Formation

Introduction au

Deep Learning

Sequence 01

“History and fundamental concepts”

FIDLE

Formation

Introduction au Deep Learning

Fidle est une formation libre, gratuite, ouverte à toutes et à tous, sans inscription :-)

FIDLE

FIDLE

Formation

Introduction au Deep Learning

FIDLE est portée par **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**, des ingénieurs du **Programme National de Recherche en Intelligence Artificielle du CNRS (PNRIA)**, 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**.

FIDLE

Formation

Introduction au Deep Learning

https://fidle.cnrs.fr

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

- Course materials (pdf)
- Practical work environment*
- Corrected notebooks
- Videos (YouTube)

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

Formation

Introduction au Deep Learning

FIDLE

Questions and answers :

https://fidle.cnrs.fr/q2a

Accompanied by :
IA Support (dream) Team of IDRIS

Directed by :
Agathe, Baptiste et Yanis - UGA/DAPI
Thibaut, Kamel - IDRIS

FIDLE

Formation

Introduction au Deep Learning

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

Fidle information list

Agoria http://fidle.cnrs.fr/agoria New !

AI exchange list

agoria@grenoble.cnrs.fr

(*) ESR is Enseignement Supérieur et Recherche, french universities and public academic research organizations

FIDLE

Formation

Introduction au Deep Learning

https://listes.services.cnrs.fr/wws/info/devlog

List of ESR* « Software developers » group

https://listes.math.cnrs.fr/wws/info/calcul

List of ESR* « Calcul » group

(*) ESR is Enseignement Supérieur et Recherche, french universities and public academic research organizations

Shall
we go ?

FIDLE

FIDLE

SAISON 23/24

Bases, Concepts et Enjeux

1. **History and Fundamental Concepts**
2. New ! **Data, models and representation's hell** — Data and models
3. New ! **Demonstration Illustration** — LLM / Text to Image
4. **AI, Law, Society and Ethics**

L'IA comme un outil,

5. New ! **Mathematics, gradients everywhere !**
6. New ! **Learning methodology**
7. **Convolutional models** — CNN
8. **Sparse (text) and sequences data** — Embedding, RNN
9. **«Attention is All You Need»** — Transformers
10. **Graph Neural Network** — GNN
11. **Autoencoder networks** — AE
12. **Variational Antoencoder** — VAE
13. **Generative Adversarial Networks** — GAN
14. **Diffusion Model** — Text to image
15. **Deep Reinforcement Learning** — RL
16. **Physics-Informed Neural Networks** — PINNS

Acteur de l'IA

17. **Learning faster and cheaper,** — Eco-Friendly
18. **Jean-Zay** — GPU acceleration
19. New ! **New models** — VLM, SM, Multimodal, ...
20. New ! **Case Study** — Experience feedback

FIDLE

SAISON 23/24

Bases, Concepts et Enjeux

1. **History and Fundamental Concepts**
2. New ! **Data, models and representation's hell** — Data and models
3. New ! **Demonstration Illustration** — LLM / Text to Image
4. **AI, Law, Society and Ethics**

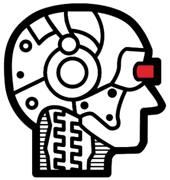
L'IA comme un outil,

5. New ! **Mathematics, gradients everywhere !**
6. New ! **Learning methodology**
7. **Convolutional models** — CNN
8. **Sparse (text) and sequences data** — Embedding, RNN
9. **«Attention is All You Need»** — Transformers
10. **Graph Neural Network** — GNN
11. **Autoencoder networks** — AE
12. **Variational Antoencoder** — VAE
13. **Generative Adversarial Networks** — GAN
14. **Diffusion Model** — Text to image
15. **Deep Reinforcement Learning** — RL
16. **Physics-Informed Neural Networks** — PINNS

Acteur de l'IA

17. **Learning faster and cheaper,** Eco-Friendly
18. **Jean-Zay** — GPU acceleration
19. New ! **New models** — VLM, SM, Multimodal, ...
20. New ! **Case Study** — Experience feedback

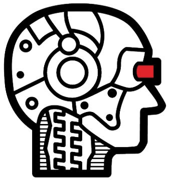
1



History and **Fundamental Concepts**

1. **Deep Learning**,
   what are we talking about?
2. From linear regression
   to **artificial neurons**
3. 80 years of history
   ...and the **victory** of **neurons**
4. Concrete **illustration** :
   - Basic Regression
   - Basic Classification

1



History and **Fundamental Concepts**

1 **Deep Learning,** what are we talking about?

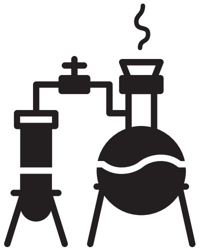
2 From linear regression to **artificial neurons**

3 80 years of history ...and the **victory** of **neurons**

4 Concrete **illustration** :
- Basic Regression
- Basic Classification

Scientific paradigms

1st paradigm



Experimental science

2nd paradigm

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

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

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

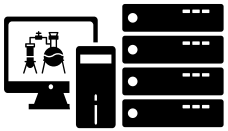
Theoretical science

3rd paradigm

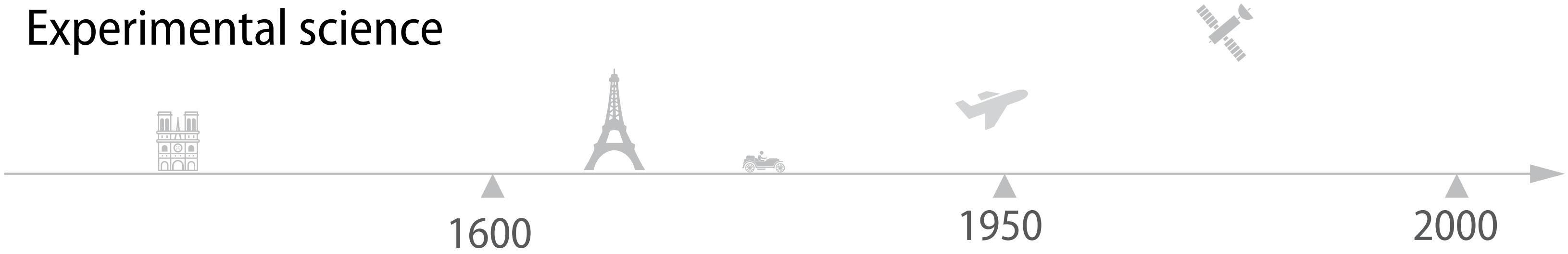
$$i\hbar\frac{d}{dt}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle$$

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

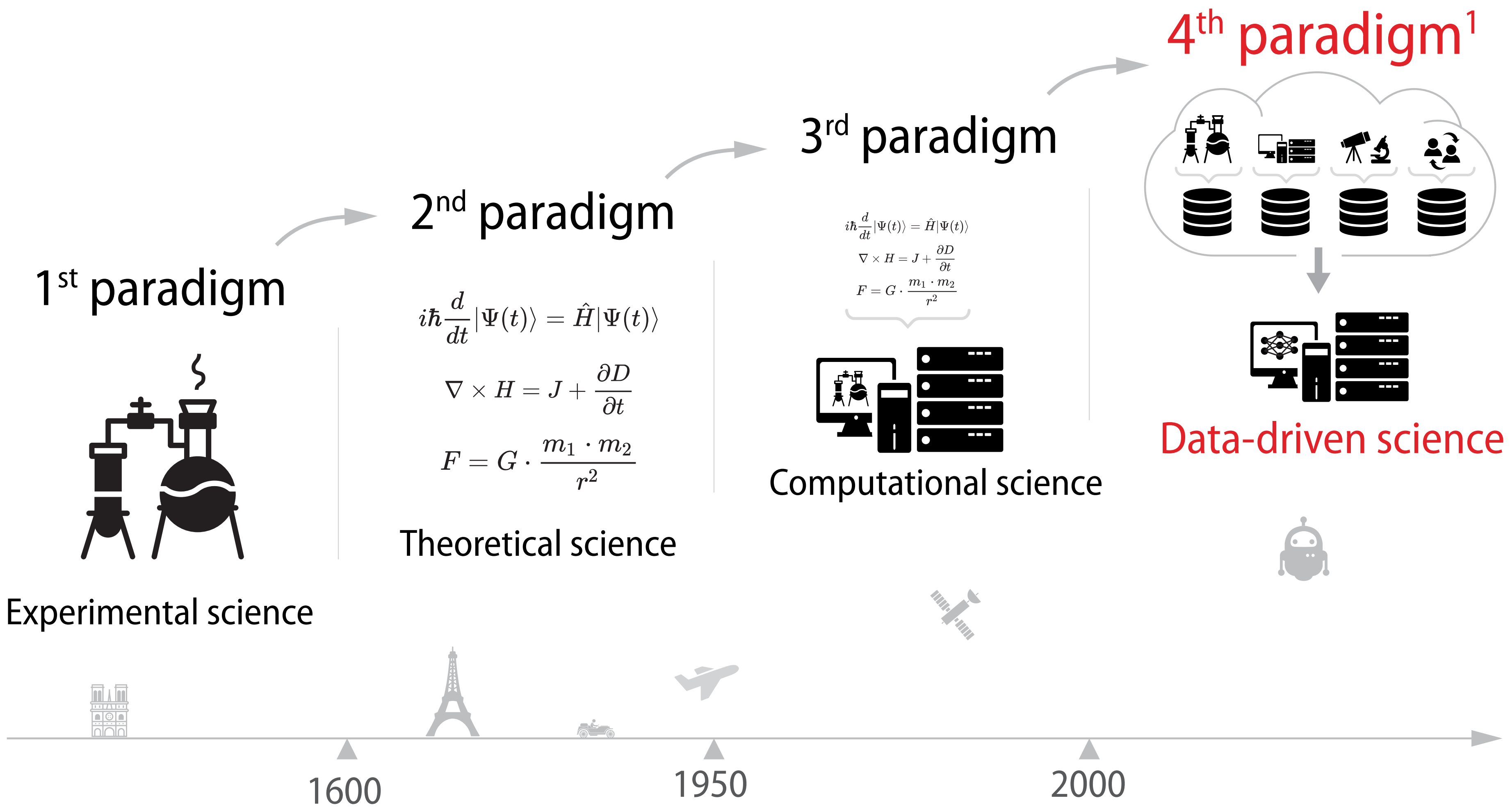
$$F = G \cdot \frac{m_1 \cdot m_2}{r^2}$$



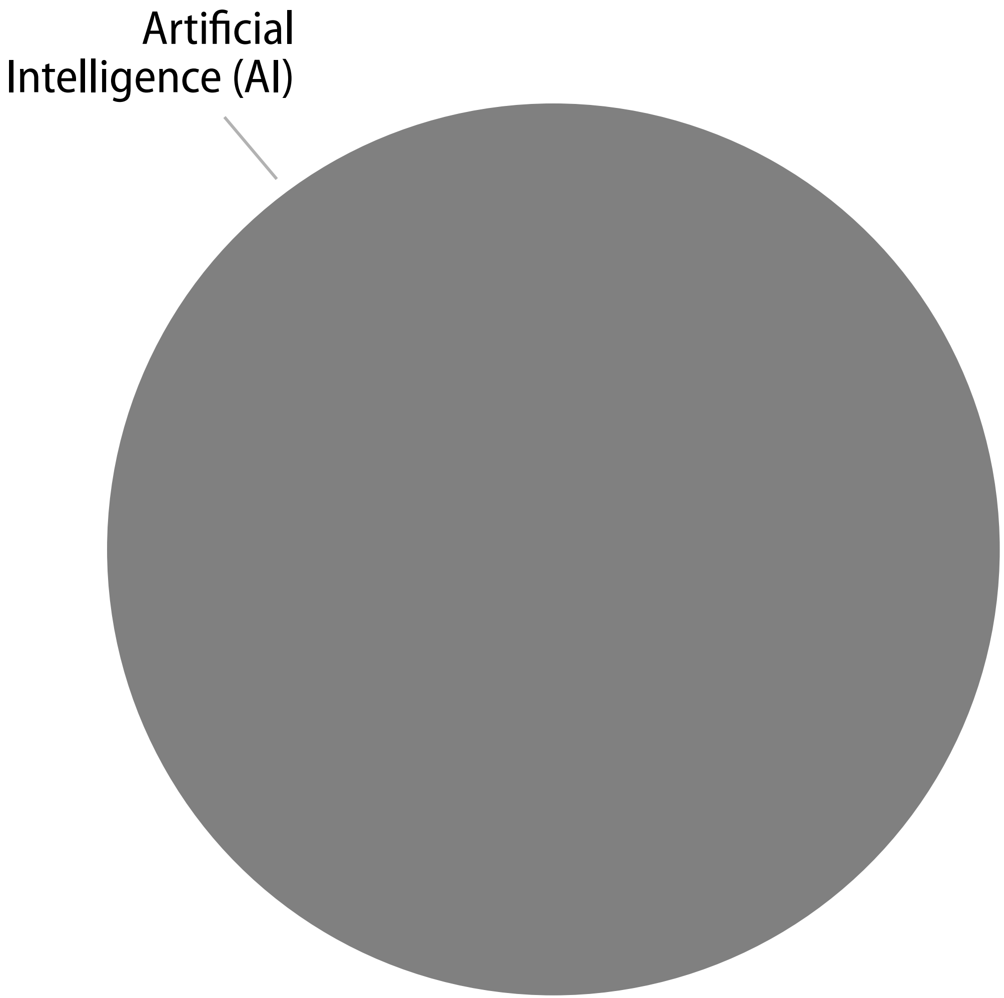
Computational science

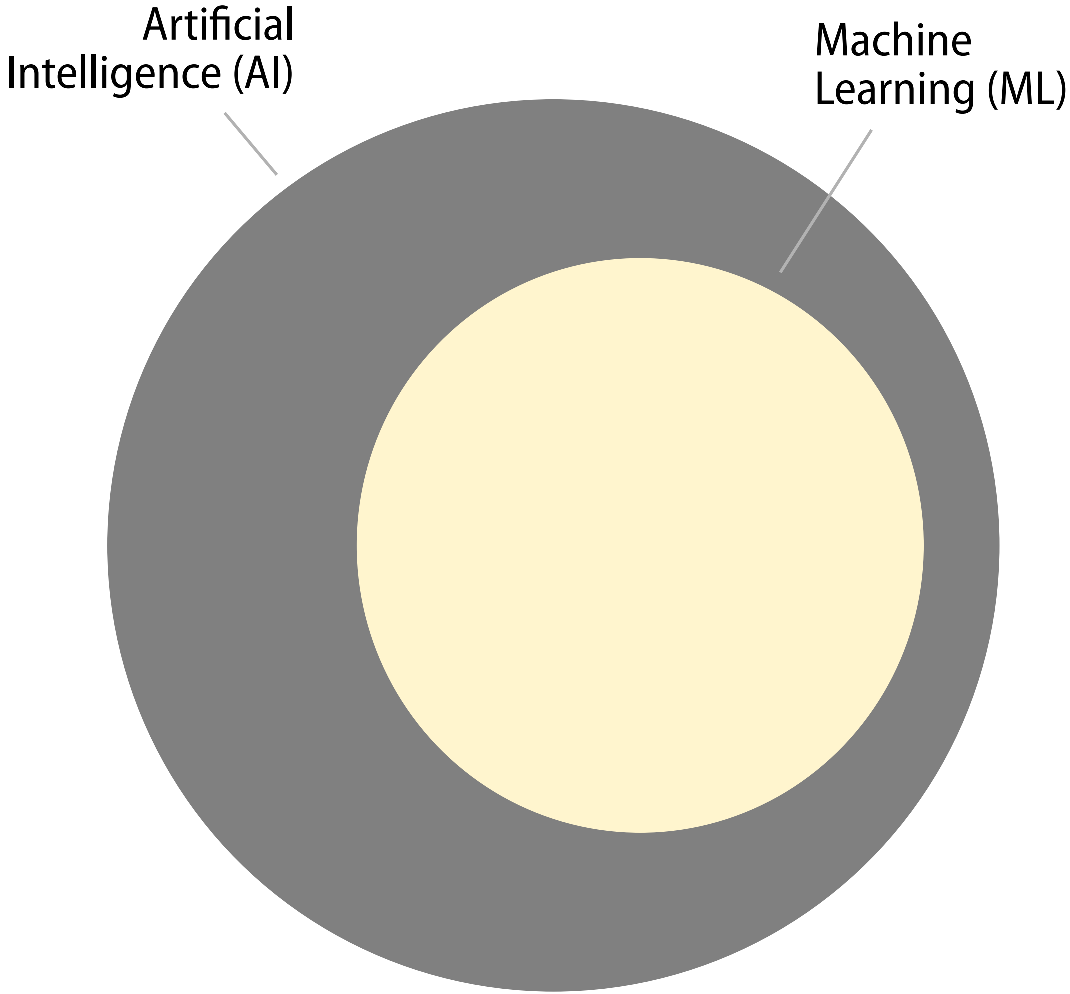


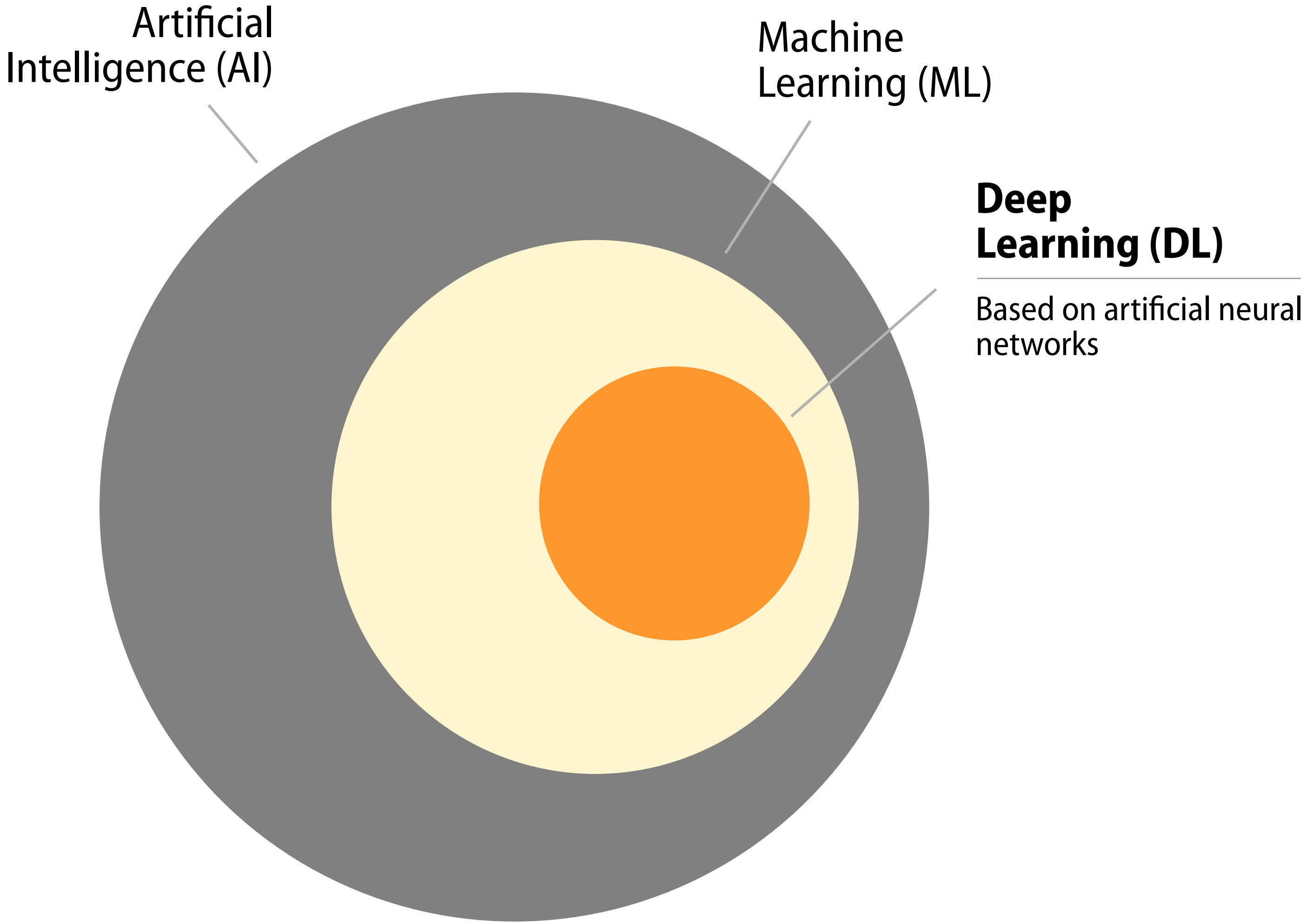
Scientific paradigms

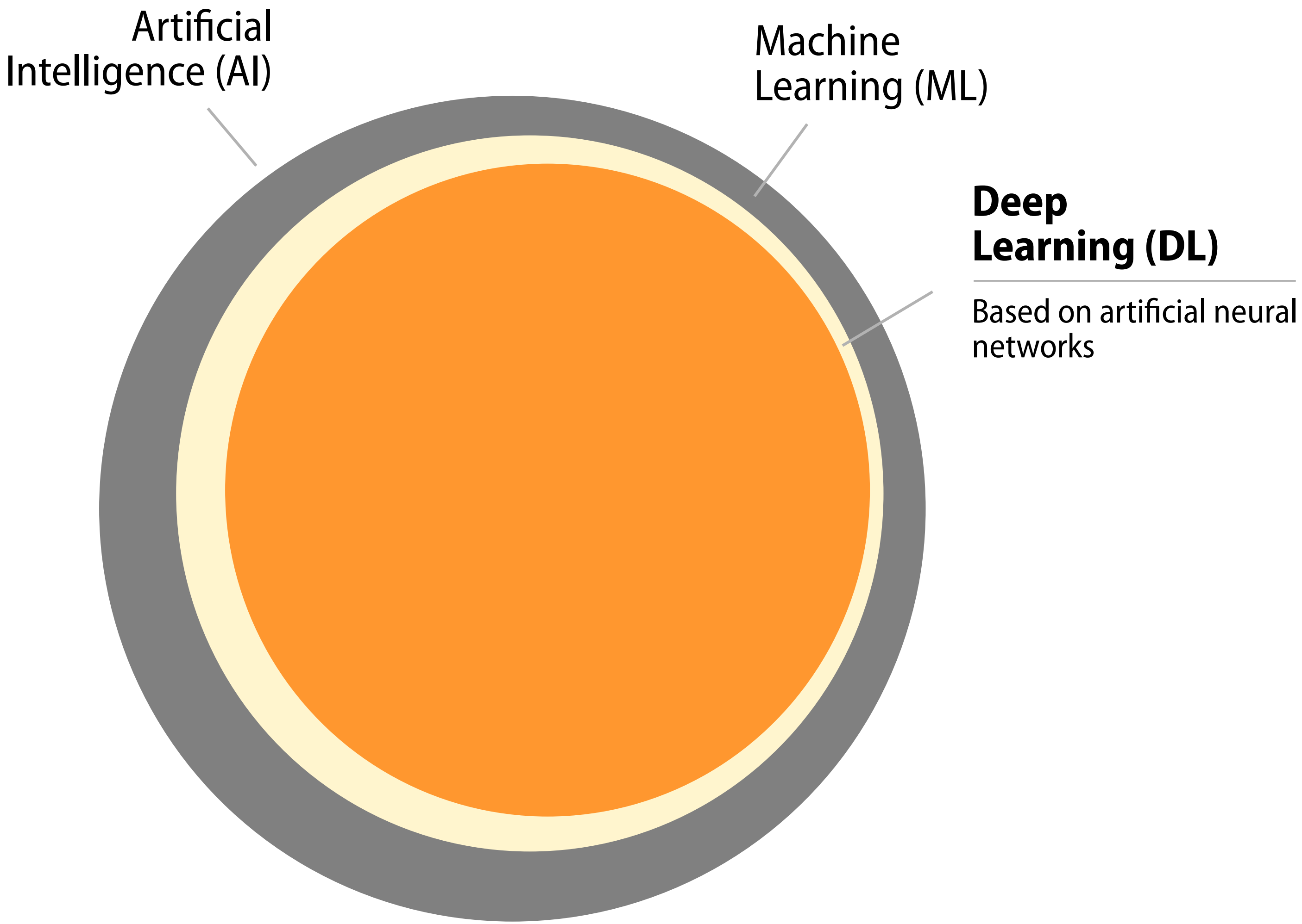


[1] Jim Gray, 2007


Artificial
Intelligence (AI)


