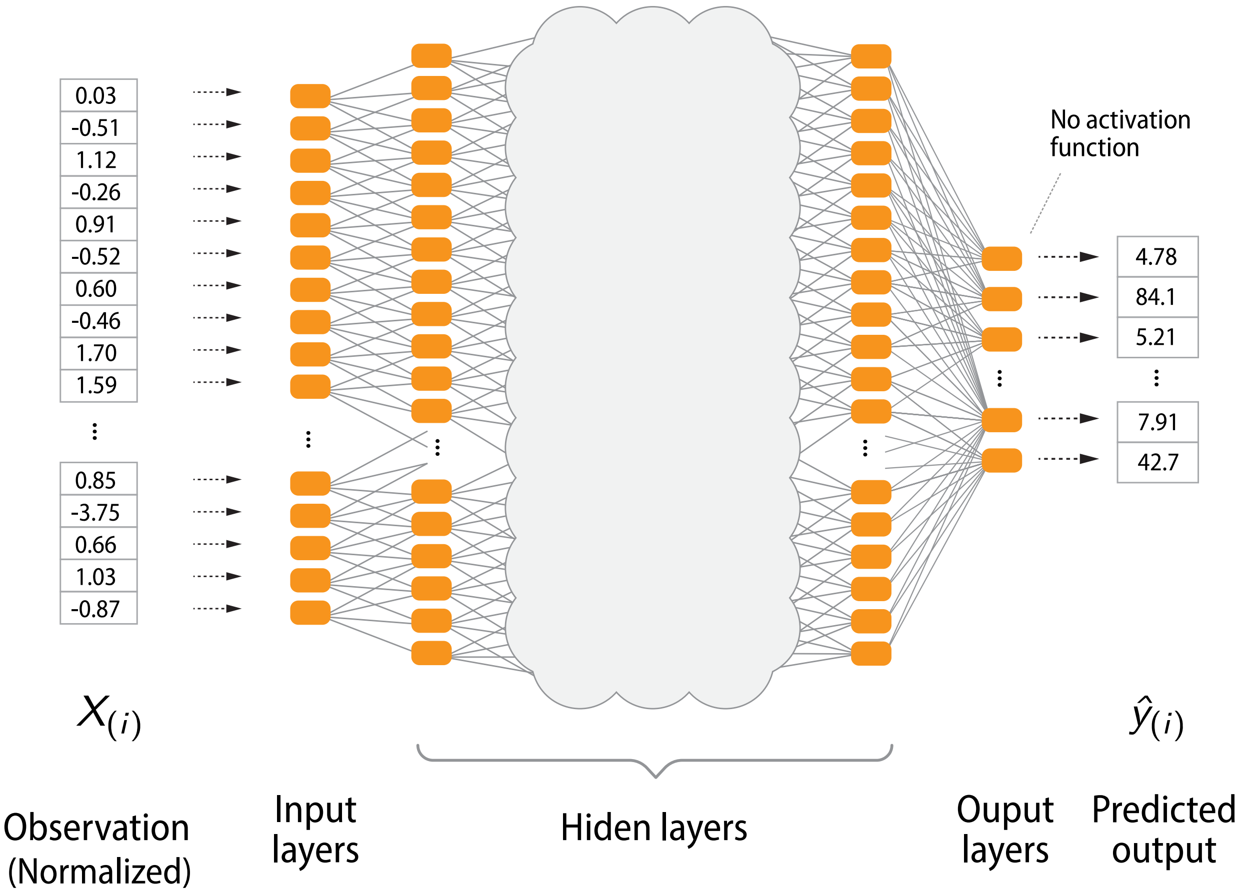
1.4 **Data and neurons**
Basic Regression
Basic Classification