Artificial
Intelligence (AI)
Machine
Learning (ML)


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

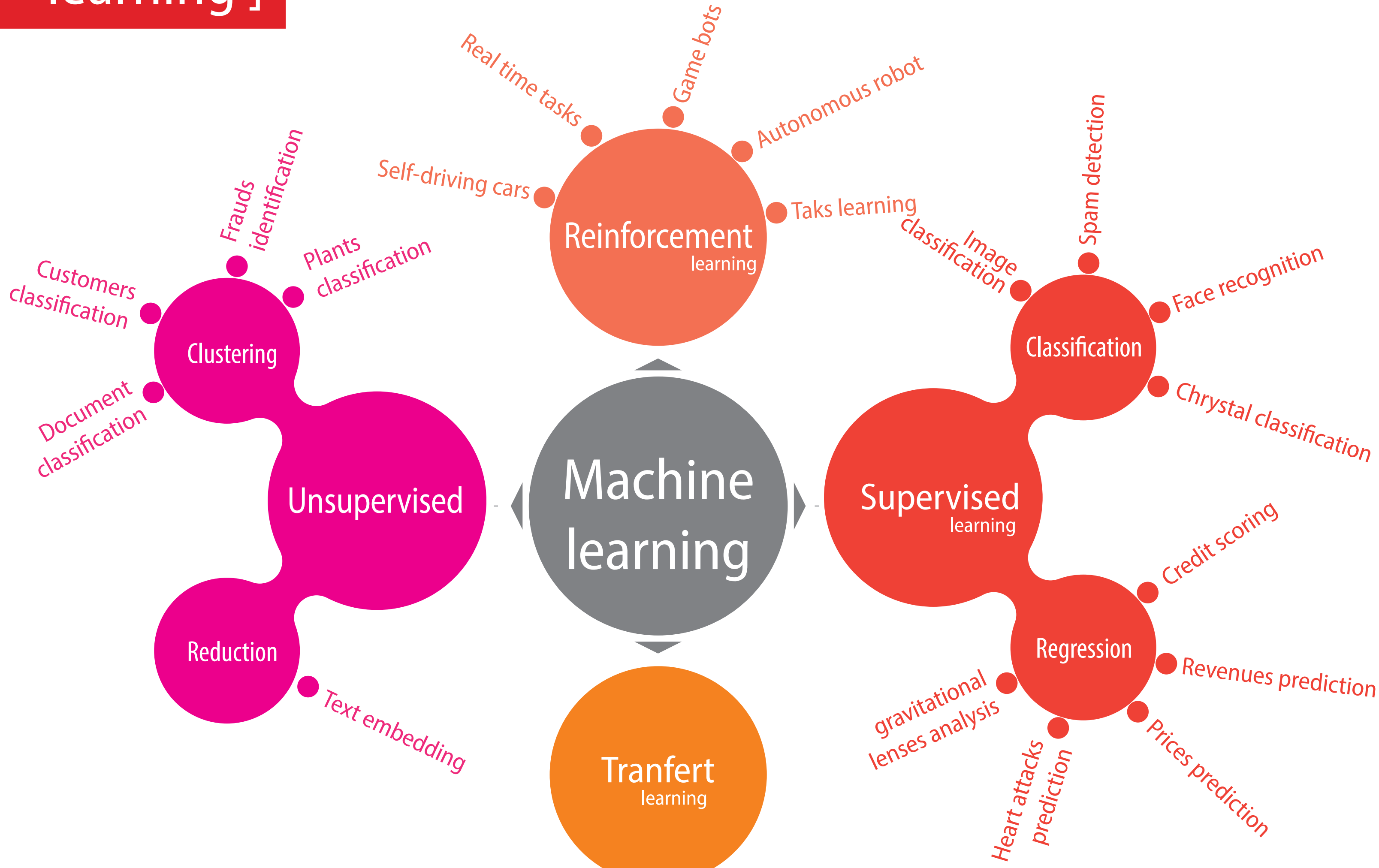

Artificial Intelligence (AI)
Machine Learning (ML)
Deep Learning (DL)
Based on artificial neural networks

Science

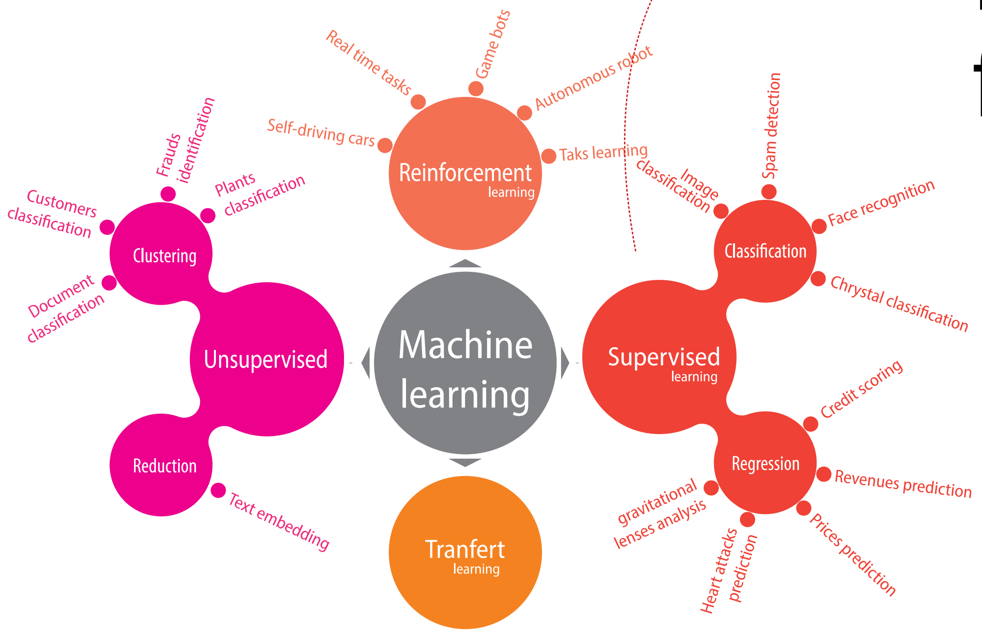
Social signal learning of the waggle dance in honey bees

Science, 9 Mars 2023
https://www.science.org/doi/10.1126/science.ade1702

[*-learning]

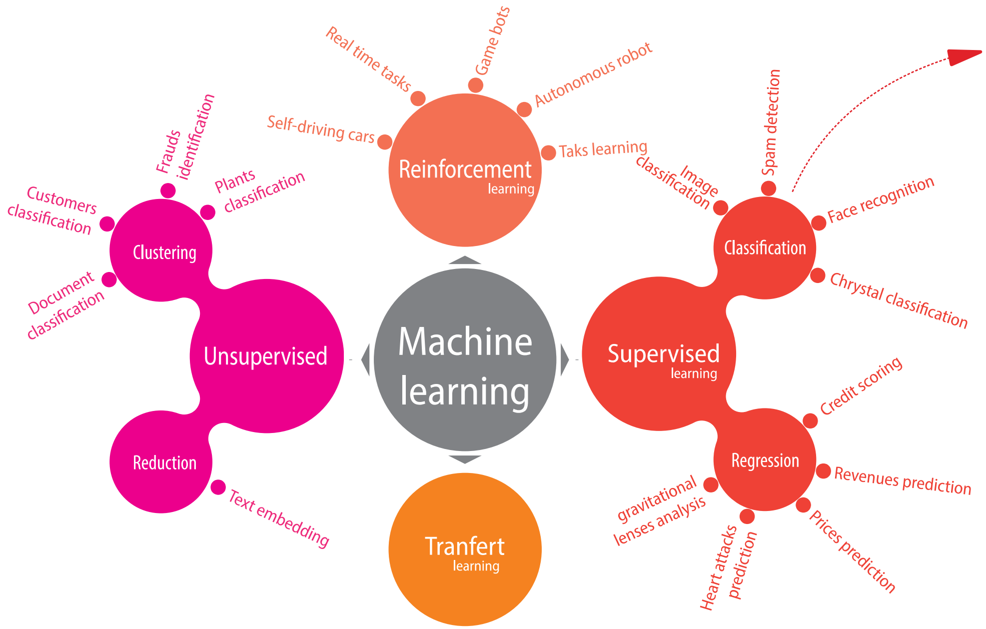


Machine learning
Reinforcement learning
Real time tasks
Game bots
Autonomous robot
Self-driving cars
Taks learning
Unsupervised
Clustering
Frauds identification
Plants classification
Customers classification
Document classification
Reduction
Text embedding
Tranfert learning
Supervised learning
Classification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



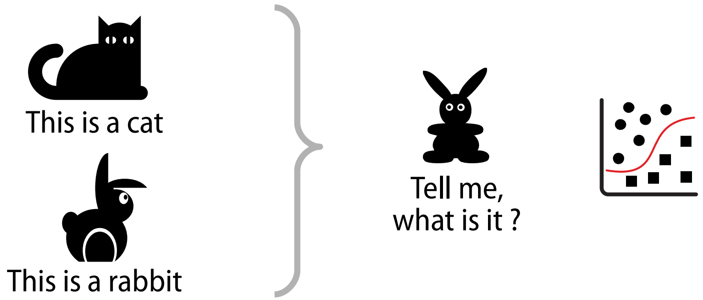
Learning from **labeled data**

Supervised learning

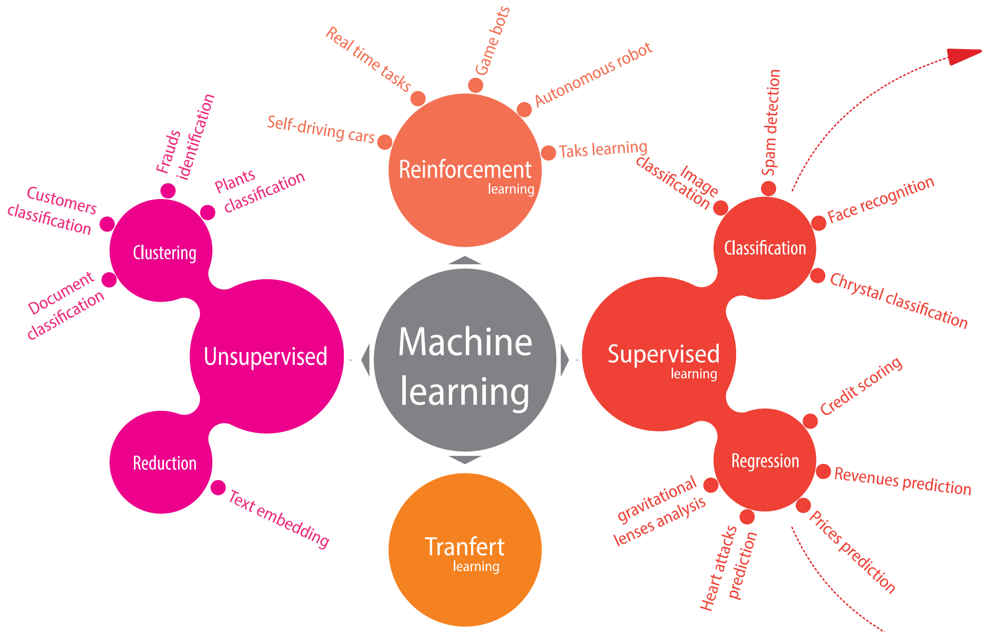


Classification :

Predict qualitative informations

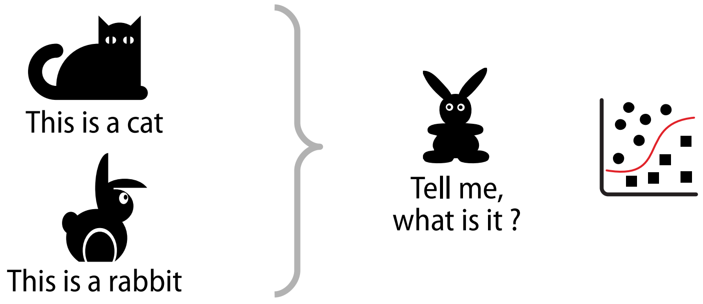


Learning from labeled data



Classification :

Predict qualitative informations

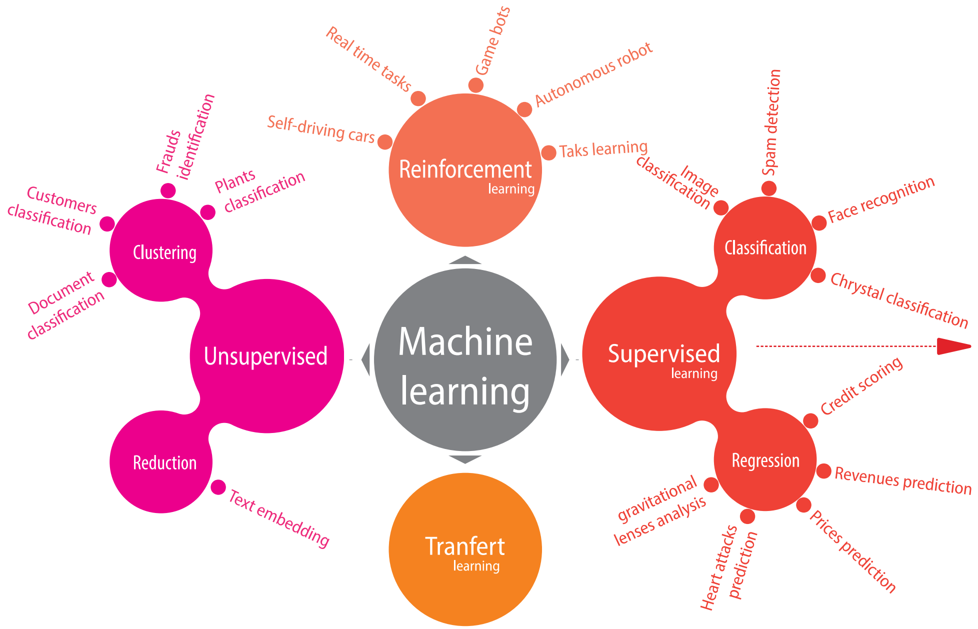


Régression :

Predict quantitative informations

Supervised learning

Learning from labeled data



Self-supervised :

Find the label inside the data !!

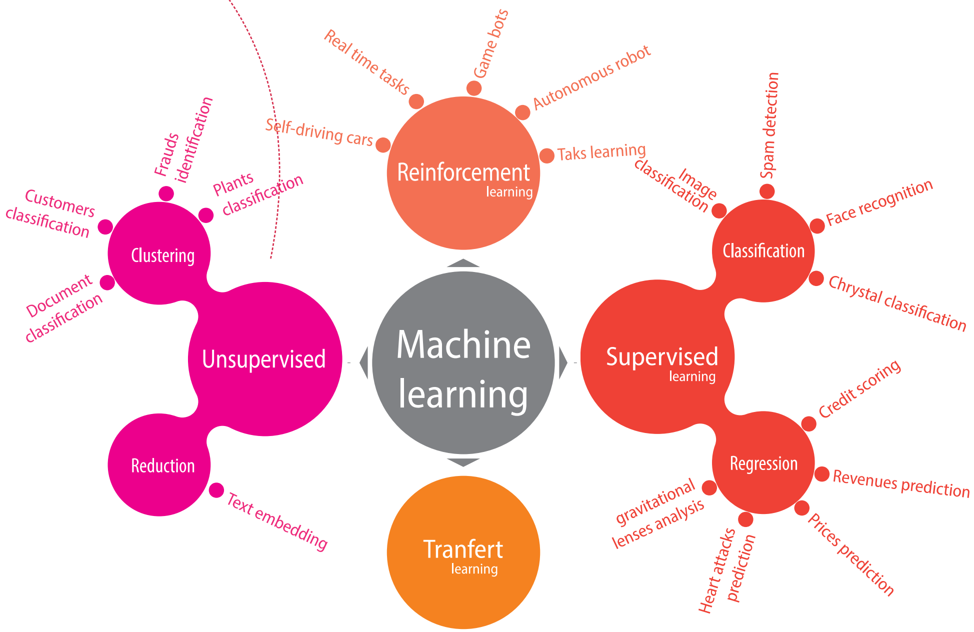
Je crois qu'il profita, pour son évasion, d'une migration d'oiseaux sauvages. Au matin du départ il mit sa planète bien en ordre. Il ramona soigneusement ses volcans en activité.

Il possédait deux volcans en activité. Et c'était bien commode pour faire chauffer le petit déjeuner du matin.

LLM, VLM

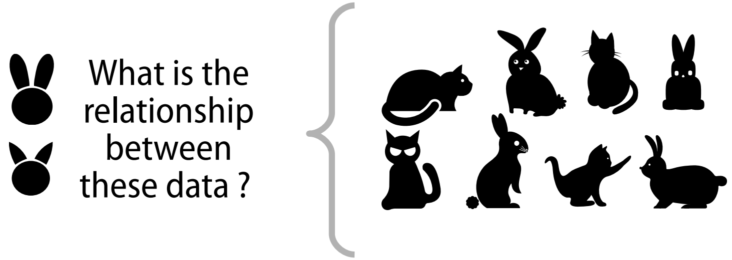
ex: chat GPT, GPT4, …

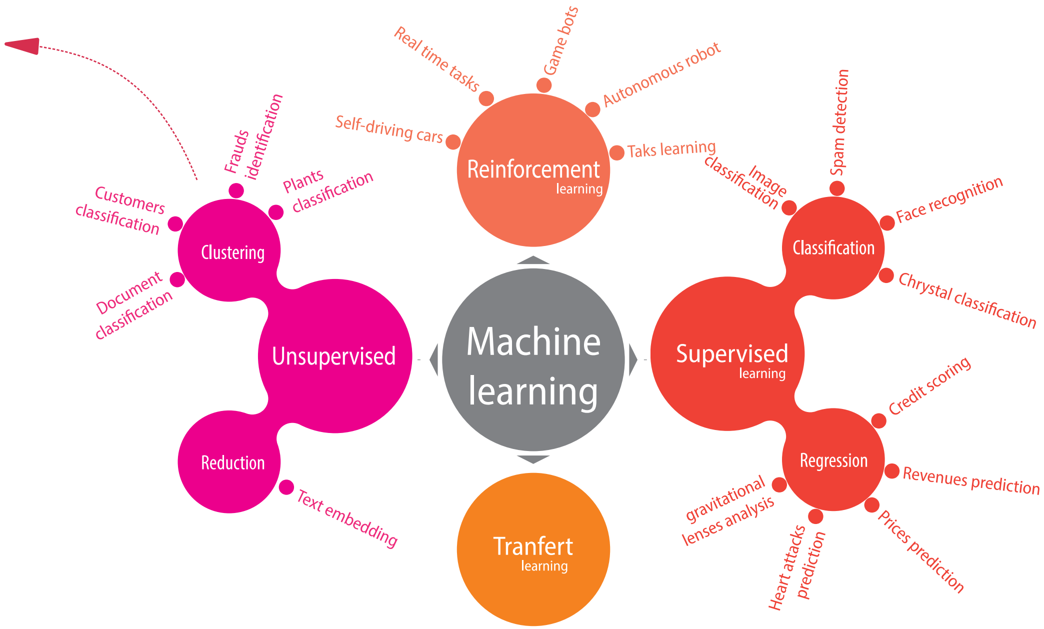
Learning
from data **alone**



Learning from data alone

Clustering :
Finding Common Relationships



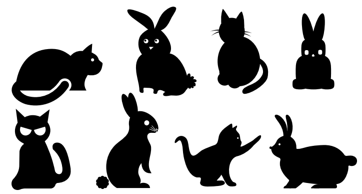


Unsupervised learning

Learning from data alone

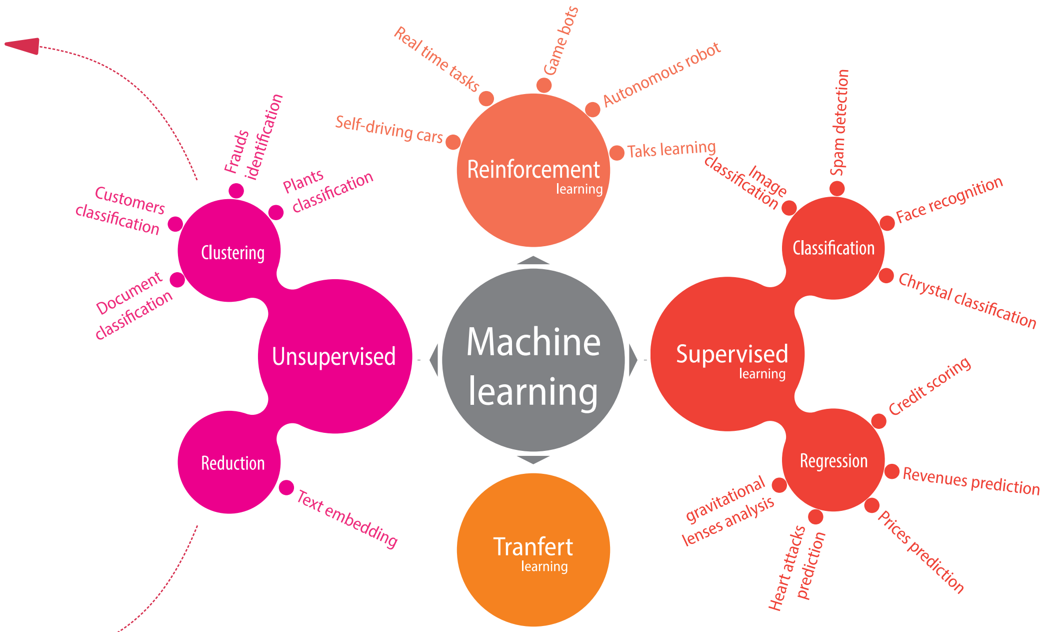
Clustering :
Finding Common Relationships

What is the relationship between these data ?

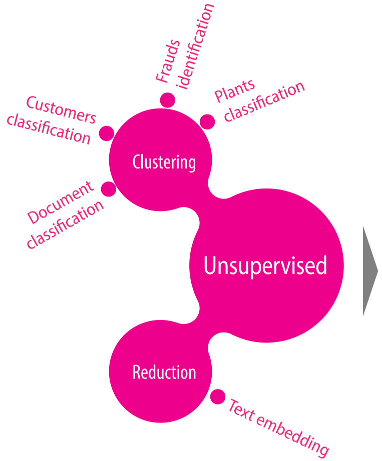


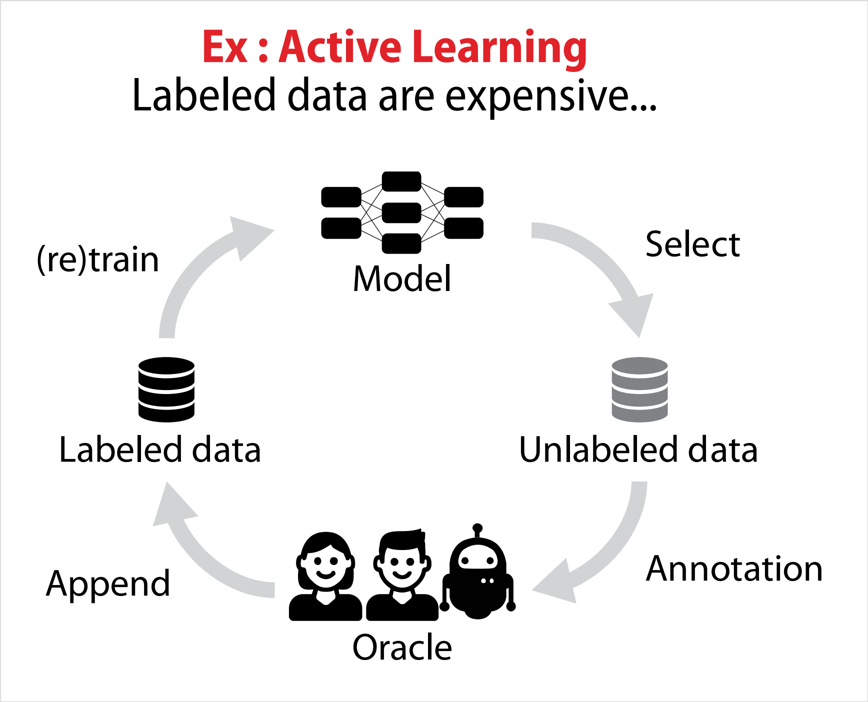
Reduction :
Reduce the number of dimensions

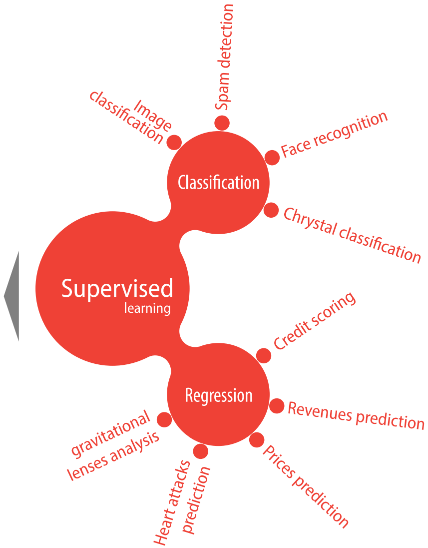
Simplify while keeping meaning



Semi-supervised learning

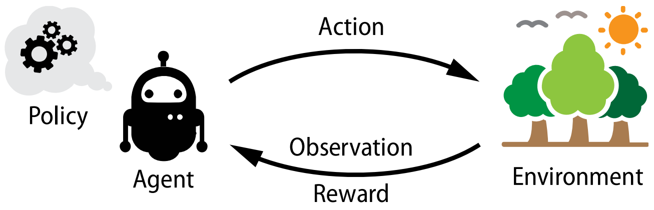


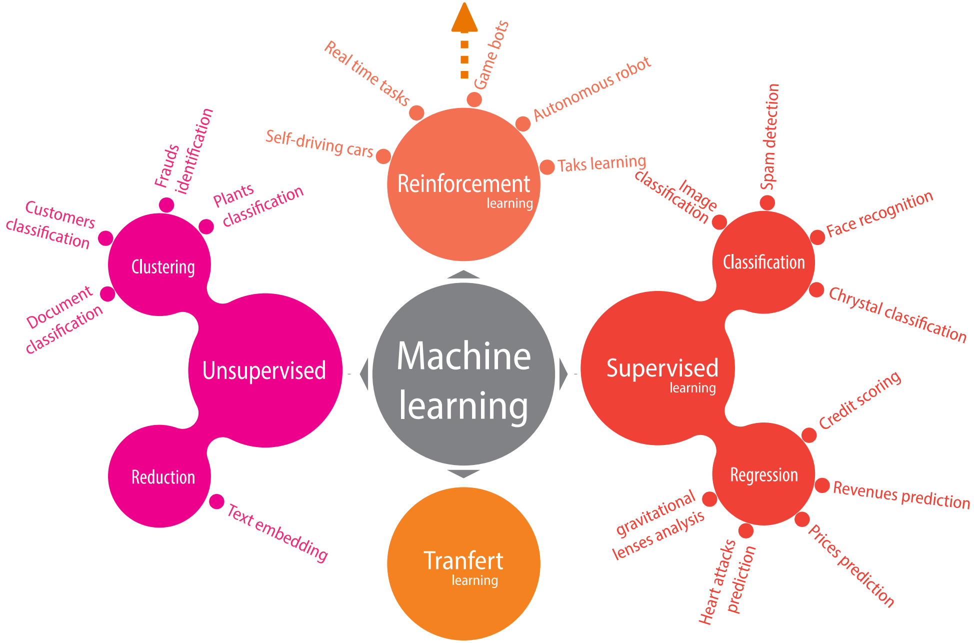




Learning from an environment

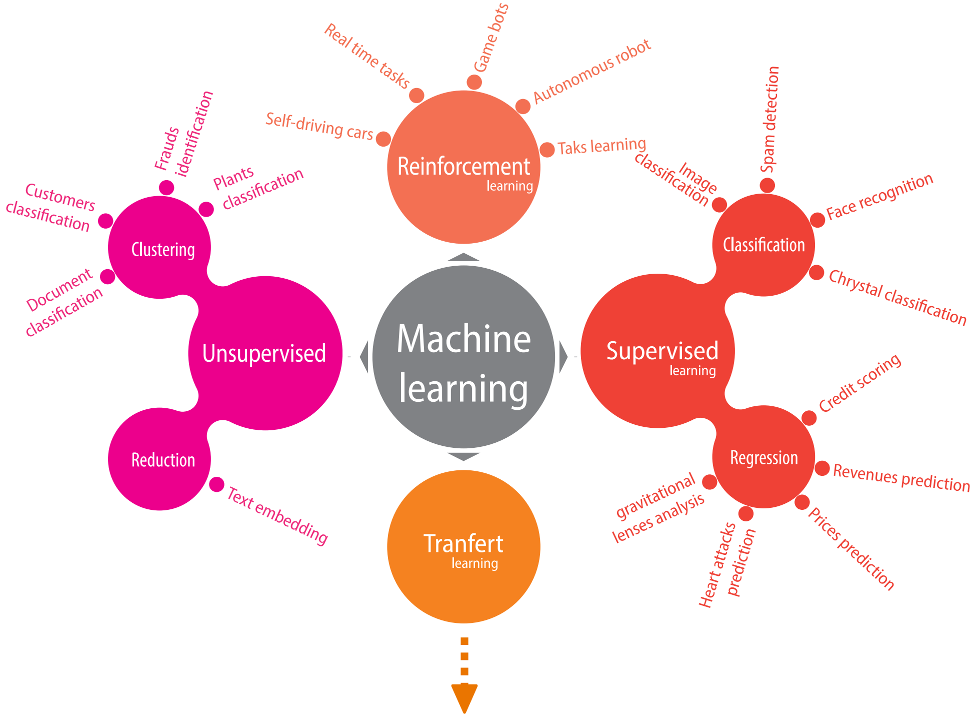
Reinforcement :
Interact "positively" in an environment



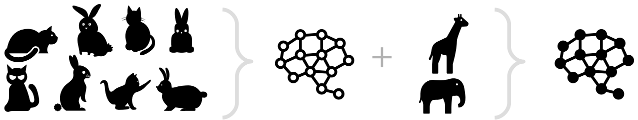


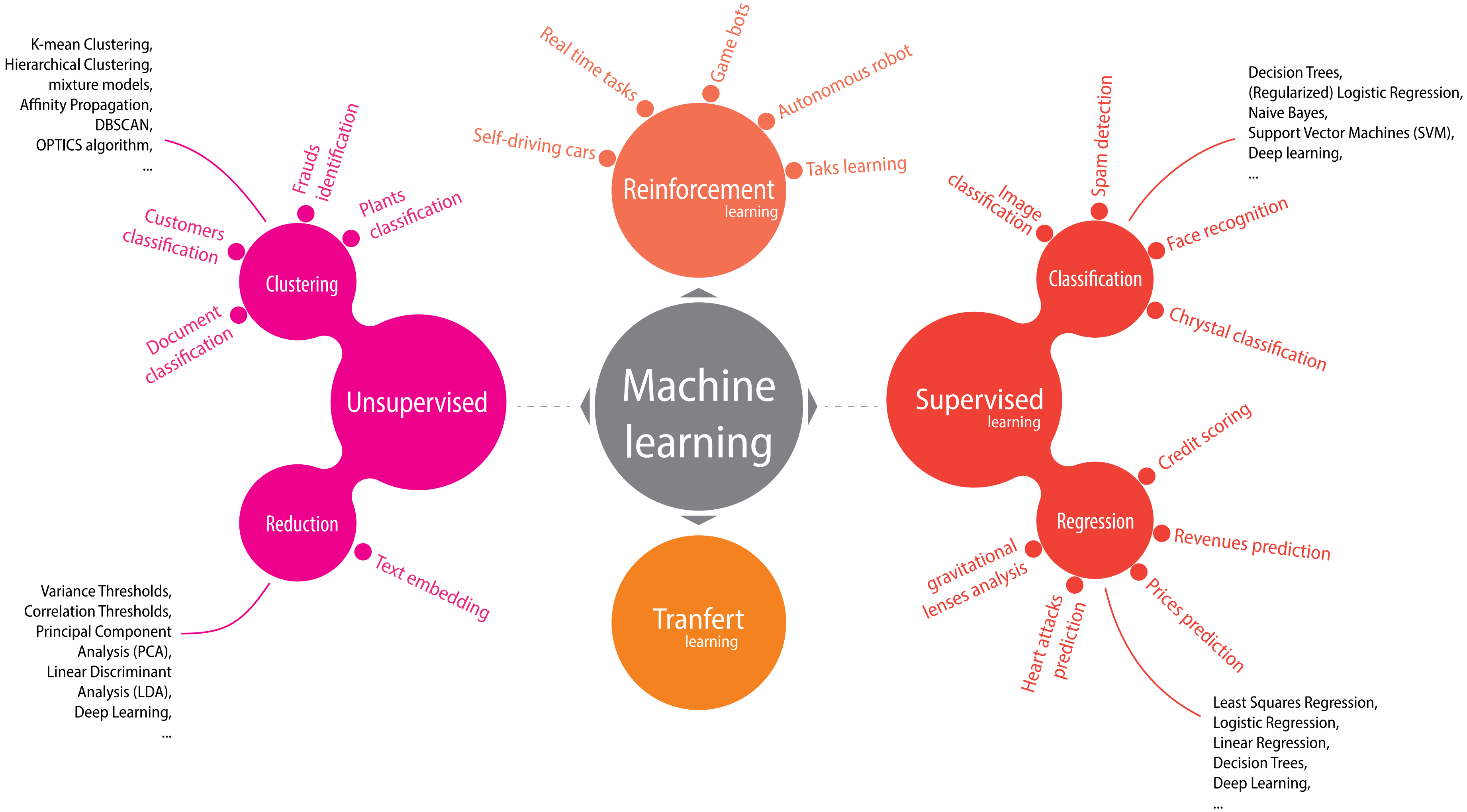
Unsupervised learning

Expand your skills



Transfer learning :
Extend, complete or specialize a training





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


ROYAL
MILAN

A Python centered world

python™ NumPy jupyter

« Deep learning for humans »

Widely used in the implementation of practical solutions

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 Lightning

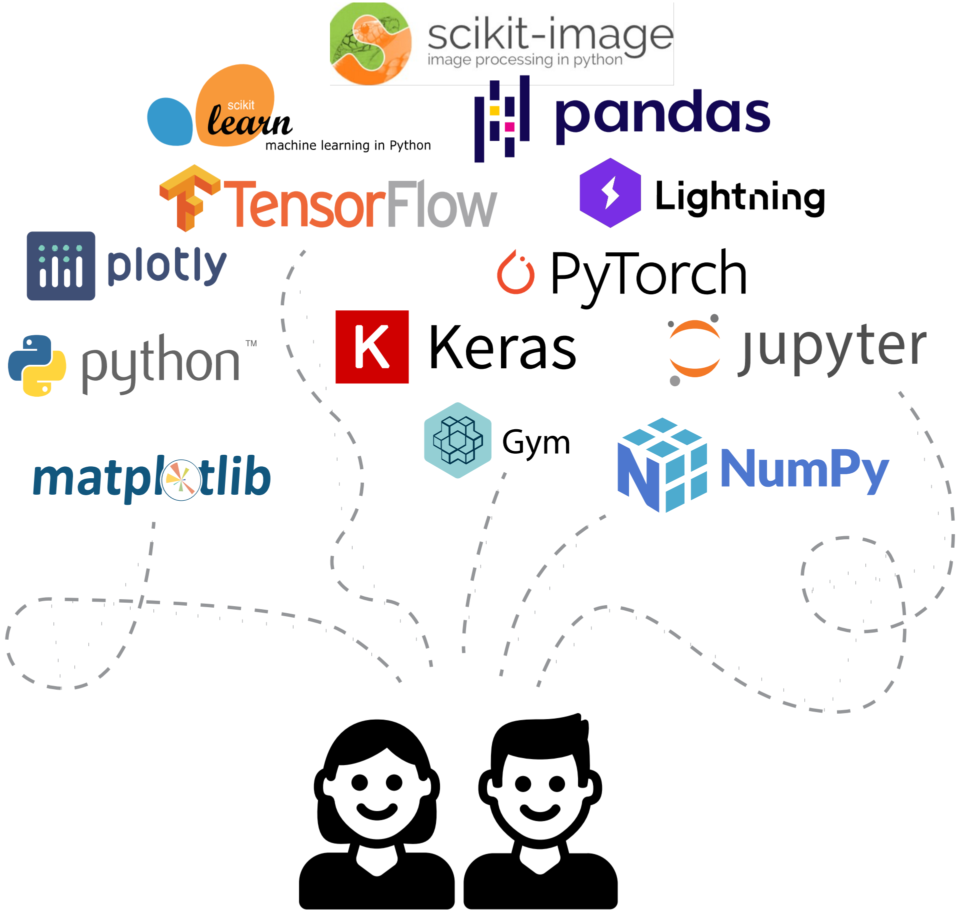
High-level interface for PyTorch, by William Falcon.
Lightning 2.0 is featuring a clean and stable API!!

PyTorch

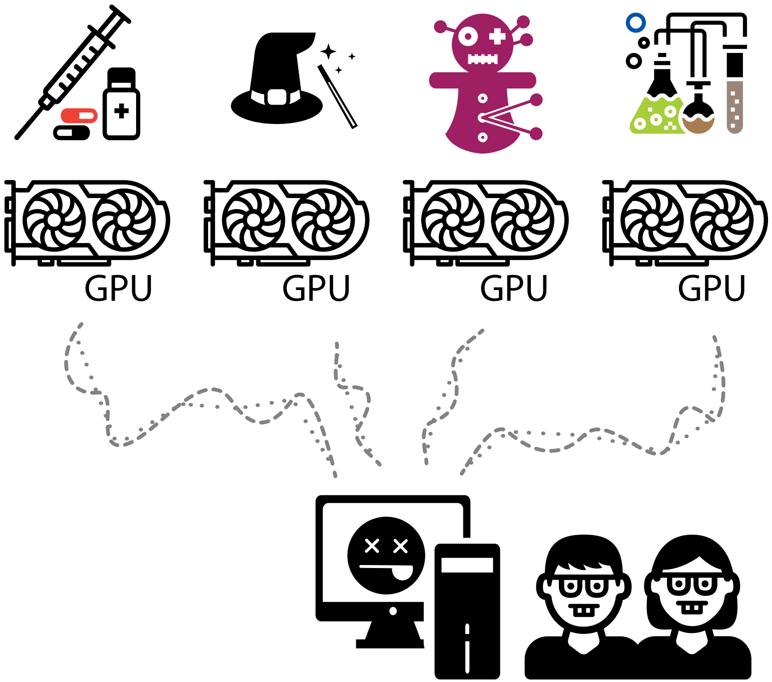
From Torch library
Supported by Facebook
BSD licence

Widely used in the field of AI research

A certain complexity...



Don't become a geek !!



A certain complexity...

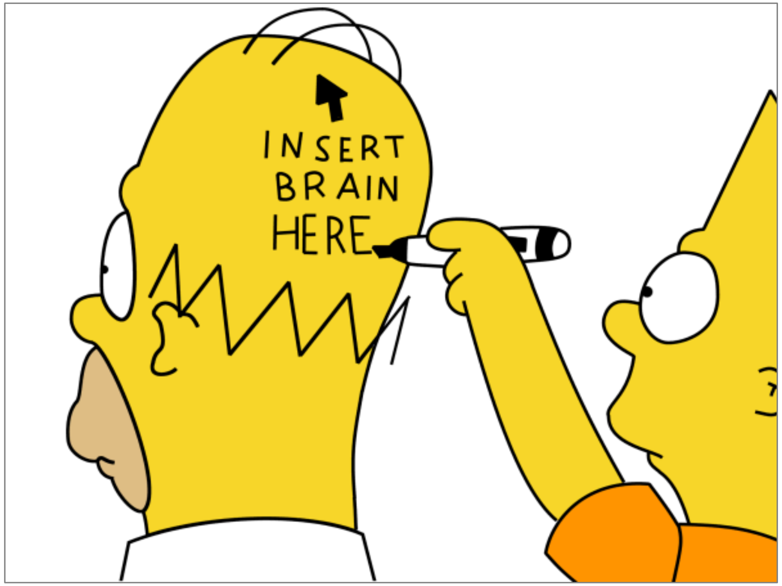
GENCI

Le calcul intensif au service de la connaissance

cnrs

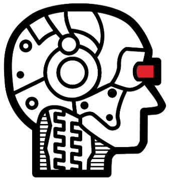
Institut du Développement et des Ressources en Informatique Scientifique

Jean Zay
3200 GPU



Fine, but Deep Learning What's that?

1



History and **Fundamental Concepts**

1 **Deep Learning**, what are we talking about?

2 From linear regression to **artificial neurons**

3 80 years of history ...and the **victory** of **neurons**

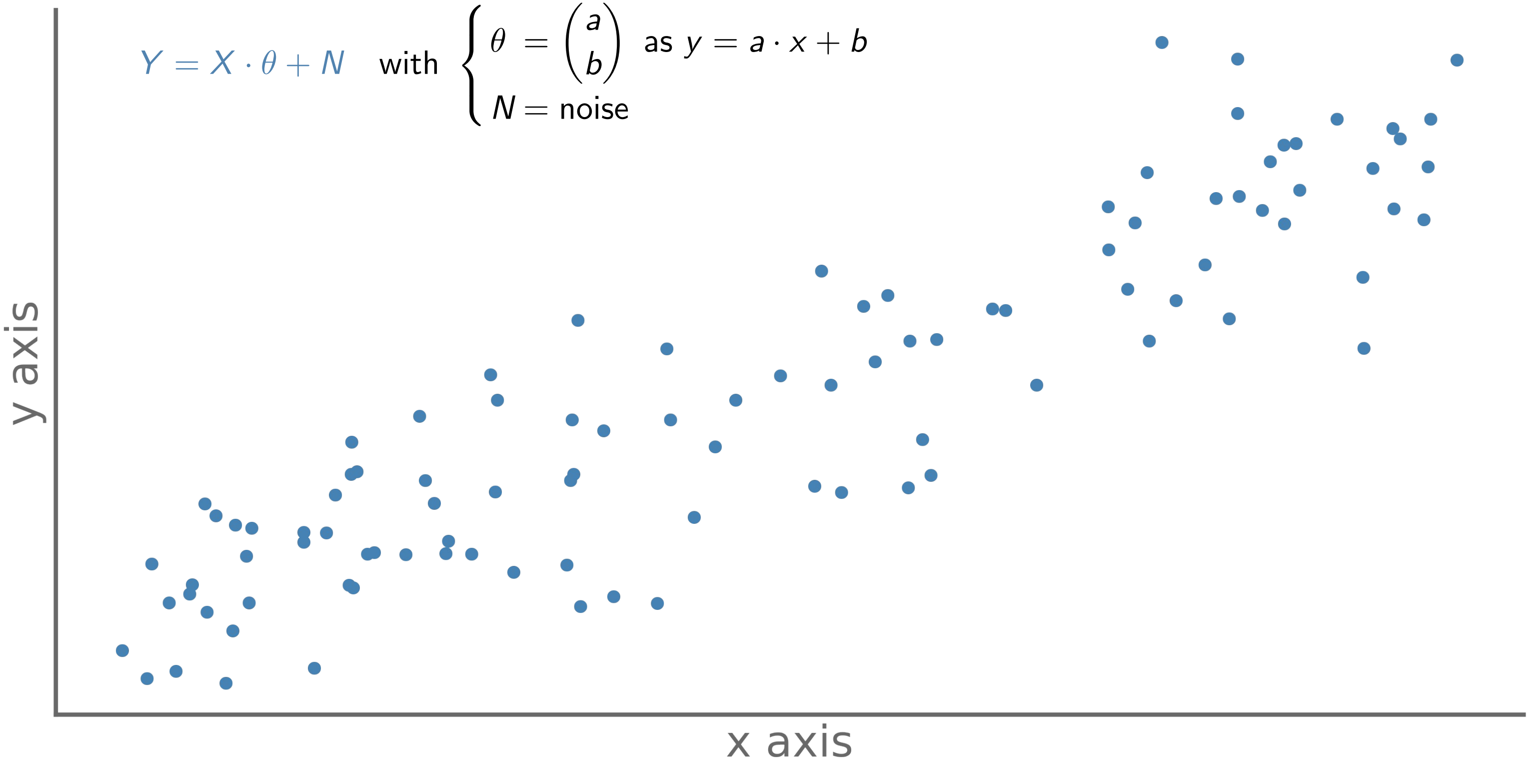
4 Concrete **illustration** :