Regression with a DNN



$X_{(i)}$: Observations

$y_{(i)}$: Expected output



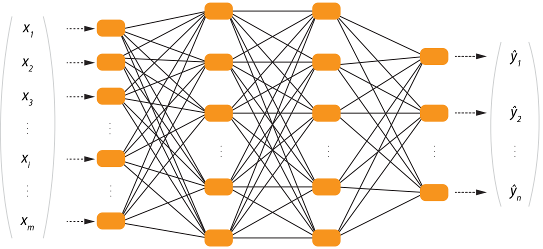
Training process - general

Observation

Feed forward

$X^{(i)}$

$\forall X^{(i)} \in \left\{X^{(1)}, X^{(2)}, \dots, X^{(m)}\right\}$



Predicted output

$\hat{y}^{(i)}$

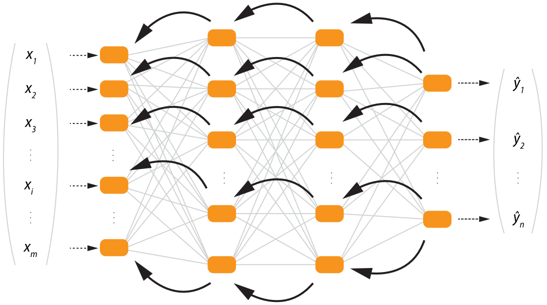
Expected output

$y^{(i)}$

loss function

$$E_{\Theta}(y, \hat{y}) = \frac{1}{2m} \sum_{i=1}^{m} \left\| y^{(i)} - \hat{y}^{(i)} \right\|^2$$

Repeat



Back-propagation

$$\Theta \leftarrow \Theta - \eta \cdot \frac{\delta E}{\delta \Theta}$$

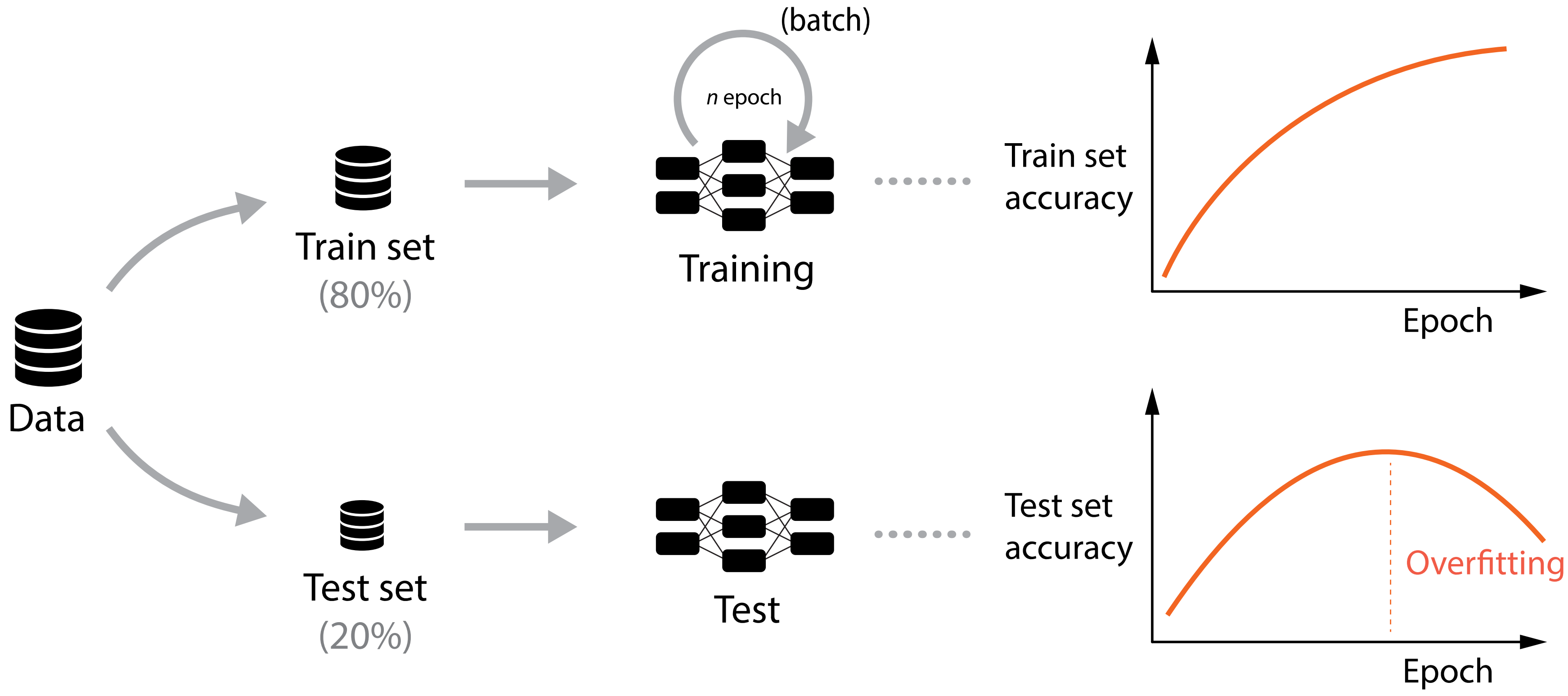
$$\begin{cases} m \leftarrow \beta m - \eta \cdot \dfrac{\delta E}{\delta \Theta} \\ \Theta \leftarrow \Theta + m \end{cases}$$

...