Basic Regression
Basic Classification

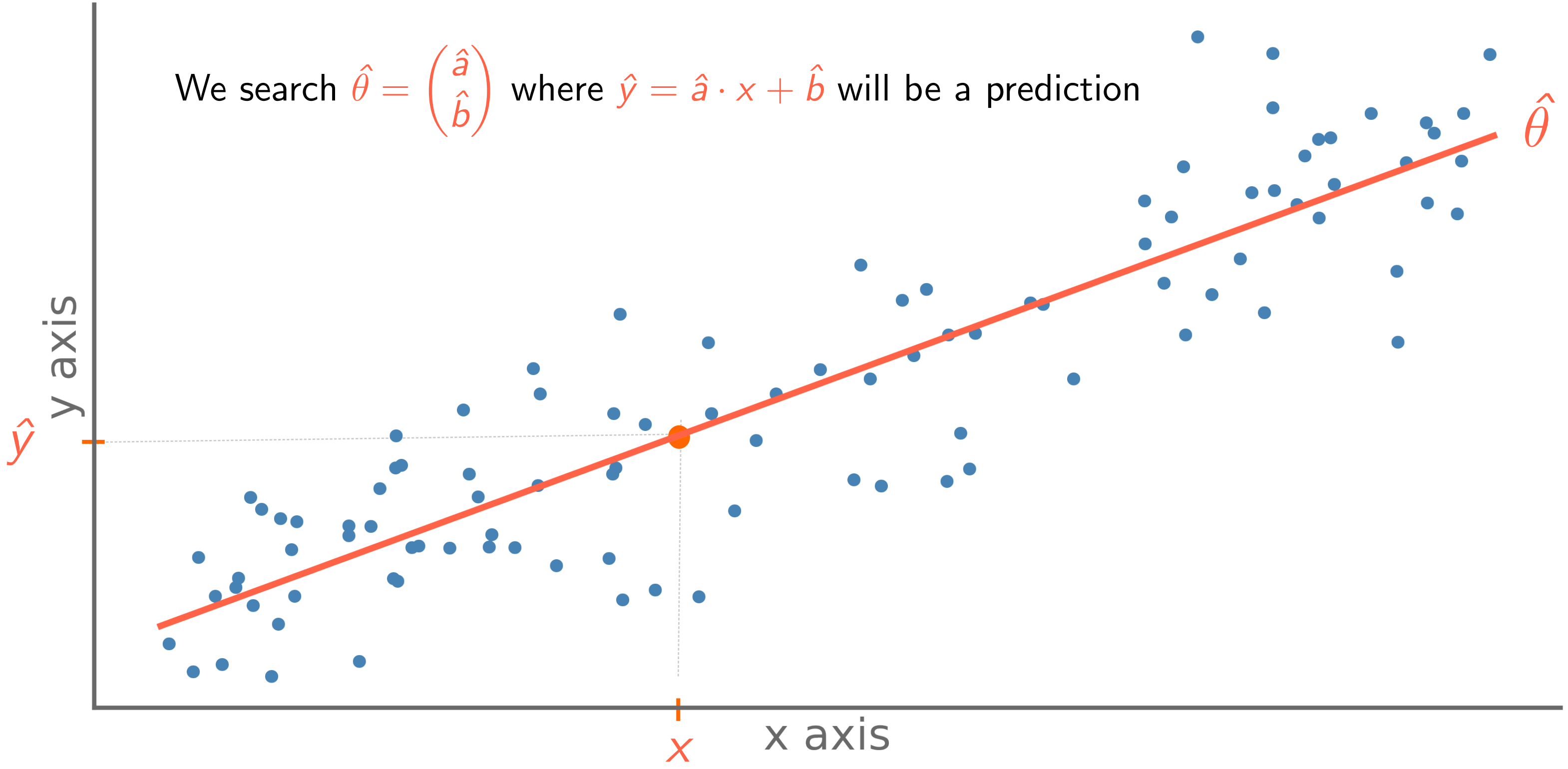
Linear regression

We have a phenomenon, for which we have observations

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

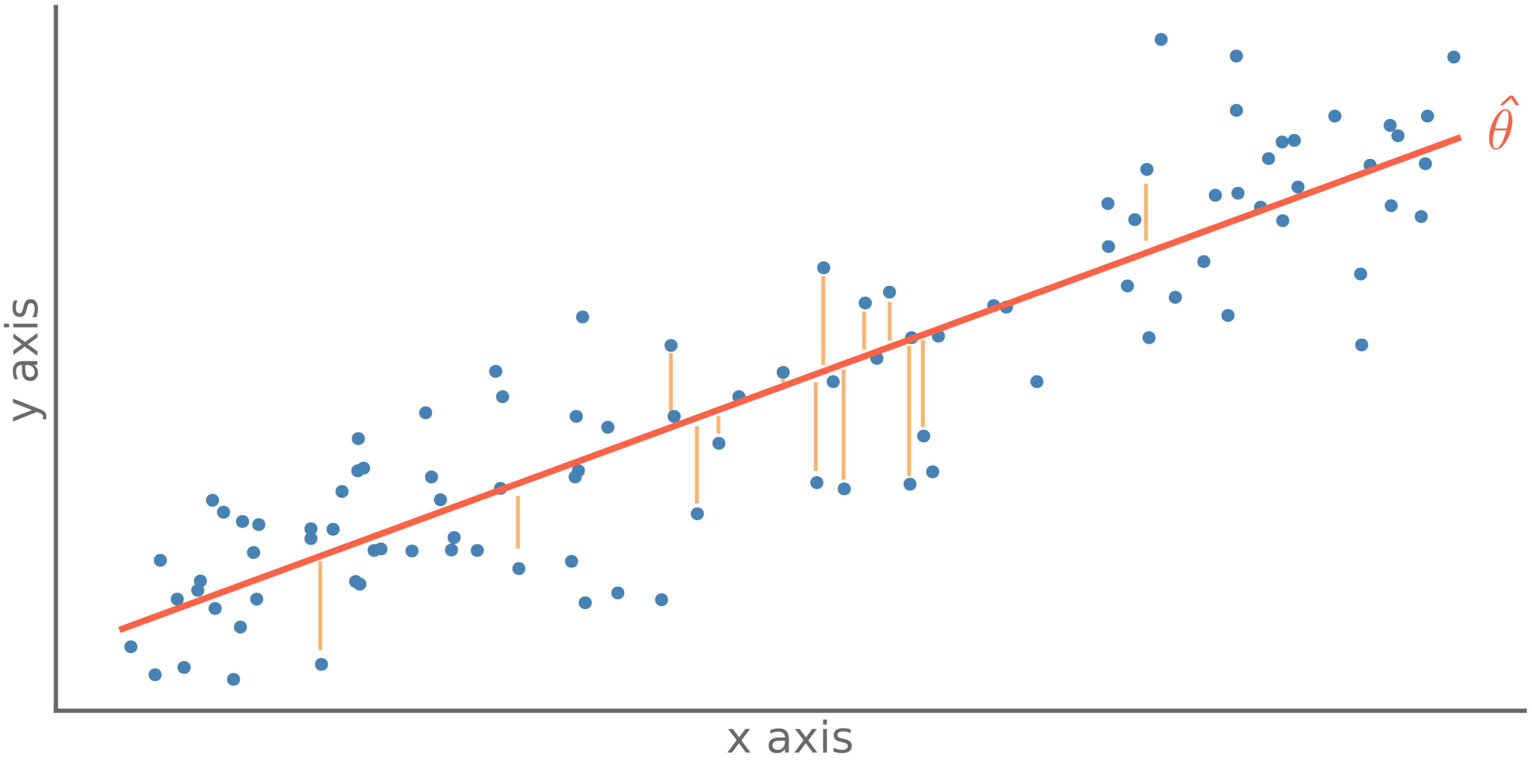


We are looking for a straight line that passes « as close as possible » to our points.



Linear regression

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

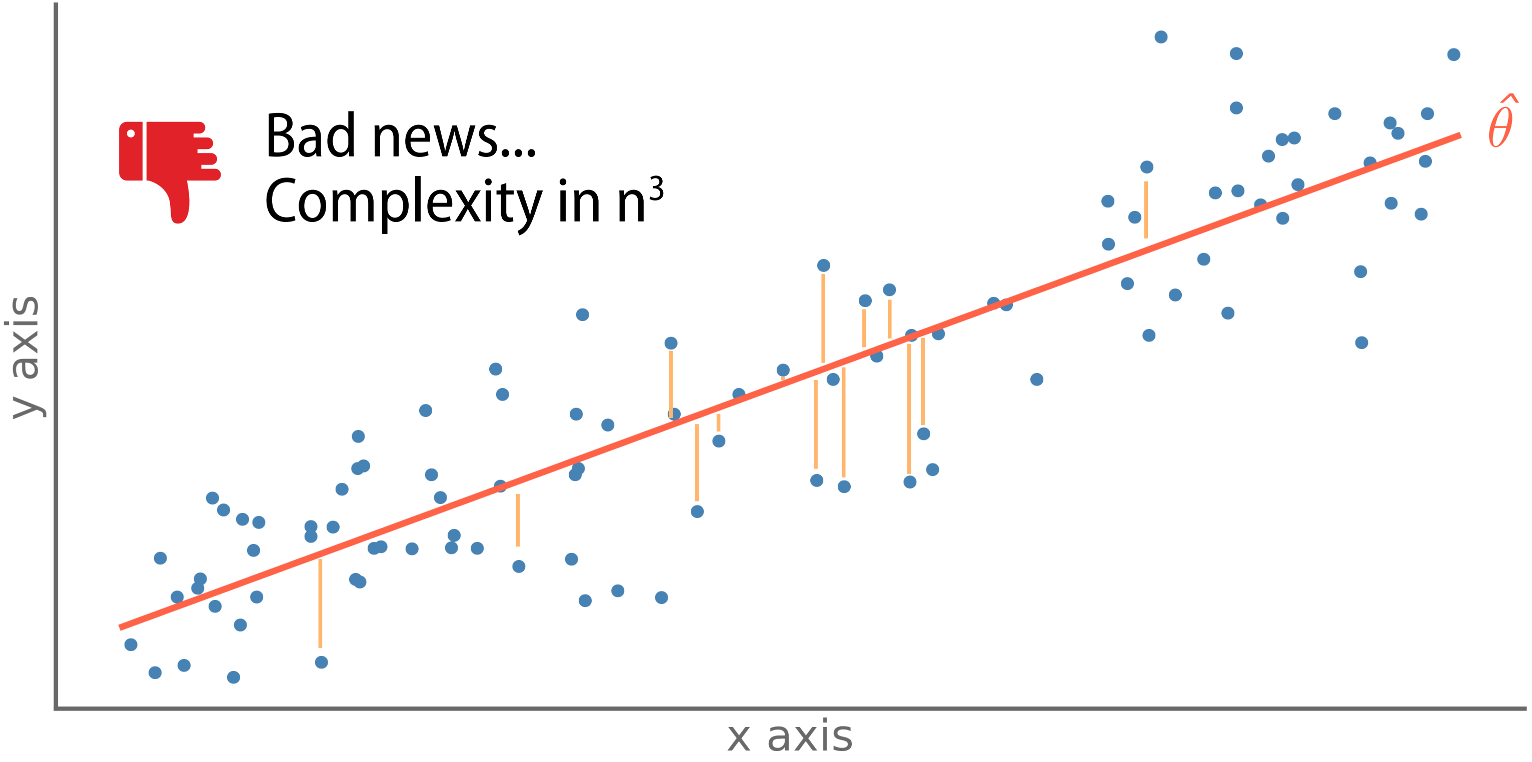


Linear regression

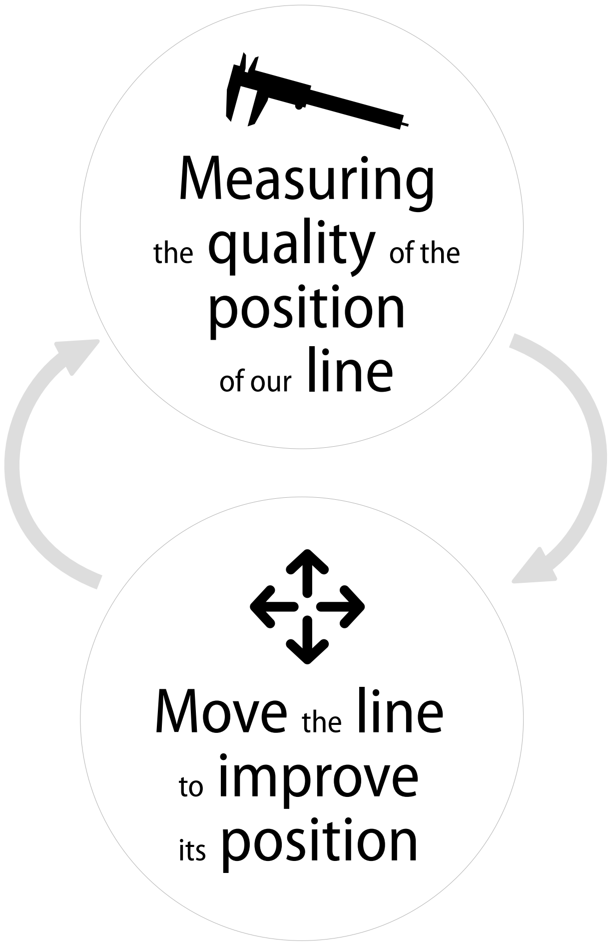
$$\hat{\theta} = (X^{-T}.X)^{-1}.X^{-T}.Y$$

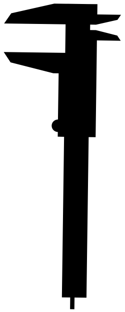
Good news !
We have a direct solution !

Bad news...
Complexity in n^3



Linear regression

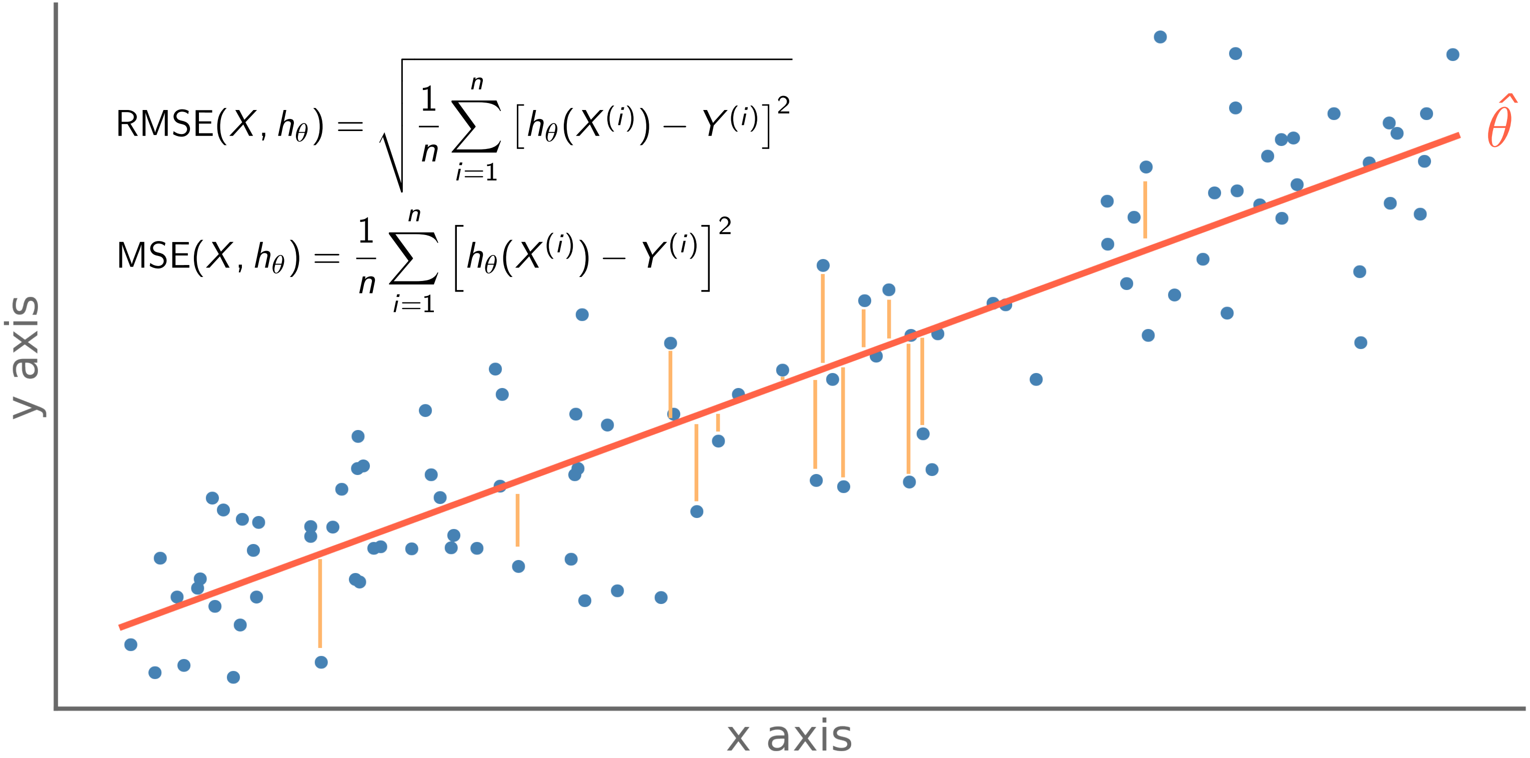




We will use an «loss function », which we will try to minimize.

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

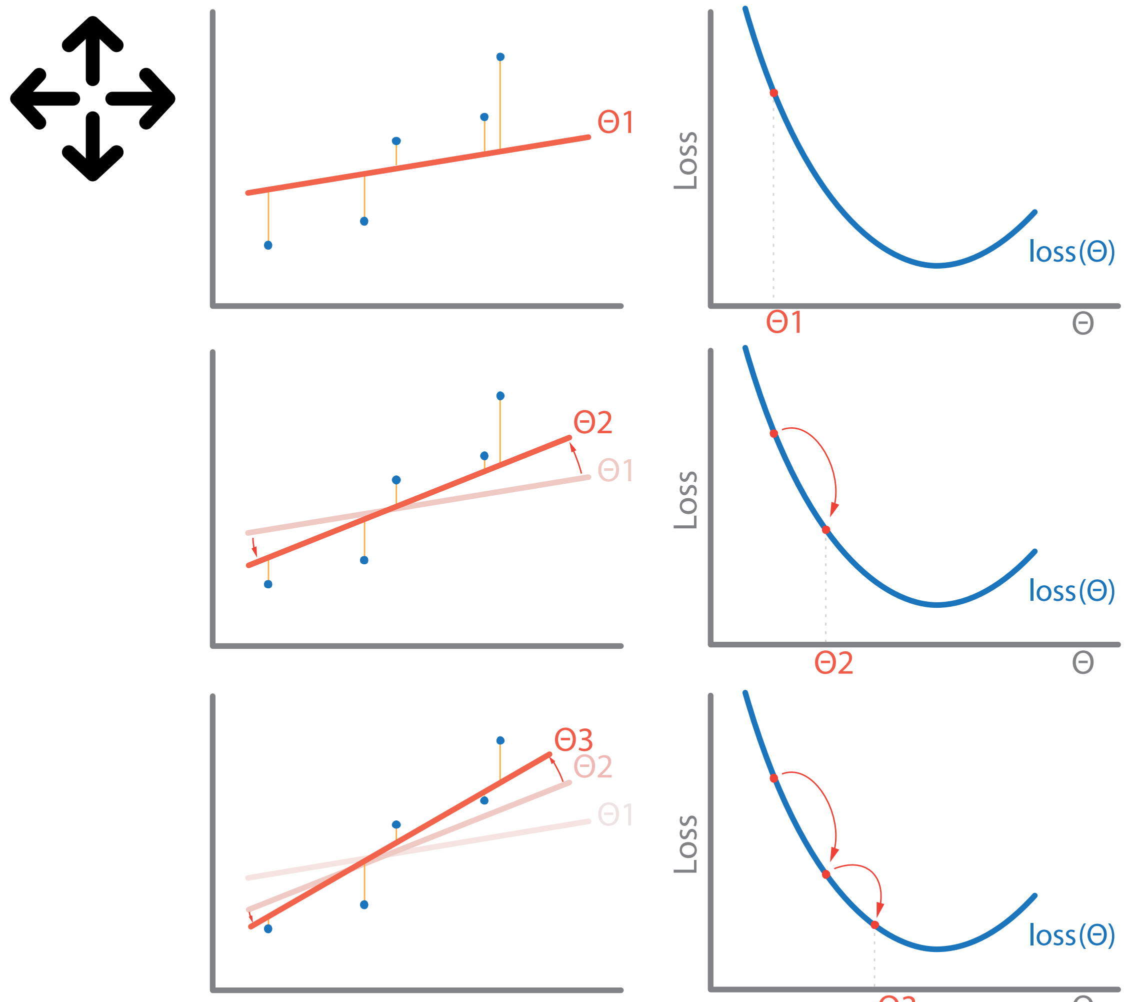
$$\mathrm{MSE}(X, h_\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[h_\theta(X^{(i)}) - Y^{(i)}\right]^2$$



RMSE : Root Mean Square Error - Erreur quadratique moyenne

MSE : Mean Squared Error - Moyenne du carré des erreurs

Optimization



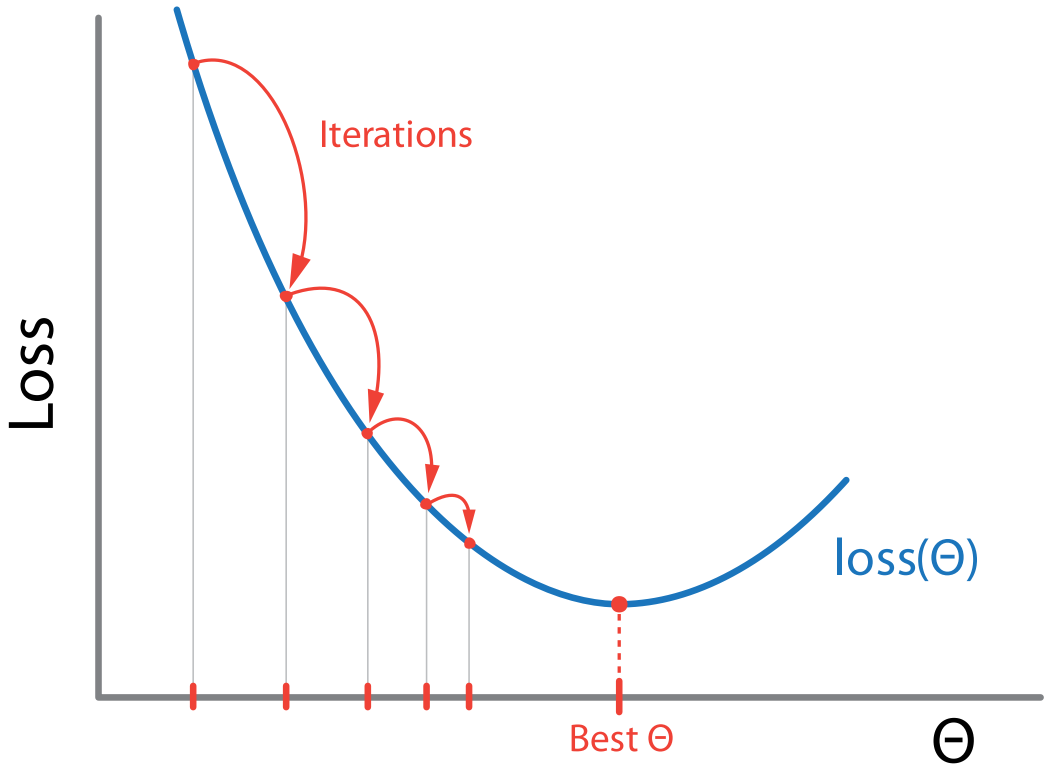
We will iteratively look for the best position of our line, by varying its parameters (Θ).

This is called optimization

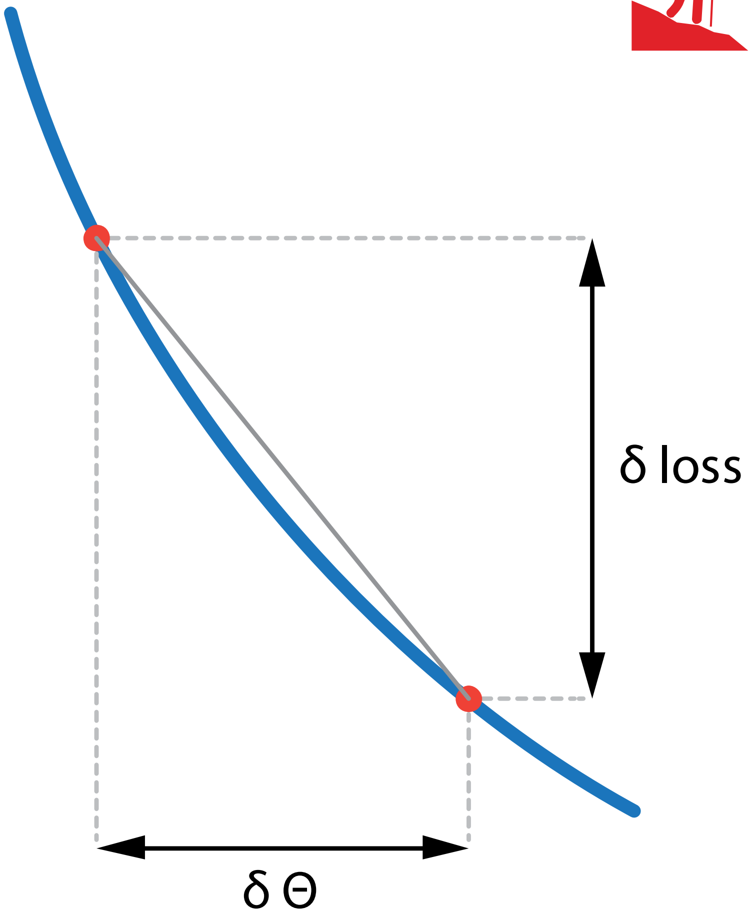
Optimization

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

This is called optimization



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

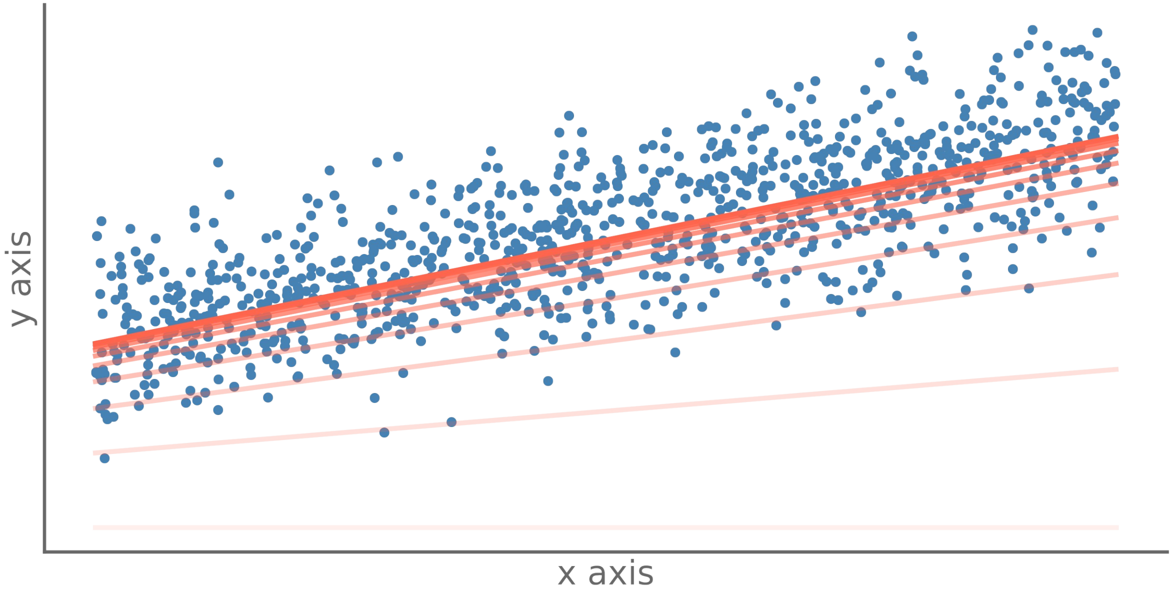
Gradient descent

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

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

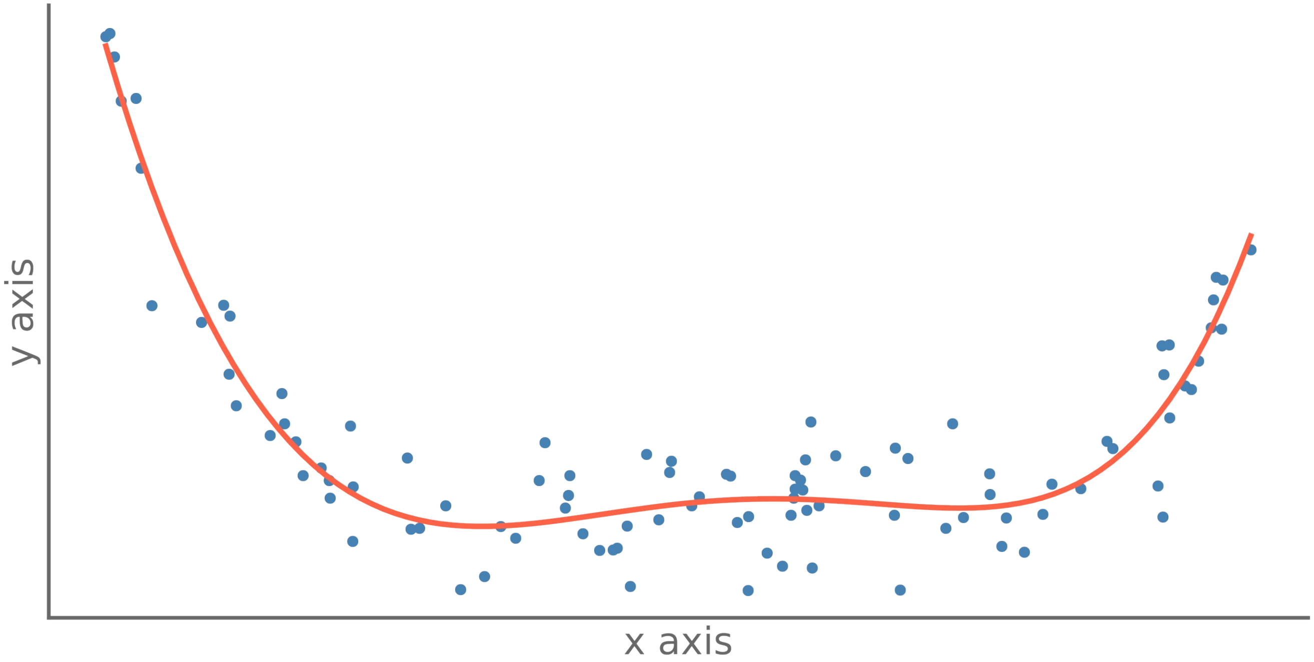
n : number of observations
m : number of characteristics

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



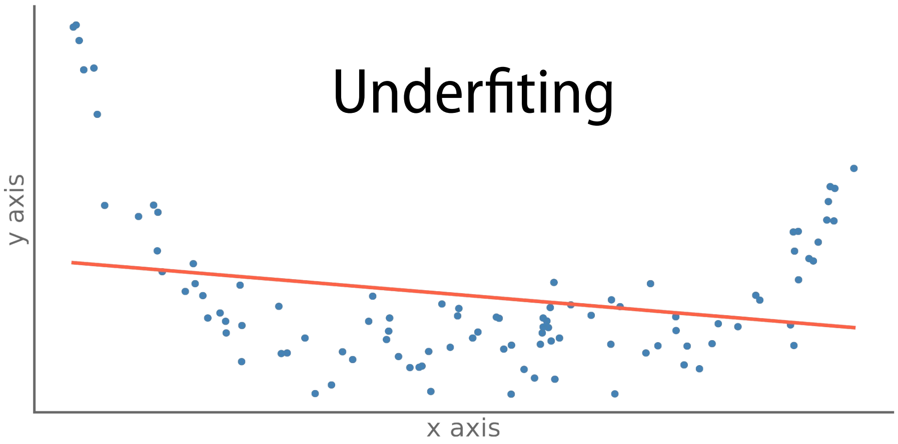
#i	Loss	Gradient		Theta	
0	+12.481	-6.777	-1.732	-3.388	+0.000
20	+4.653	-4.066	-1.039	-2.033	+0.346
40	+1.835	-2.440	-0.624	-1.220	+0.554
60	+0.821	-1.464	-0.374	-0.732	+0.679
80	+0.455	-0.878	-0.224	-0.439	+0.754
100	+0.324	-0.527	-0.135	-0.263	+0.799
120	+0.277	-0.316	-0.081	-0.158	+0.826
140	+0.260	-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

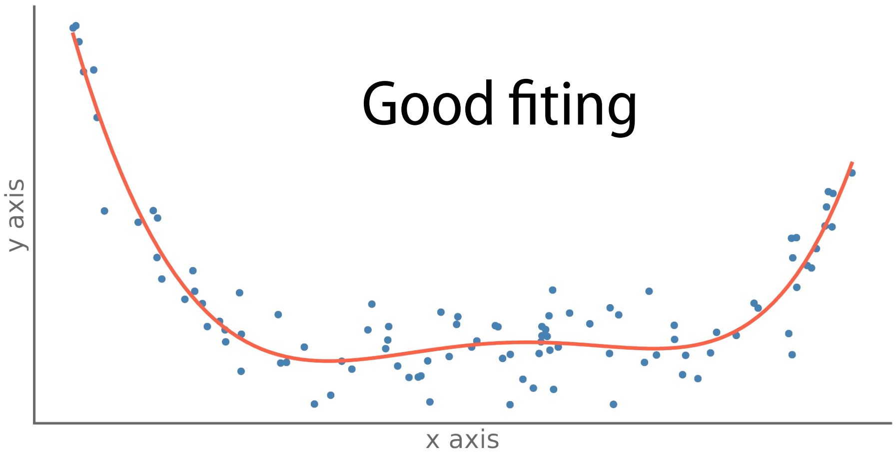
Polynomial regression

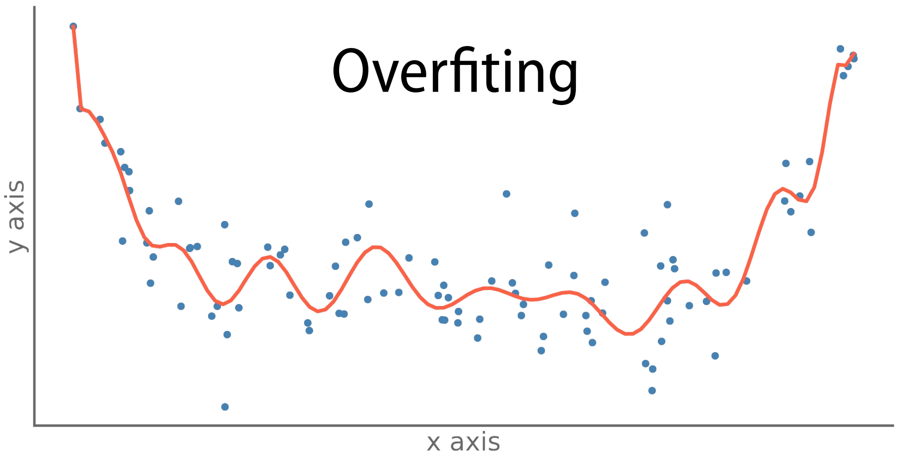


$$P_n(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \cdots + a_n \cdot x^n = \sum_{i=0}^{n} a_i \cdot x^n$$

Underfiting and overfiting



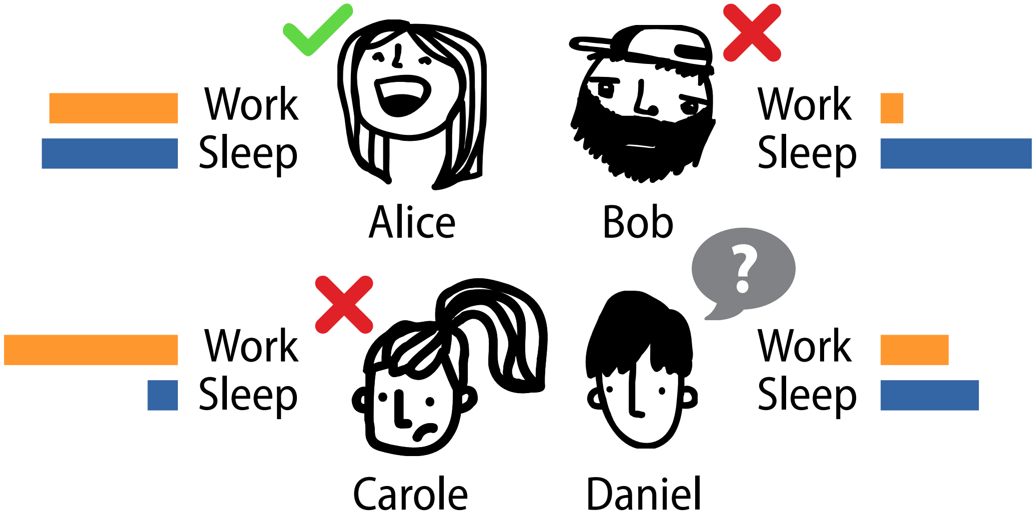




Great, but where's my first artificial neuron?

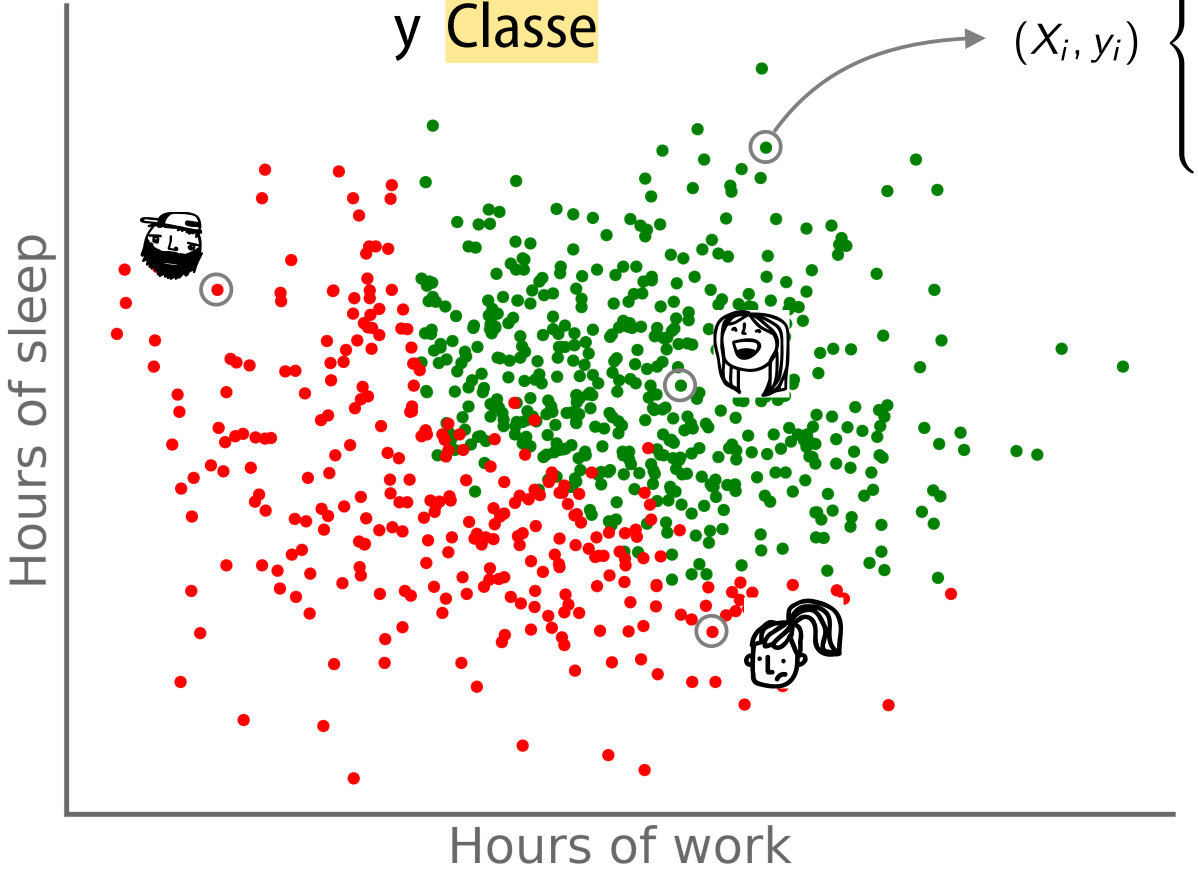
Okay, let's look at an example !

Let's try to predict our students' success in an exam, based on their work time and sleep time:



Logistic regression

Dataset : X Observations
y Classe

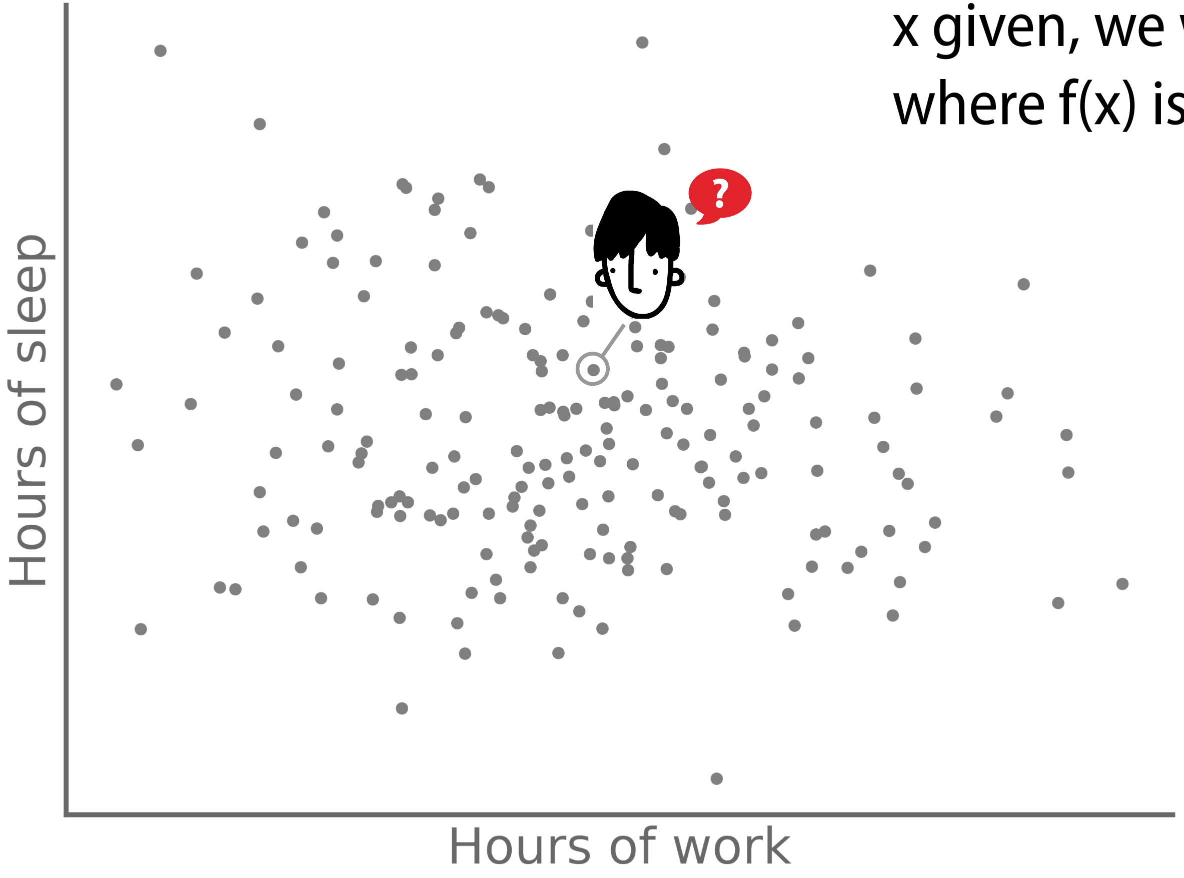


$$(X_i, y_i) \begin{cases} X_i = \begin{pmatrix} x_{i1} = \text{Hours of work} \\ x_{i2} = \text{Hours of sleep} \end{pmatrix} \\ y_i = \begin{cases} 1 & \text{Successful examination} \\ 0 & \text{Failed exam} \end{cases} \end{cases}$$

Objective : Predict the class !

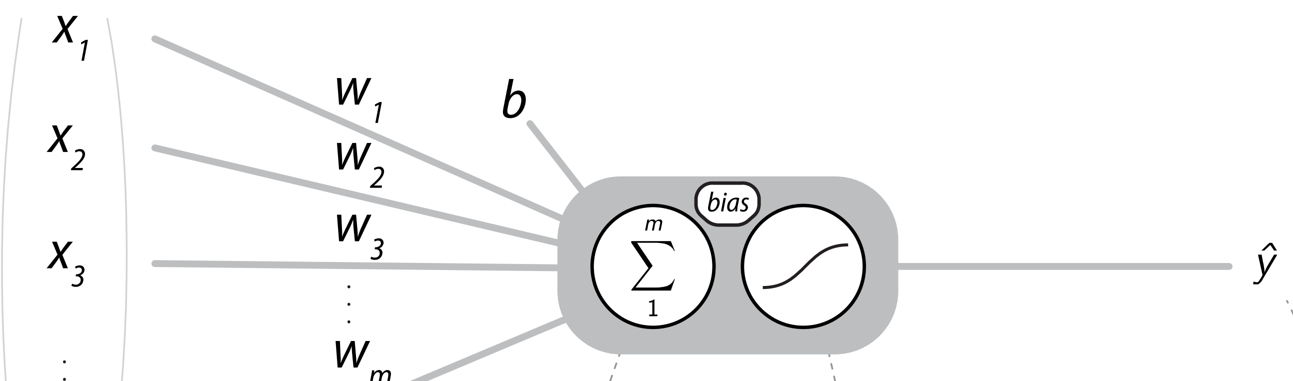
x given, we want to predict : $\mathbf{y_{pred} = f(x)}$

where f(x) is a linear function



Logistic regression

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



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

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

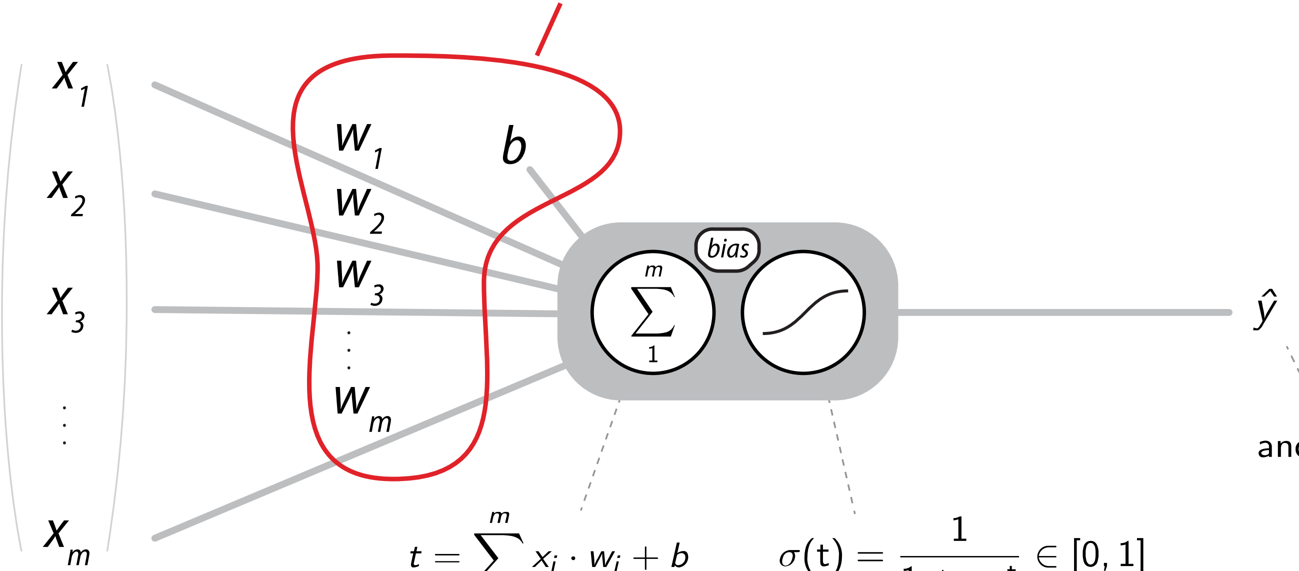
$$\text{and } \bar{y} = \begin{cases} 0 & \text{if } \hat{y} < 0.5 \text{ Failed} \\ 1 & \text{if } \hat{y} \geq 0.5 \text{ Success} \end{cases}$$

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

Logistic regression

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

Determined by the minimisation of a cost function J(Θ)



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

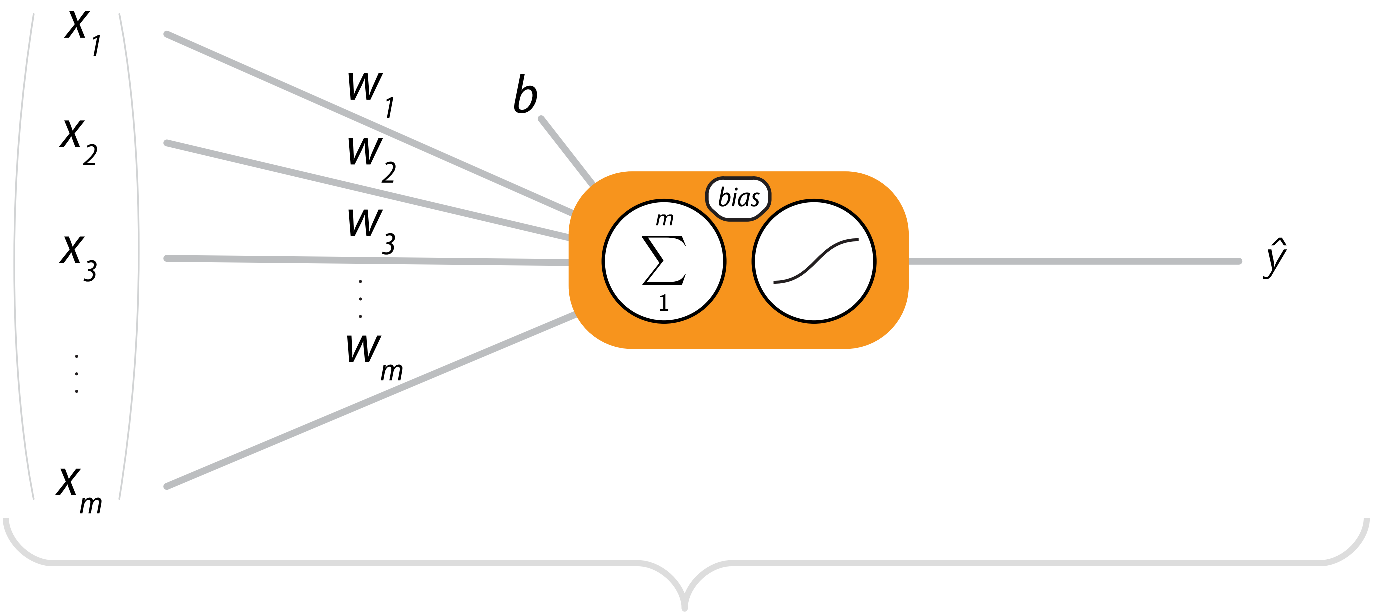
$$\sigma(\mathrm{t}) = \frac{1}{1 + \mathrm{e}^{-\mathrm{t}}} \in [0, 1]$$