η : Learning rate
β : Momentum

To define :
- Optimization
- Activation
- Loss
- Metrics
- ...

Training process - general



Regression with a Dense Network (DNN)

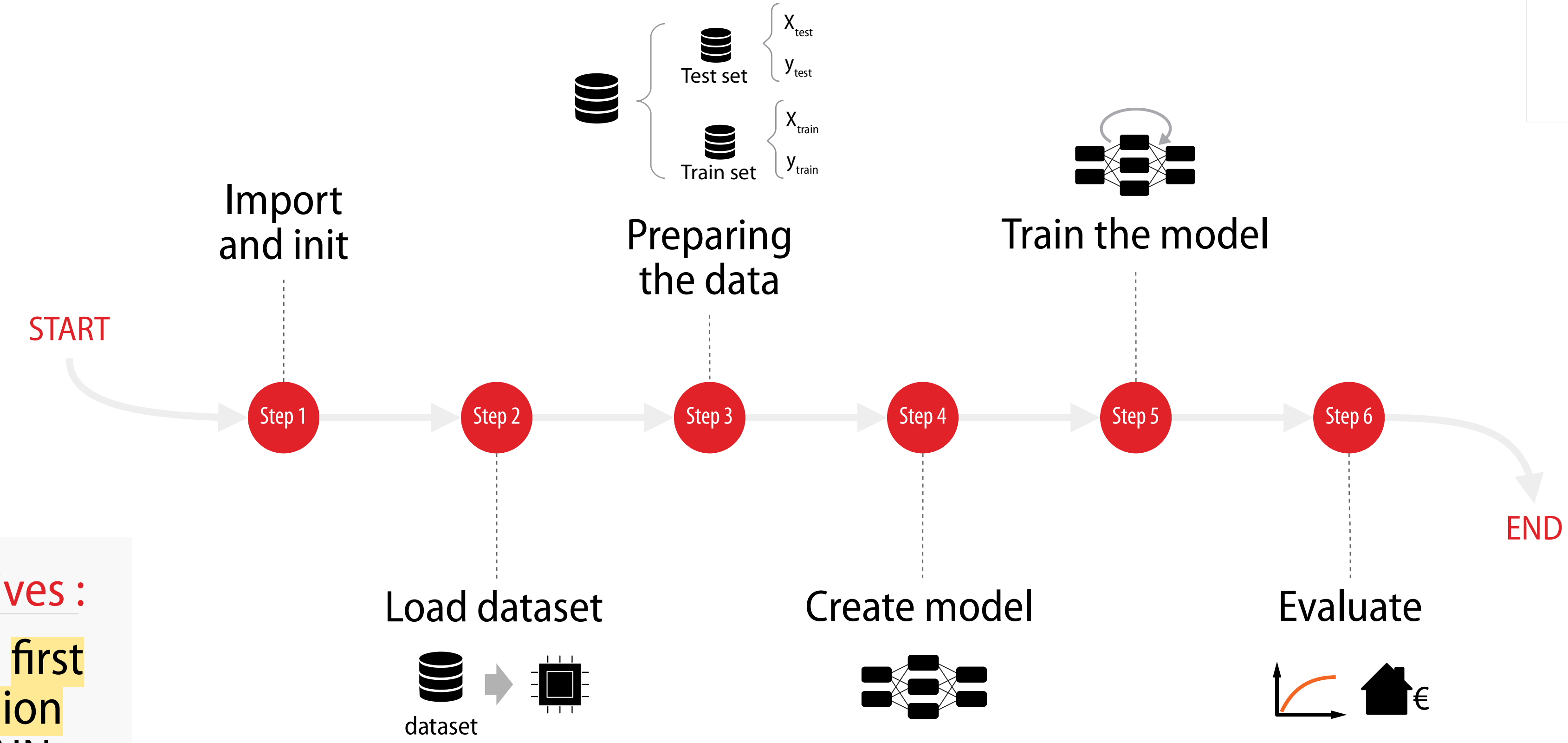
Notebook : [Wine1]

Objective :
Wine quality prediction with a Dense Network (DNN)

Dataset :
Wine Quality datasets,
provide by :Paulo Cortez, University of Minho, Guimarães, Portugal,
http://www3.dsi.uminho.pt/pcortez

WINE1

50s
(5 epochs)

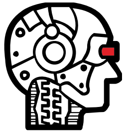


Objectives :

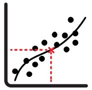
Make a first regression via a DNN network