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

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

Logistic regression

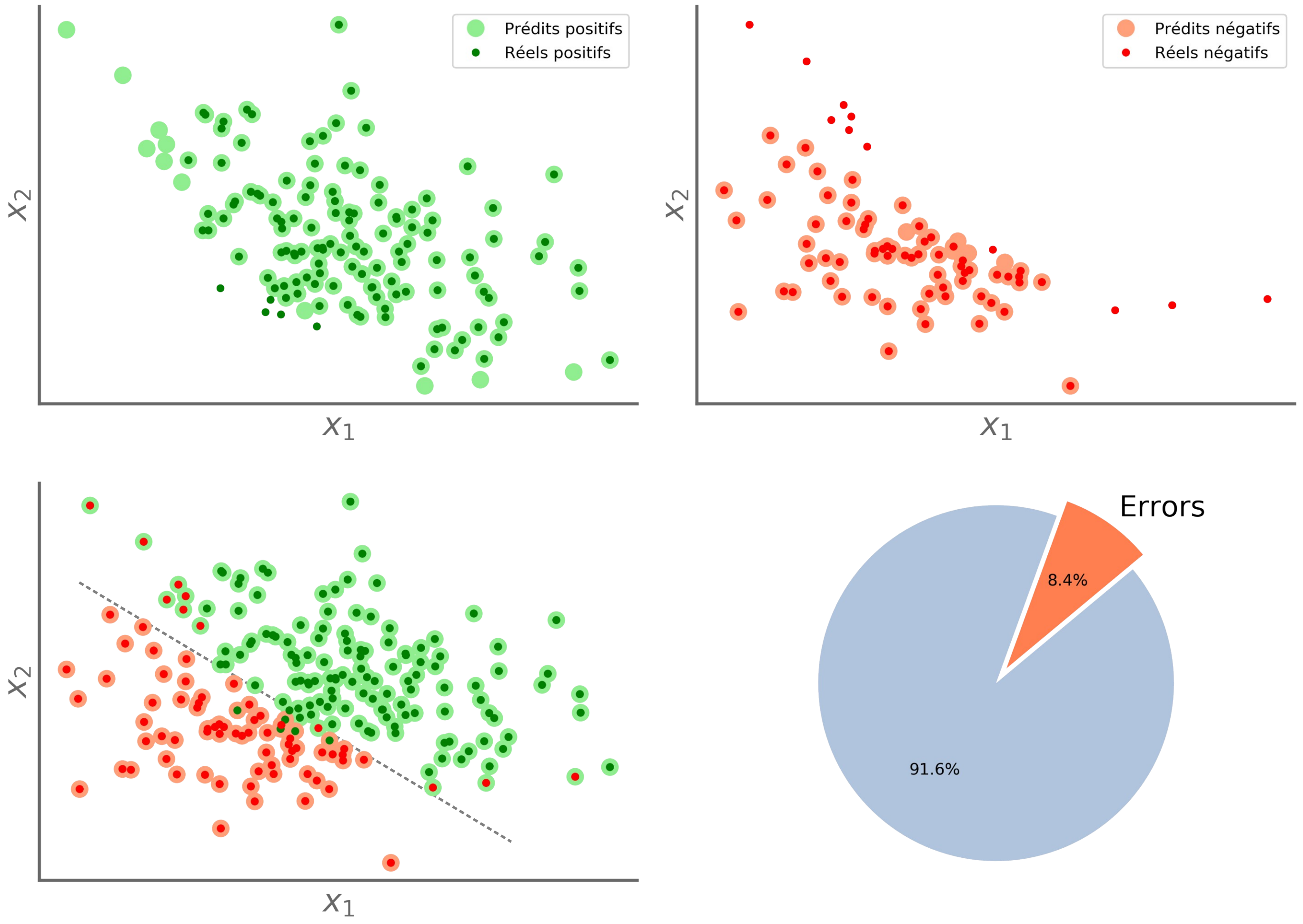
$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$



That's an « artificial neuron » !

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

Logistic regression

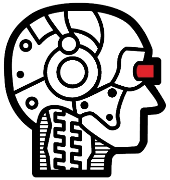


Fidle notebook [LOGR1] - https://fidle.cnrs.fr

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

Help !
No more maths…
Pleaaase ! ;-)

1



History and **Fundamental Concepts**

1. **Deep Learning**, what are we talking about?
2. From linear regression to **artificial neurons**
3. 80 years of history ...and the **victory** of **neurons**
4. Concrete **illustration** :
   - Basic Regression
   - Basic Classification

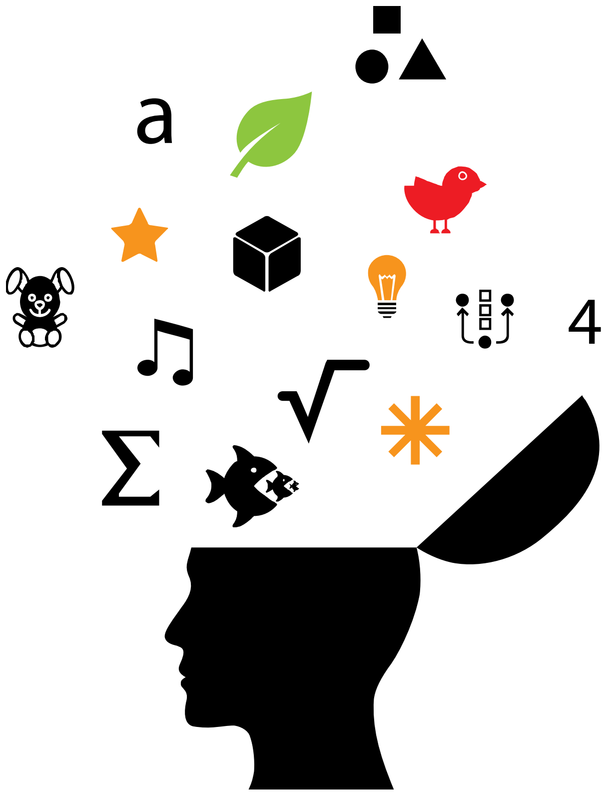
[intelligence]

[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



[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**.

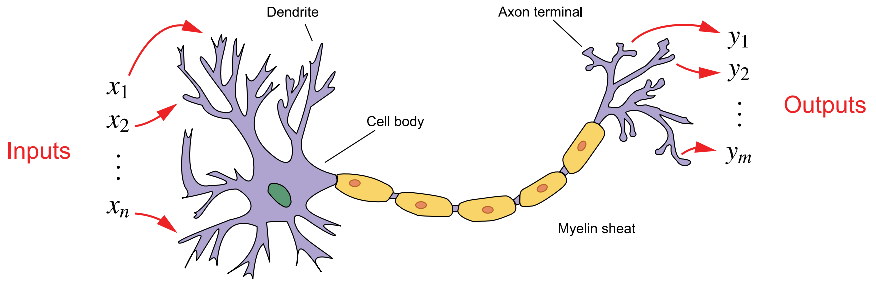
Connectionnism vs Symbolic

Connectionnism

Modelling the brain
Modéliser le cerveau

Symbolic

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), La revanche des neurones
https://dx.doi.org/10.3917/res.211.0173

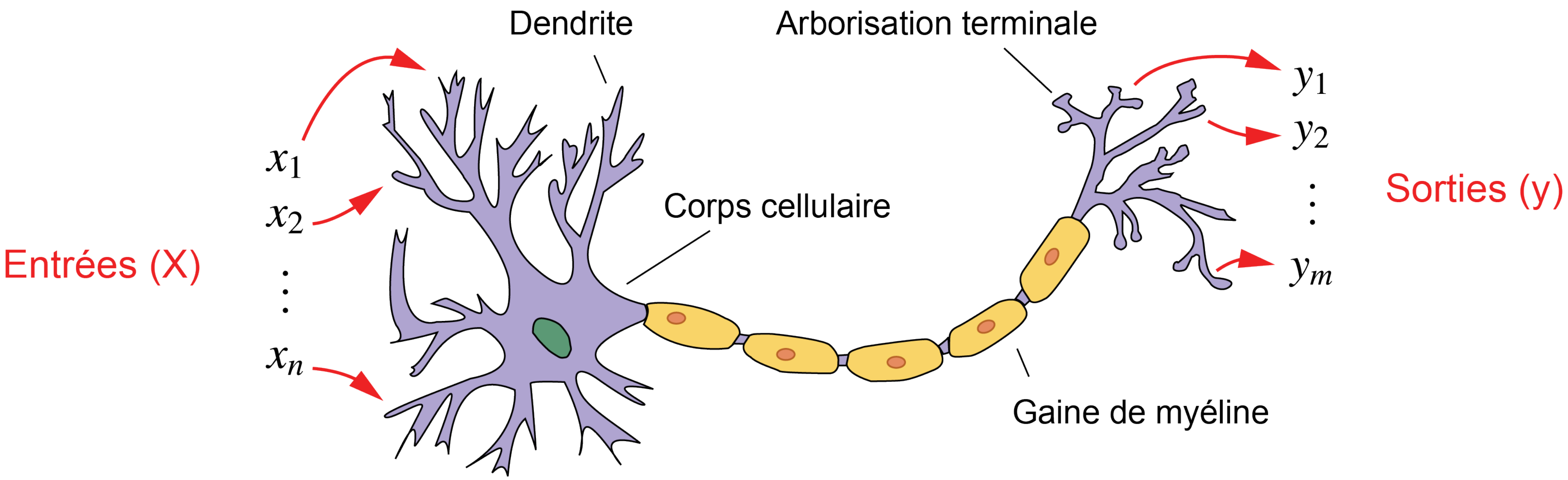
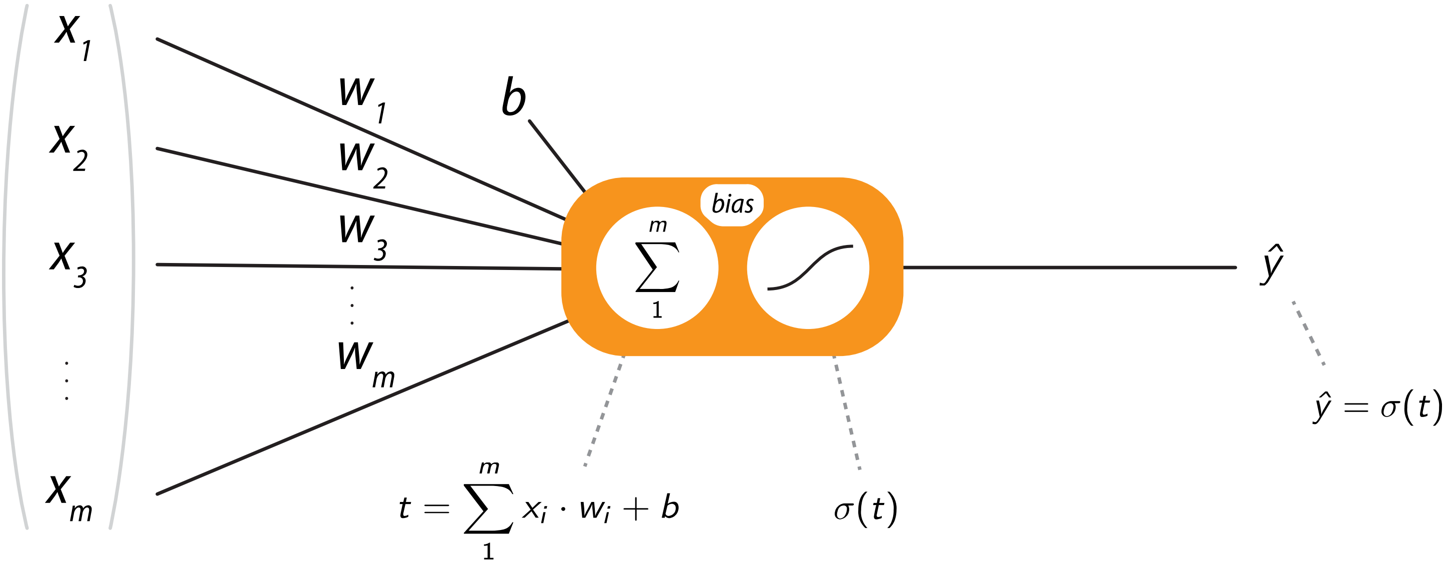


Illustration : Wikipédia

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



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 :

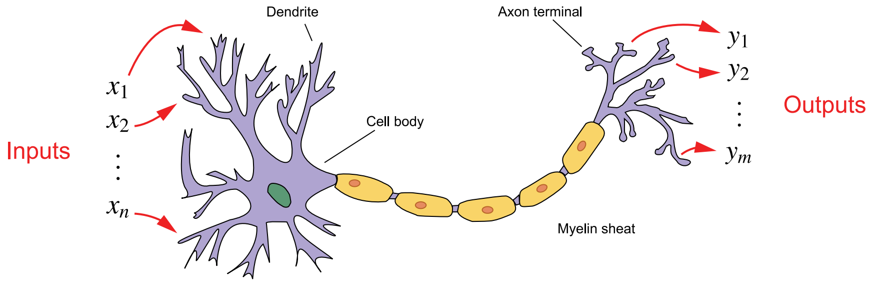
« Penser, c'est calculer des **symboles** qui ont à la fois une réalité matérielle et une valeur sémantique de représentation »[1]

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), La revanche des neurones https://dx.doi.org/10.3917/res.211.0173

The big Controversy

Inductive approach

Input → Program (model)
Output

Deductive approach

Input → Output
Program

Connectionnism vs **Symbolic**

Facts ▶ Rules and laws

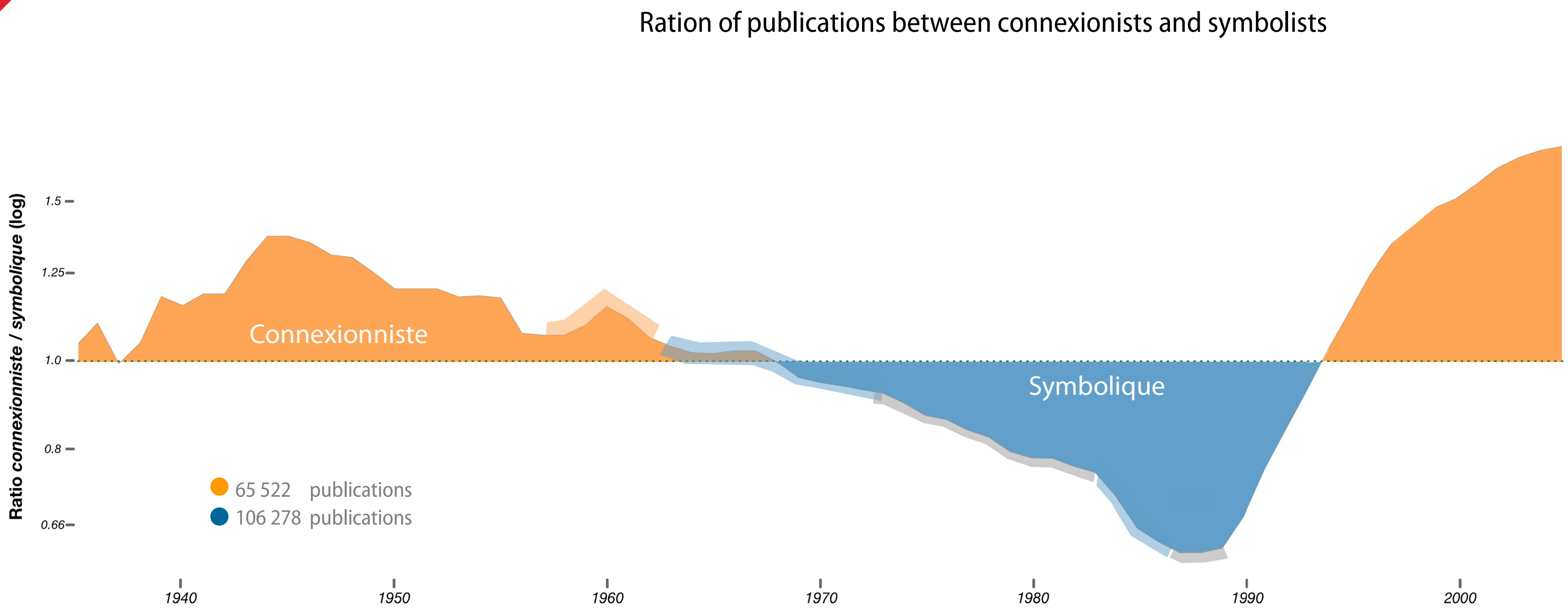
Model

Expert

Rules and laws ▶ Special case

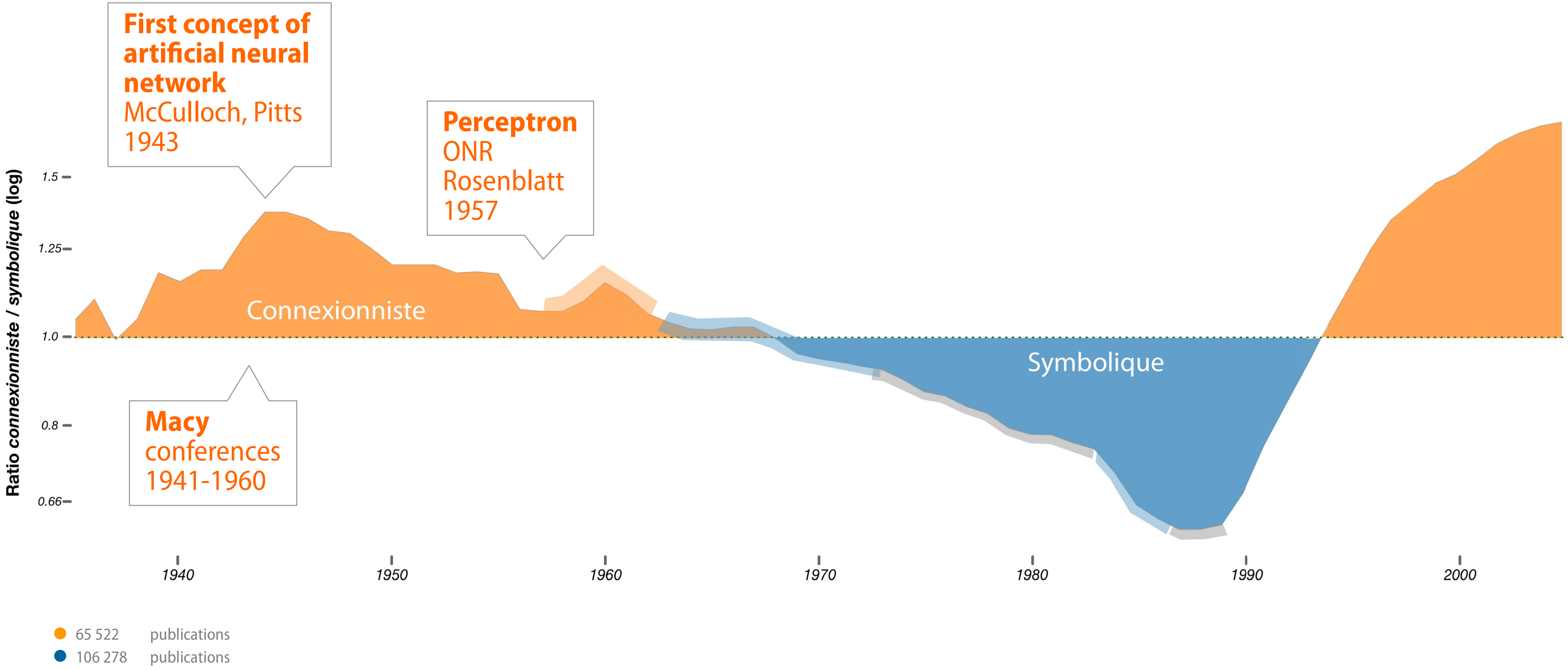
The big Controversy

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



[1] D Cardon, JP Cointet, A Mazieres (2018), La revanche des neurones https://dx.doi.org/10.3917/res.211.0173

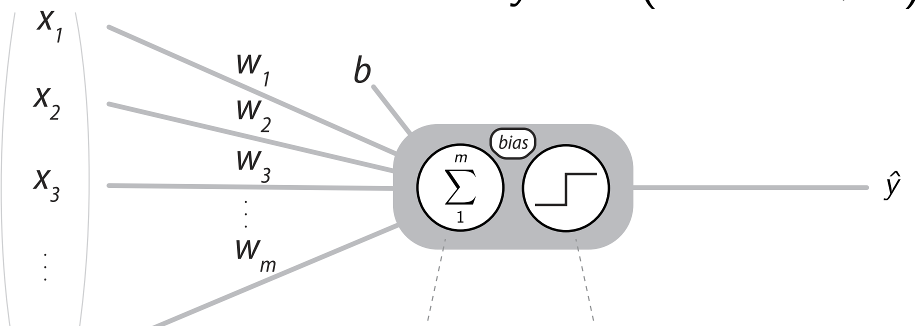
Evolution of the academic influence of connexionist and symbolic approaches[1]



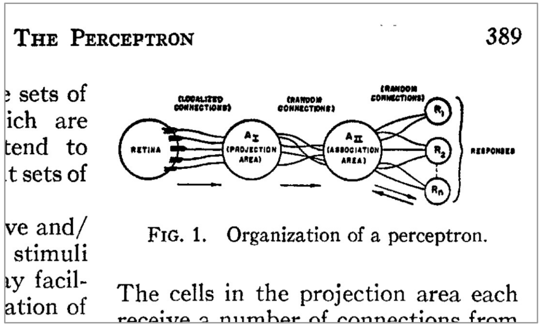
[1] D Cardon, JP Cointet, A Mazieres (2018), La revanche des neurones
https://dx.doi.org/10.3917/res.211.0173

Perceptron

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



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



THE PERCEPTRON 389

e sets of
ich are
tend to
t sets of

ve and/
stimuli
ly facil-
ation of

FIG. 1. Organization of a perceptron.

The cells in the projection area each

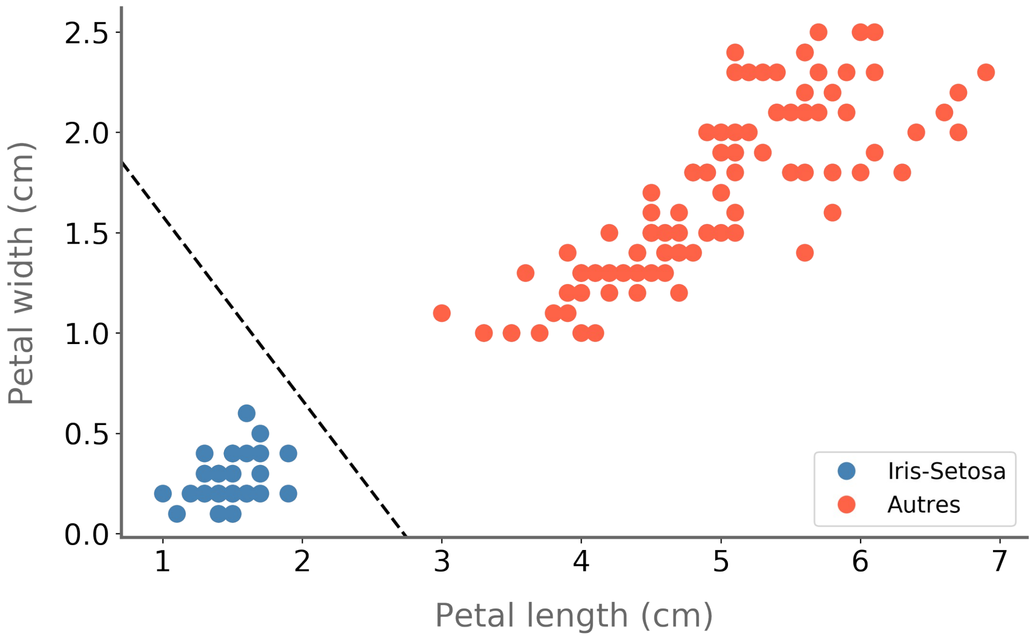
Perceptron
Frank Rosenblatt
1958

Linear and binary classifier

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. https://doi.org/10.1037/h0042519

Iris plants dataset

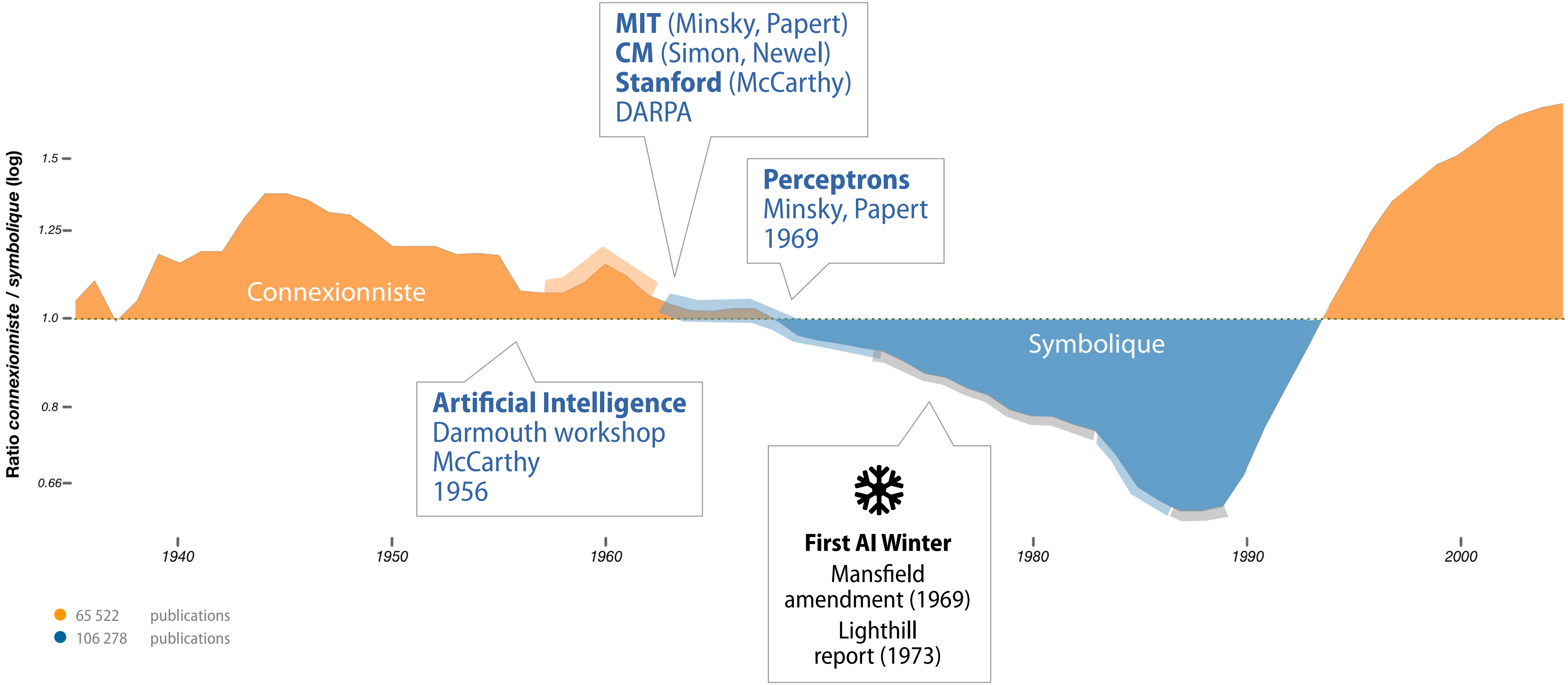
Dataset from : Fisher, R.A. "The use of multiple measurements in taxonomic problems" Annual Eugenics, 7, Part II, 179-188 (1936)



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
(...)		

The big Controversy

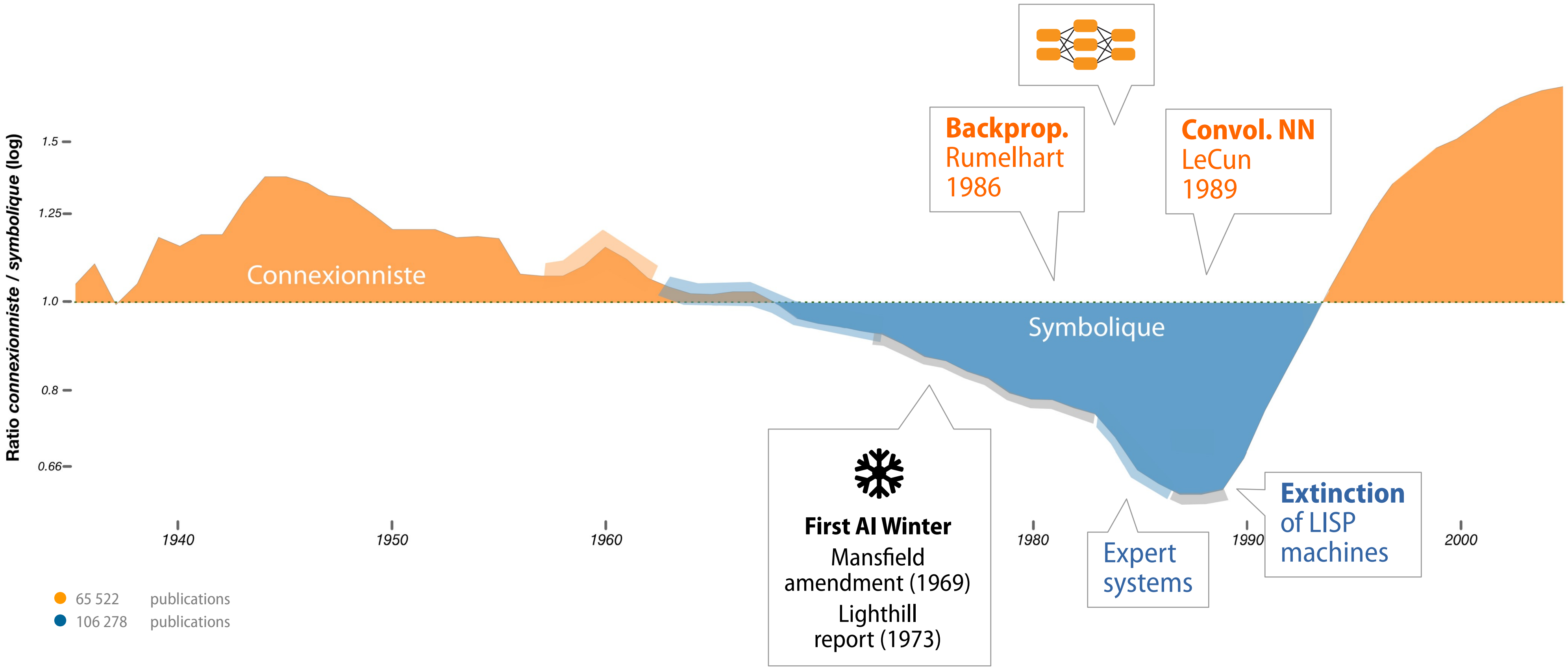
Evolution of the academic influence of connexionist and symbolic approaches[1]



[1] D Cardon, JP Cointet, A Mazieres (2018), La revanche des neurones
https://dx.doi.org/10.3917/res.211.0173

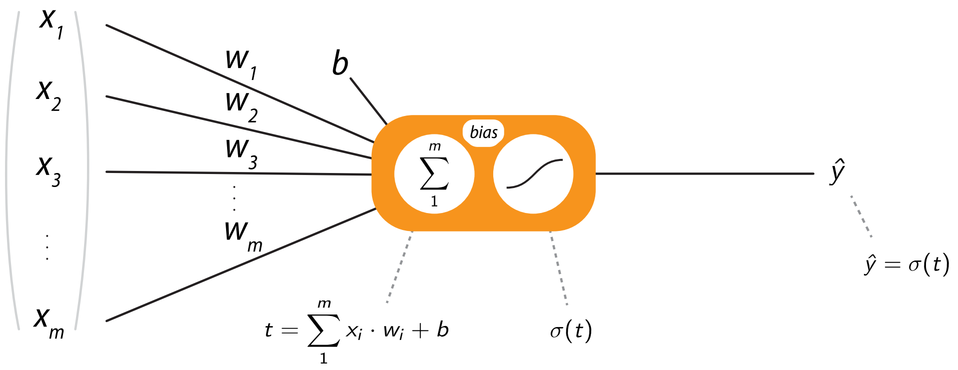
The big Controversy

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

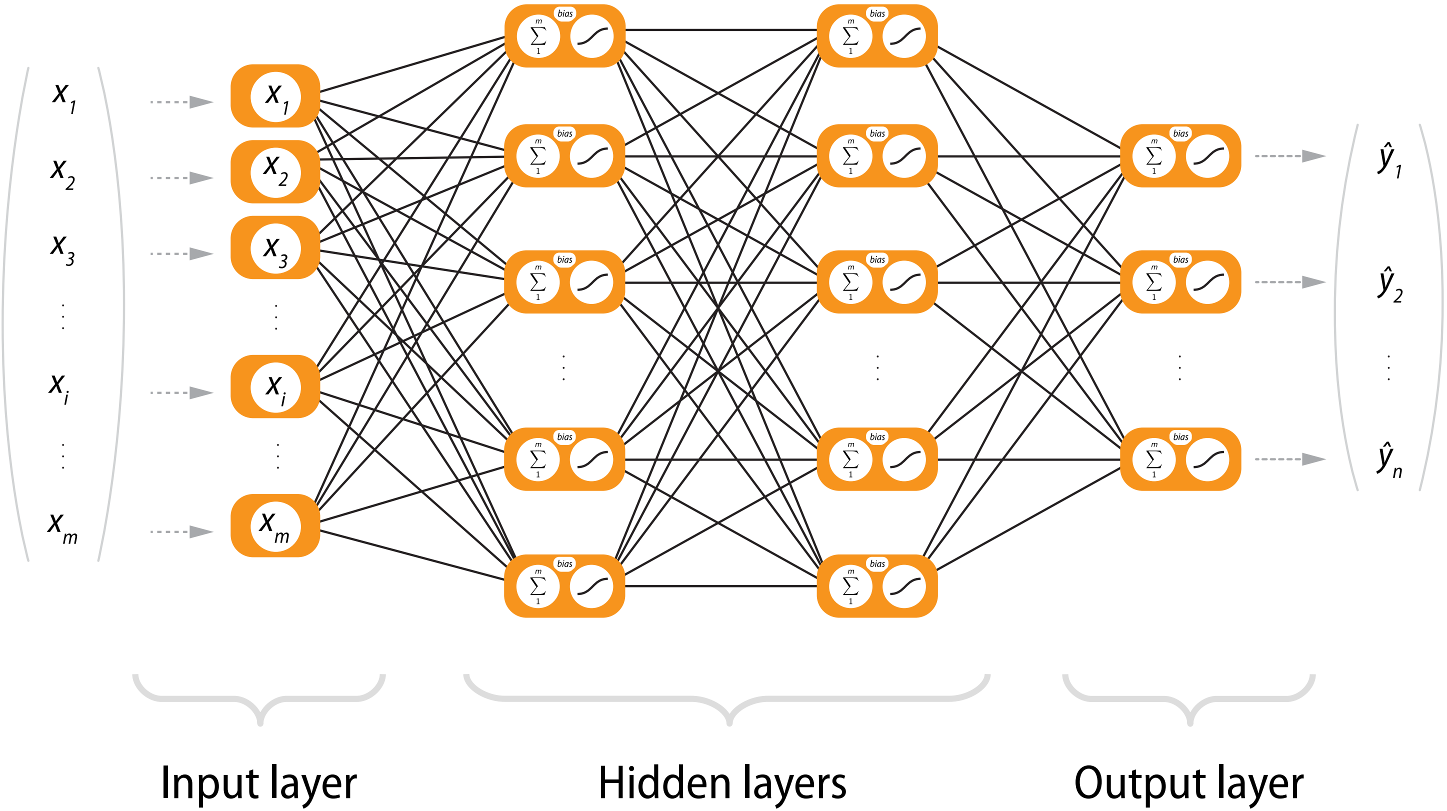


[1] D Cardon, JP Cointet, A Mazieres (2018), La revanche des neurones https://dx.doi.org/10.3917/res.211.0173

Deep Neural Networks



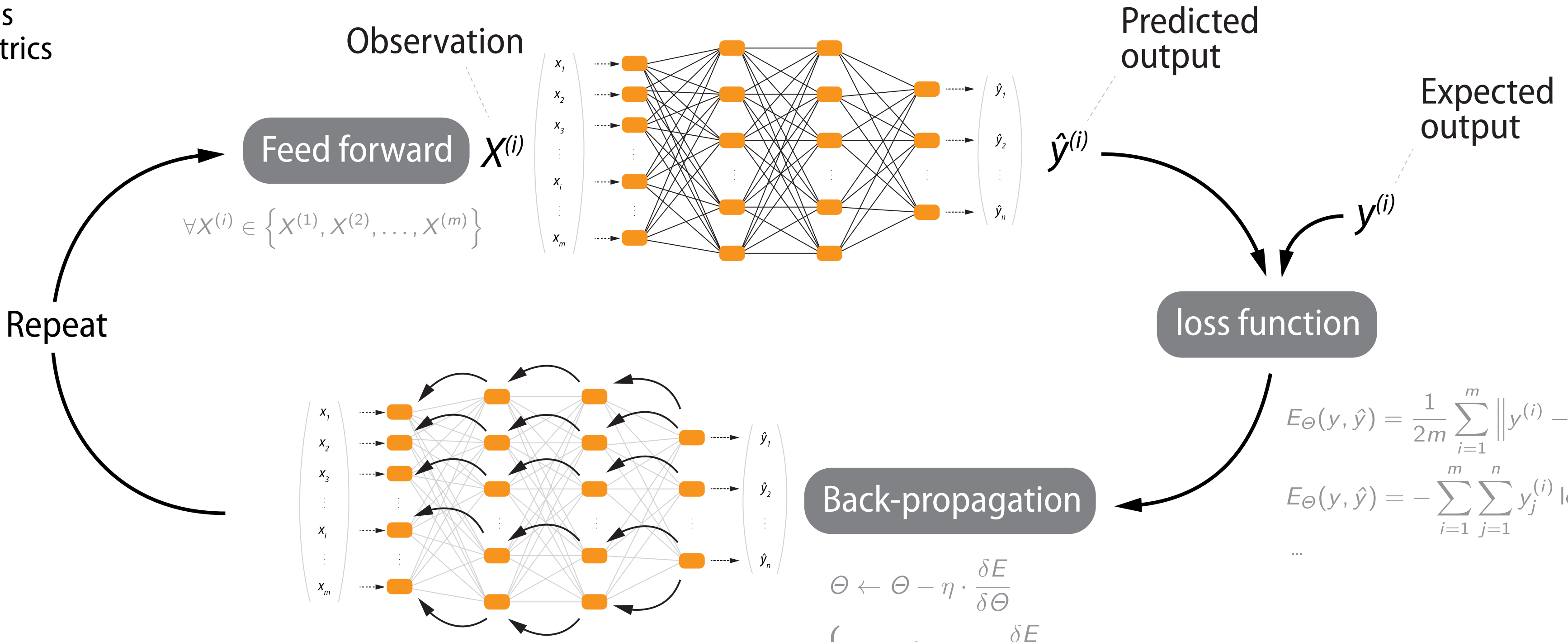
Deep Neural Networks



Deep Neural Networks

To define :

- Optimization
- Activation
- Loss
- Metrics
- ...



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

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

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

...

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

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

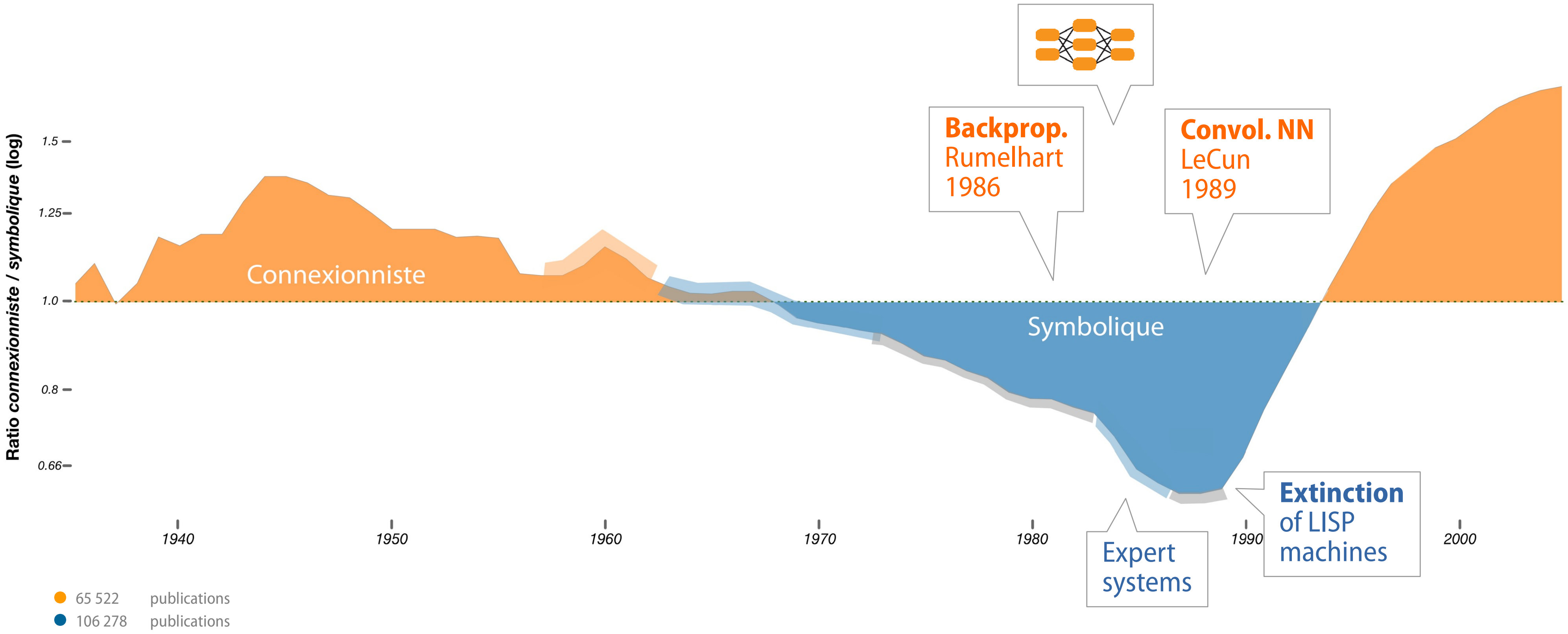
η : Learning rate
β : Momentum

Back-propagation

Learning process

The big Controversy

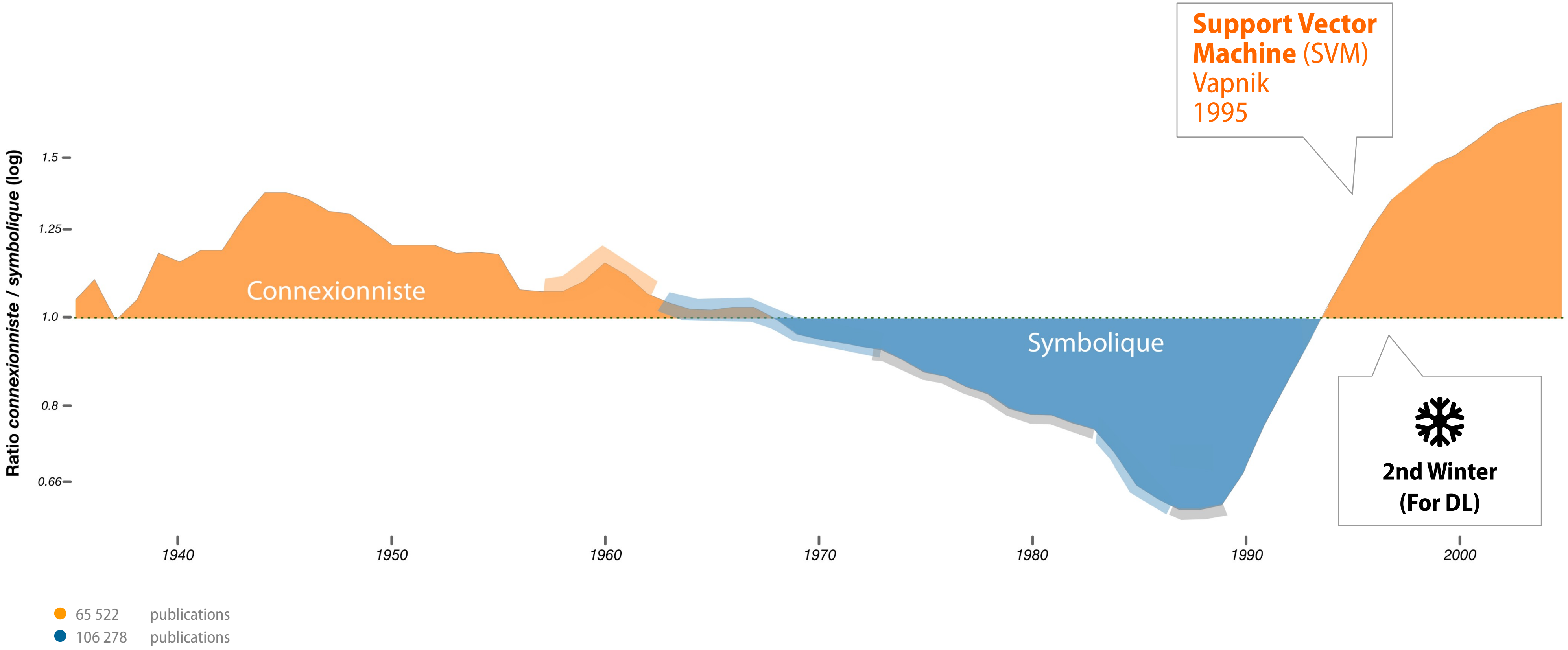
Evolution of the academic influence of connexionist and symbolic approaches[1]



[1] D Cardon, JP Cointet, A Mazieres (2018), La revanche des neurones
https://dx.doi.org/10.3917/res.211.0173

The big Controversy

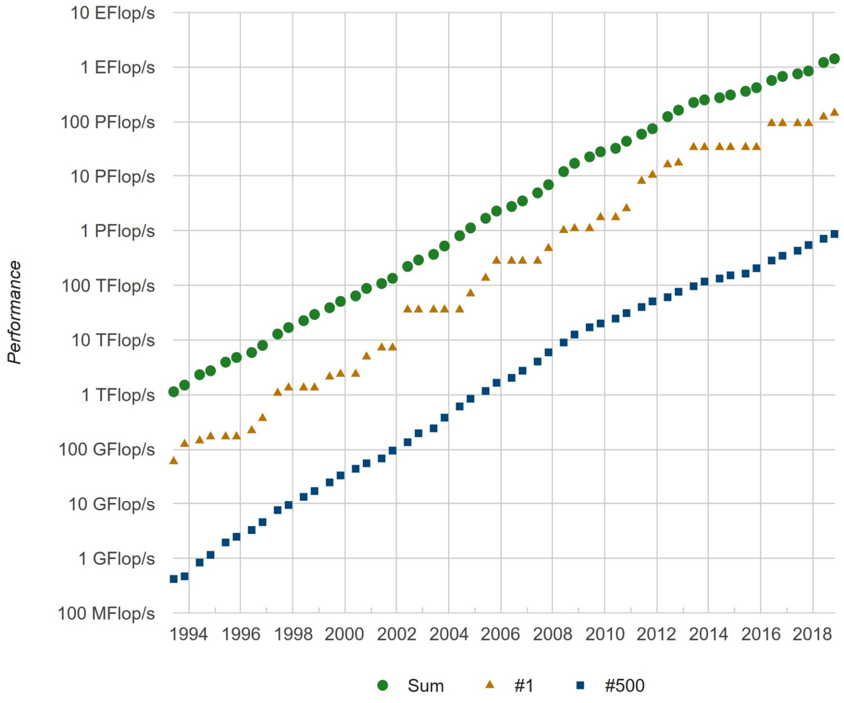
Evolution of the academic influence of connexionist and symbolic approaches[1]