Roadmap

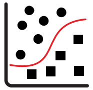
1



History, Fundamental Concepts



Basic Regression
DNN



Basic Classification
DNN

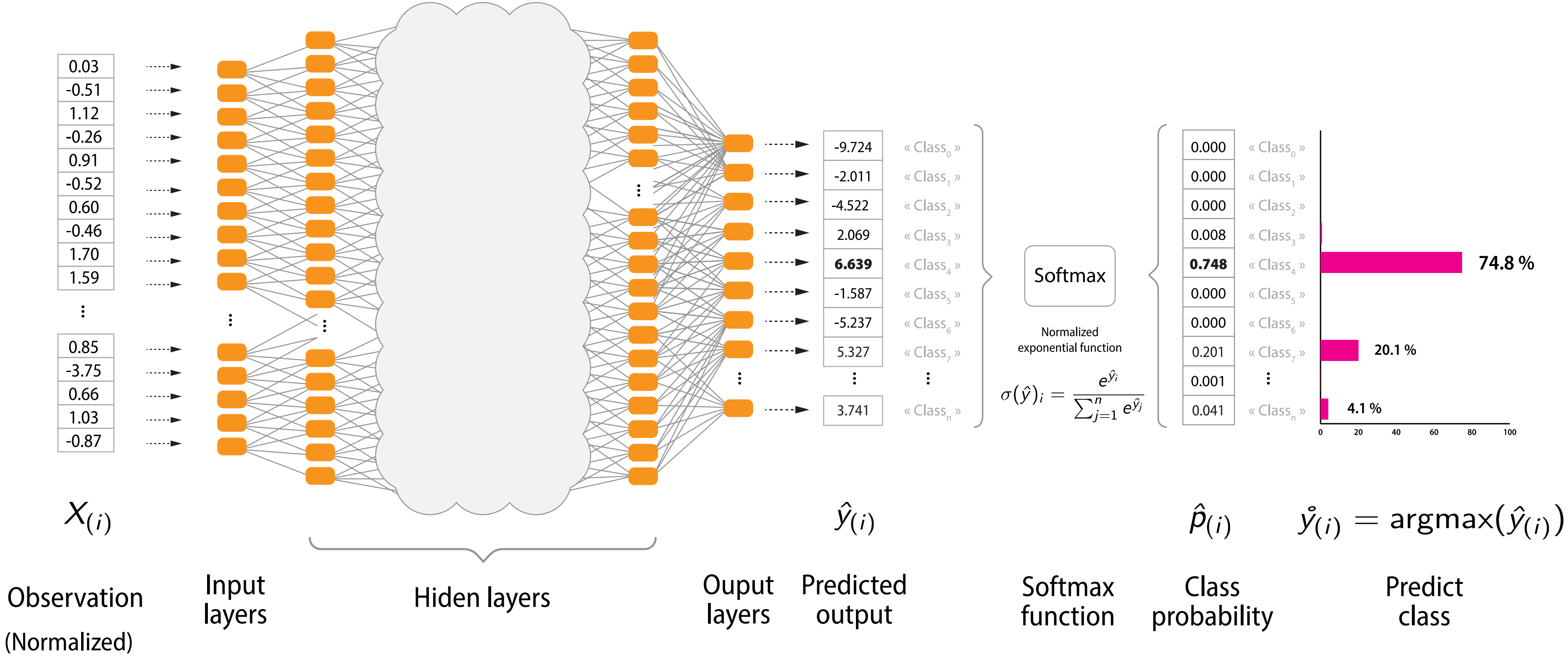
1.1 Introduction
Context, tools and ressources

1.2 **From the liner regression to the first neuron**

1.3 **Neurons in controversy**

1.4 **Data and neurons**
Basic Regression
Basic Classification

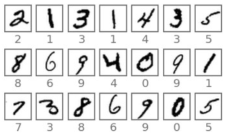
Classification with a DNN



Simple classification with DNN

Notebook : [MNIST1]

Objective :
Recognizing handwritten numbers



Dataset :
Modified National Institute of Standards and Technology (MNIST)

MNIST1

50s
(5 epochs)

START

Step 1 — Import and init

Step 2 — Load dataset

dataset

Step 3 — Preparing the data

Test set: X_{test}, y_{test}

Train set: X_{train}, y_{train}

Step 4 — Create model

Step 5 — Train the model

Step 6 — Evaluate

END

Objectives :

Make a first classification via a DNN network

Next, on Fidle :

2 3

Hight Dimensionnal Data

CNN

Jeudi 24 novembre,
14h00

Séquence 2 :

Réseaux convolutifs, partie 1

Quand la dimension des données augmente...
Images et réseaux convolutifs

- Principes et concepts des réseaux convolutifs (CNN)
- Convolutions - Dropout - Pooling

Exemple proposé :

Classification de chiffres manuscrits

Durée : 2h00

Next on Fidle :

2 3

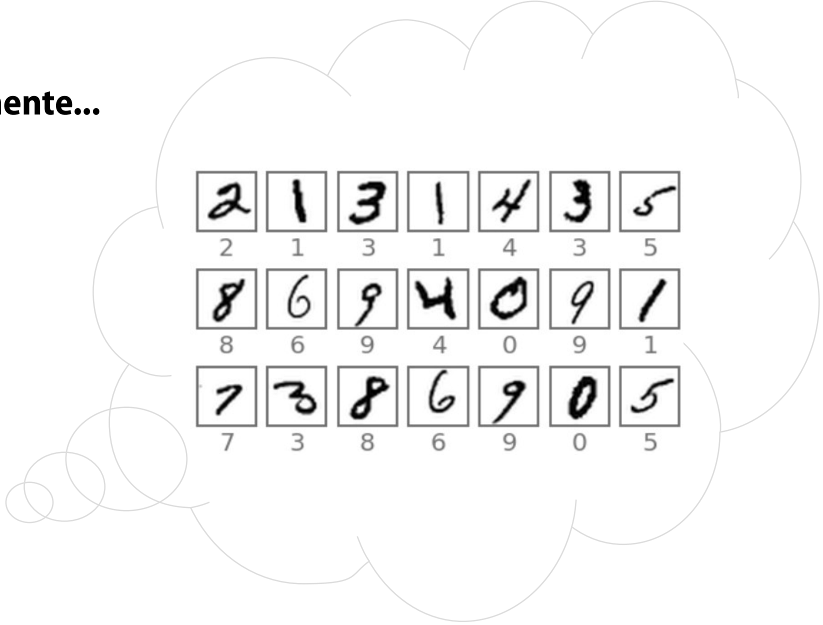
Hight Dimensionnal Data

CNN

Jeudi 24 novembre, 14h00

Séquence 2 :

Réseaux convolutifs, partie 1
Quand la dimension des données augmente...
Images et réseaux convolutifs



To be continued...

Contact@fidle.cnrs.fr
FIDLE https://fidle.cnrs.fr
YouTube https://fidle.cnrs.fr/youtube

FIDLE

Références

[JGRAY] Gray, J. (2001), from « The Fourth Paradigm: Data-Intensive Scientific Discovery » Tony Hey, Stewart Tansley, Kristin Tolle (2009). Published by Microsoft Research.
ISBN: 978-0-9825442-0-4

[FROS] Rosenblatt, Frank. (1958). « The perceptron: A probabilistic model for information storage and organization in the brain. » Psychological Review, 65(6), 386-408.

[MIPA] Minsky, Marvin; Papert, Seymour. (1969). « Perceptrons : An Introduction to Computational Geometry », MIT Press

[DRUM] Rumelhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J. (1986). « Learning representations by back-propagating errors ». Nature. 323 (6088): 533–536. doi:10.1038/323533a0.

[YLEC1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, « Backpropagation Applied to Handwritten Zip Code Recognition », AT&T Bell Laboratories

[LRDN] Dominique Cardon, Jean-Philippe Cointet, Antoine Mazieres. (2018). « La revanche des neurones », Réseaux, La Découverte, 5 (211), <10.3917/res.211.0173>. <hal-01925644>

[TOP500] Statistics on top 500 high-performance computers. (2018) « Exponential growth of supercomputing power as recorded by the TOP500 list ». https://www.top500.org

[WKP1] Wikipedia/en. (2018) « List of datasets for machine-learning research ». https://en.wikipedia.org

[WOS1] Core database : TS=("support vector machine*" OR ("SVM" AND "classification") OR ("SVM" AND "regression") OR ("SVM" AND "classifier") OR "support vector network*" OR ("SVM" AND "kernel trick*"))

[WOS2] Core database : TS=("deep learning" OR "deep neural network*" OR ("DNN" AND "neural network*") OR "convolutional neural network*" OR ("CNN" AND "neural network*") OR "recurrent neural network*" OR ("LSTM" AND "neural network*") OR ("RNN*" AND "neural network*"))

Illustrations :

Potato plant From *Die Giftpflanzen Deutschlands,* Peter Esser, 1910, via iconspng.com

Neuron Wikimedia Commons, the free media repository.

Icons thenounproject.com