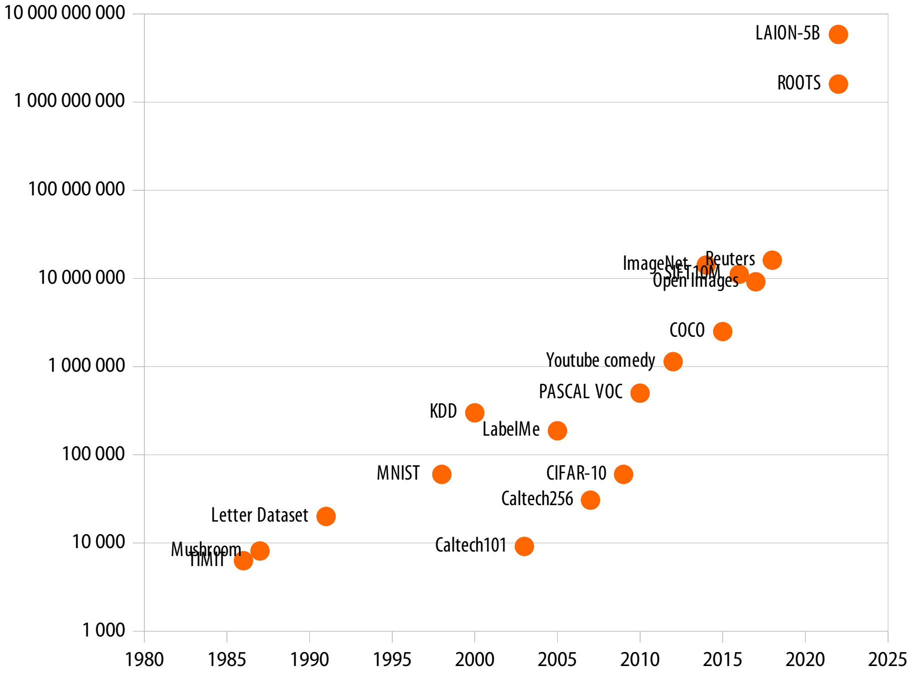
[1] D Cardon, JP Cointet, A Mazieres (2018), La revanche des neurones https://dx.doi.org/10.3917/res.211.0173

The big Controversy

Performance Development[1]



Datasets for machine-learning[2]

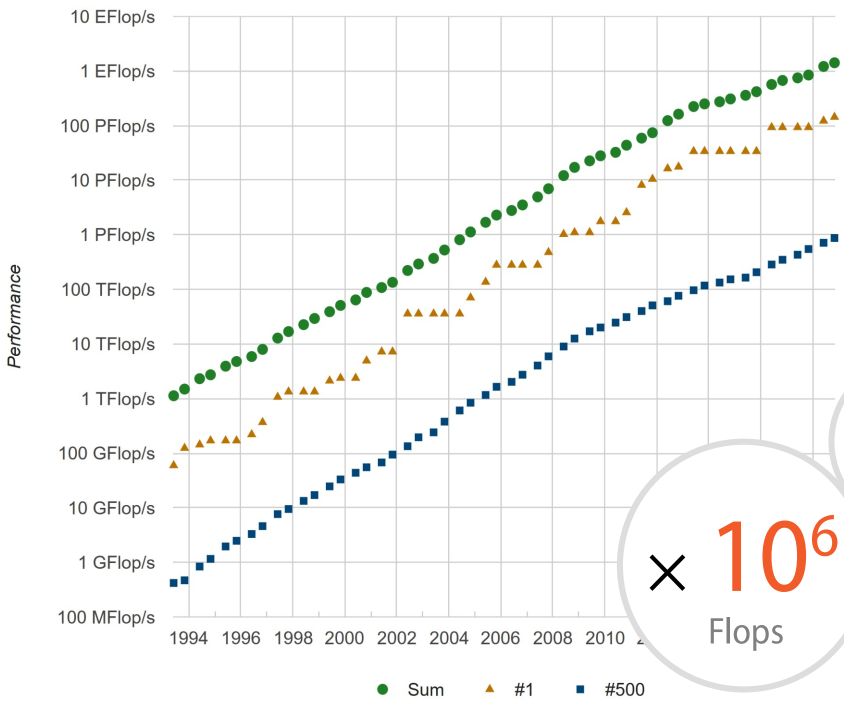


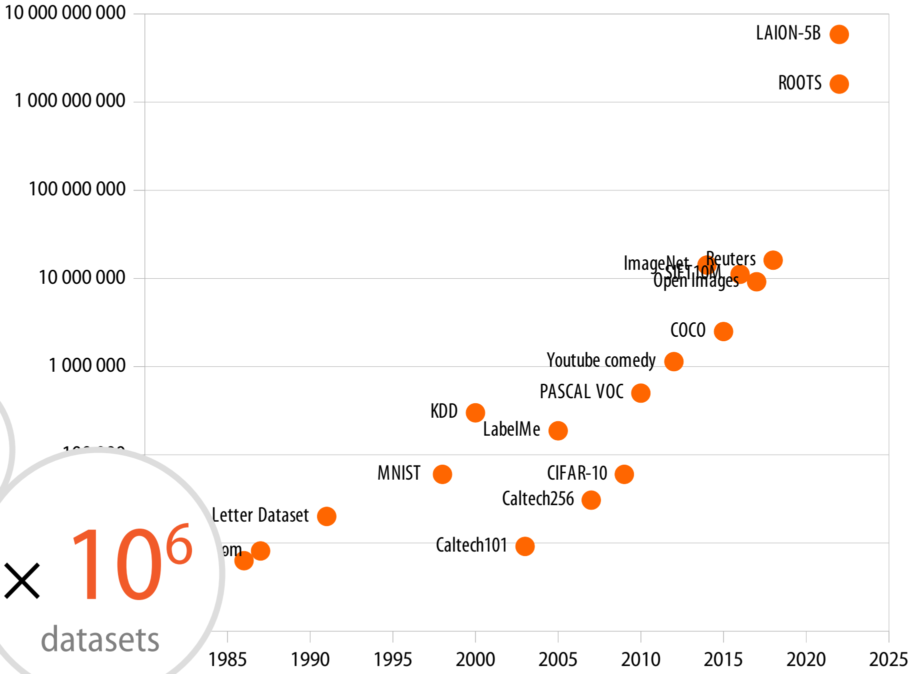
Laboratoire
Cas particulier
→
Monde réel

[1] TOP500 List - https://www.top500.org/

[2] Wikipedia

The big Controversy





× 10^6 Flops

25 ans

× 10^6 datasets

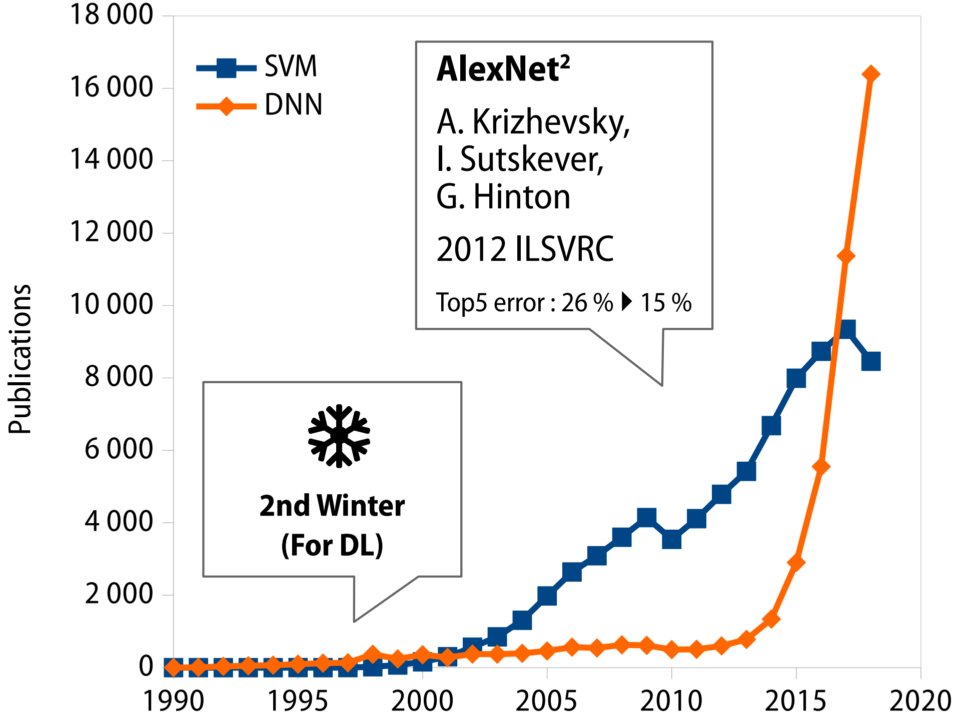
Laboratoire
Cas particulier → Monde réel

[1] TOP500 List - https://www.top500.org/

[2] https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research

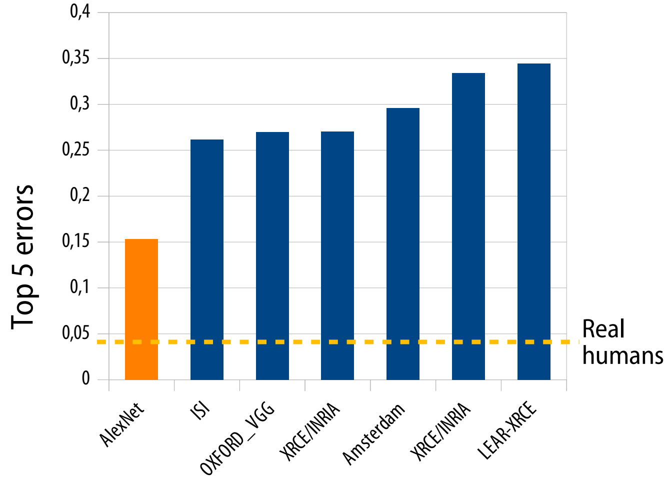
The big Controversy

Publications SVM vs DNN[1]



DNN

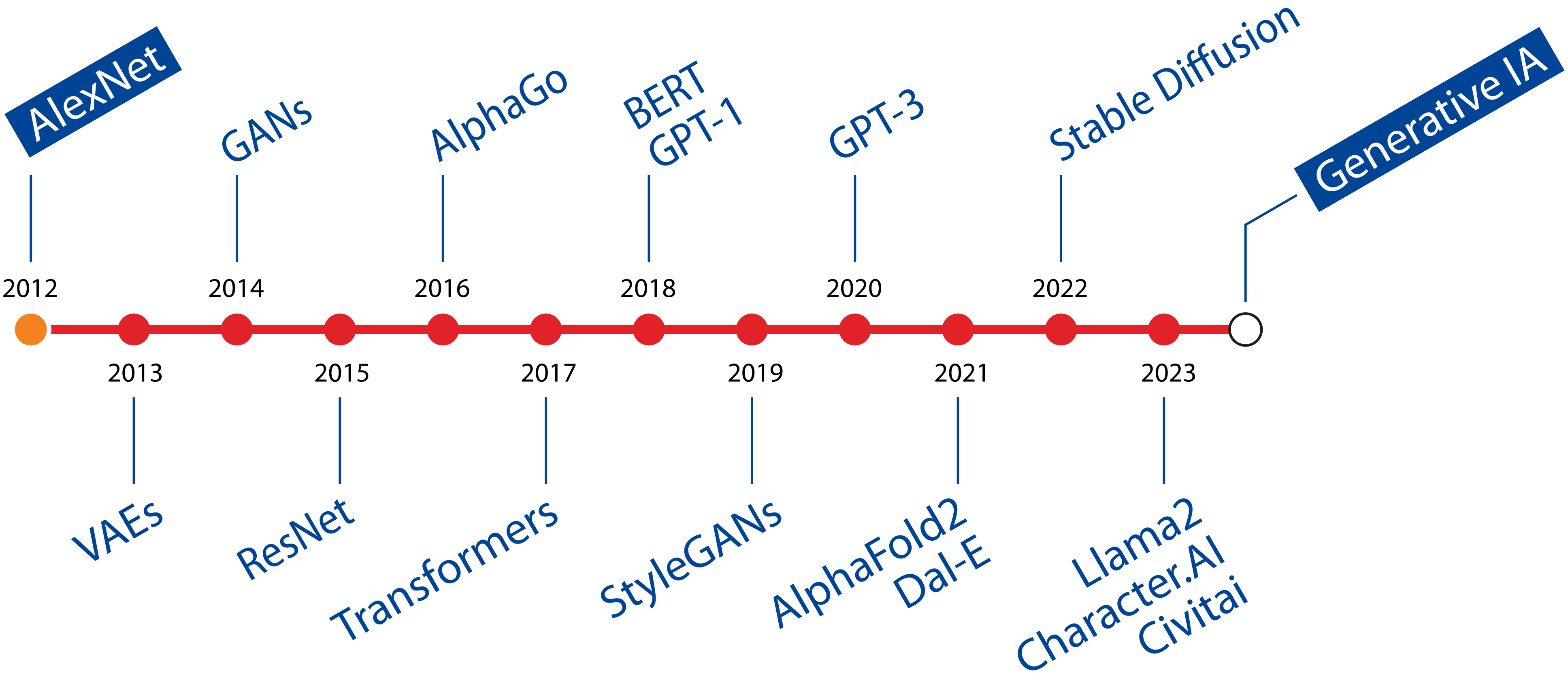
Images classification
Top 5 error at ILSVRC 2012[3,4]



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

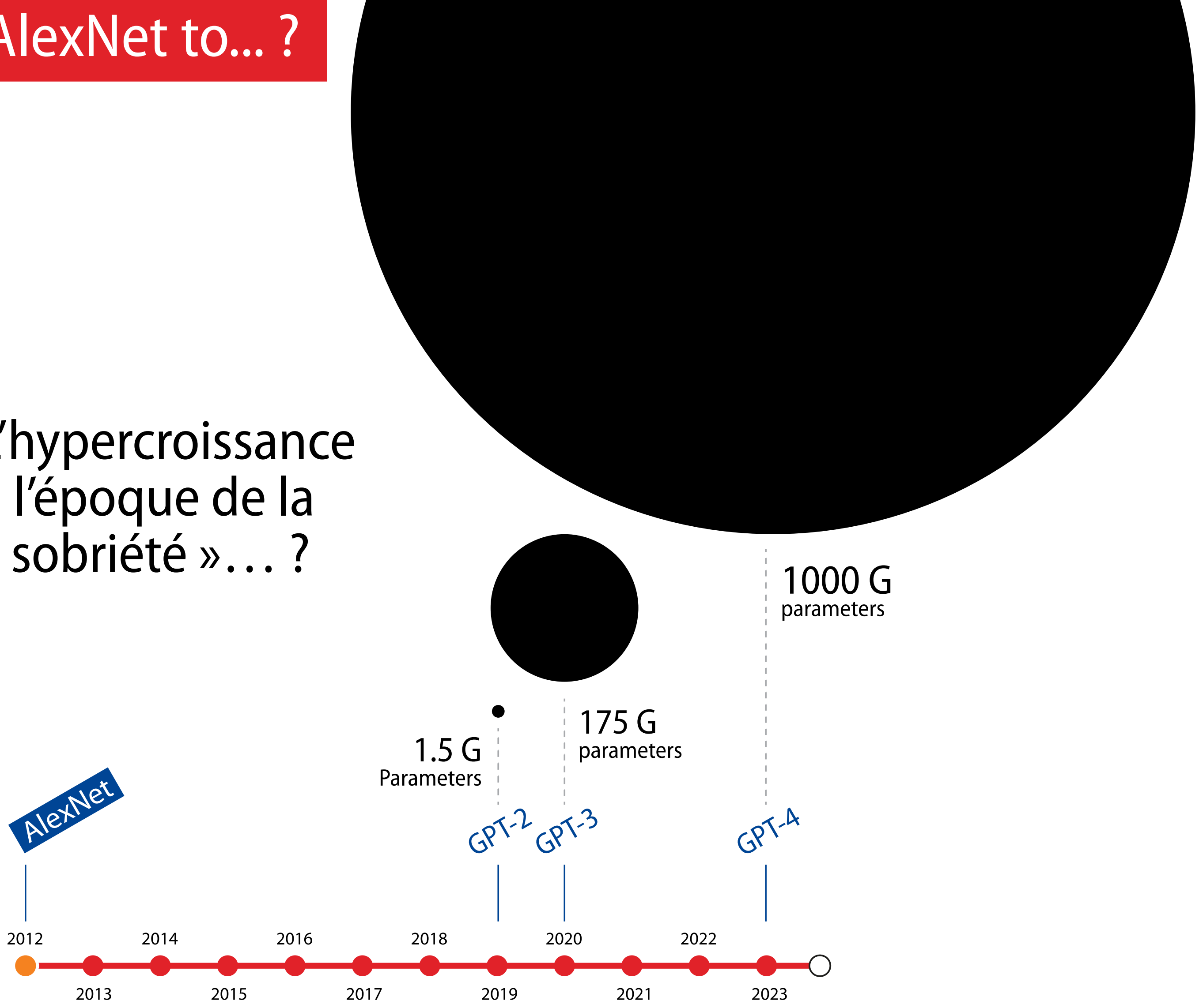
[1] Web of Science
[2] AlexNet at ILSVRC, https://image-net.org/challenges/LSVRC/2012/results.html
[3] ImageNet Large Scale Visual Recognition, https://www.image-net.org/challenges/LSVRC/
[4] Similar evolution in NLP, translation, board games, etc.
See : DeepL.com, AlphaGo, AlphaZero, ...

From AlexNet to... ?



From AlexNet to... ?

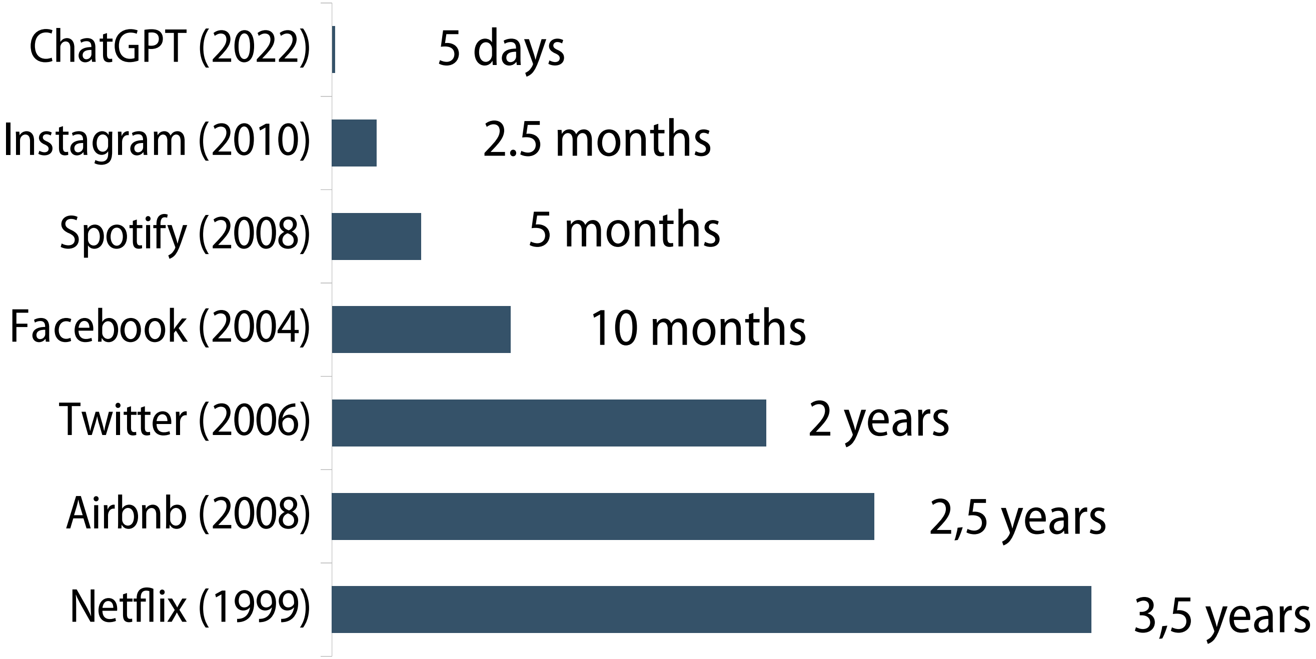
L'hypercroissance à l'époque de la « sobriété »… ?



From AlexNet to... ?

An unprecedented social appropriation.

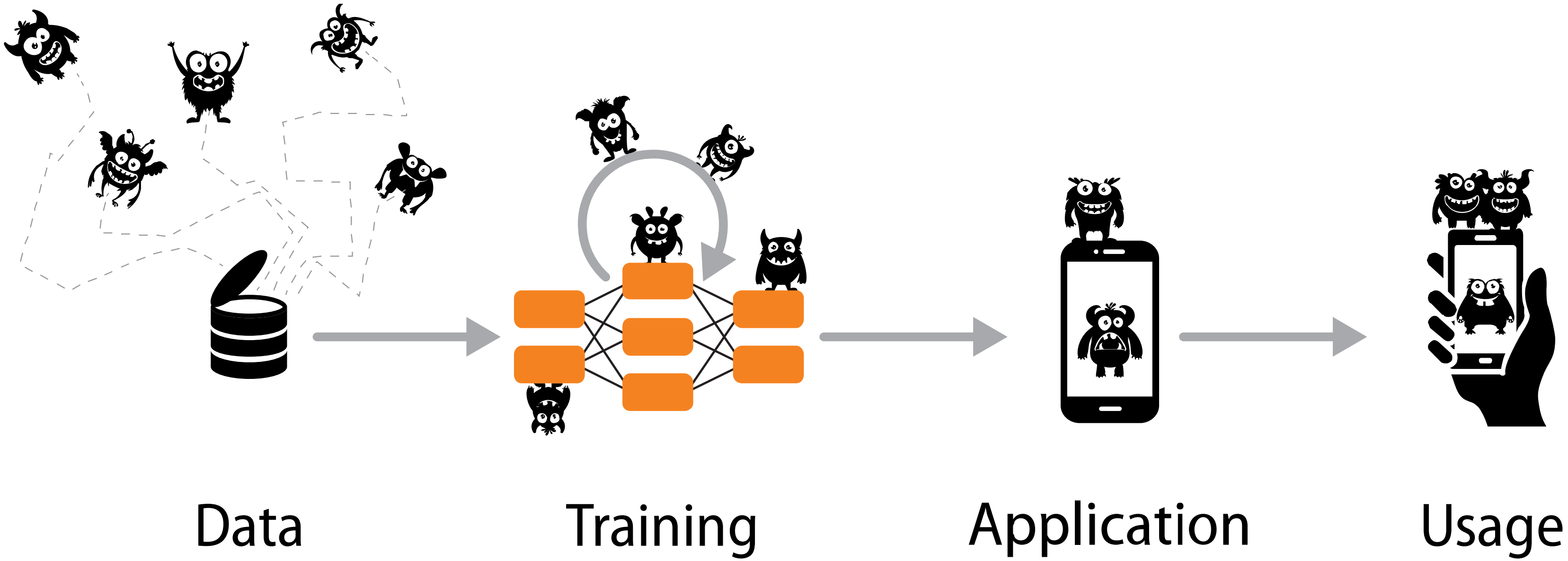
Time by service to reach 1 million users



https://www.statista.com/chart/29174/time-to-one-million-users/

From AlexNet to... ?

There are biases in my data...
...and elsewhere too!



From AlexNet to... ?

But also :

Explainability ?

Self-Consuming Generative Models Go MAD* ?

Transposition into the real world ?

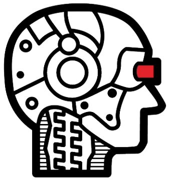
Social impact ?

Ability to keep up ?

9,000 publications per month....

(*) Model Autophagy Disorder (MAD) : https://doi.org/10.48550/arXiv.2307.01850

1



History and **Fundamental Concepts**

1 **Deep Learning**, what are we talking about?

2 From linear regression to **artificial neurons**

3 80 years of history ...and the victory of **neurons**

4 Concrete **illustration** :
- Basic Regression
- Basic Classification

Illustration #1

Predicting data : Regression

Using a Dense Neural Network (DNN)

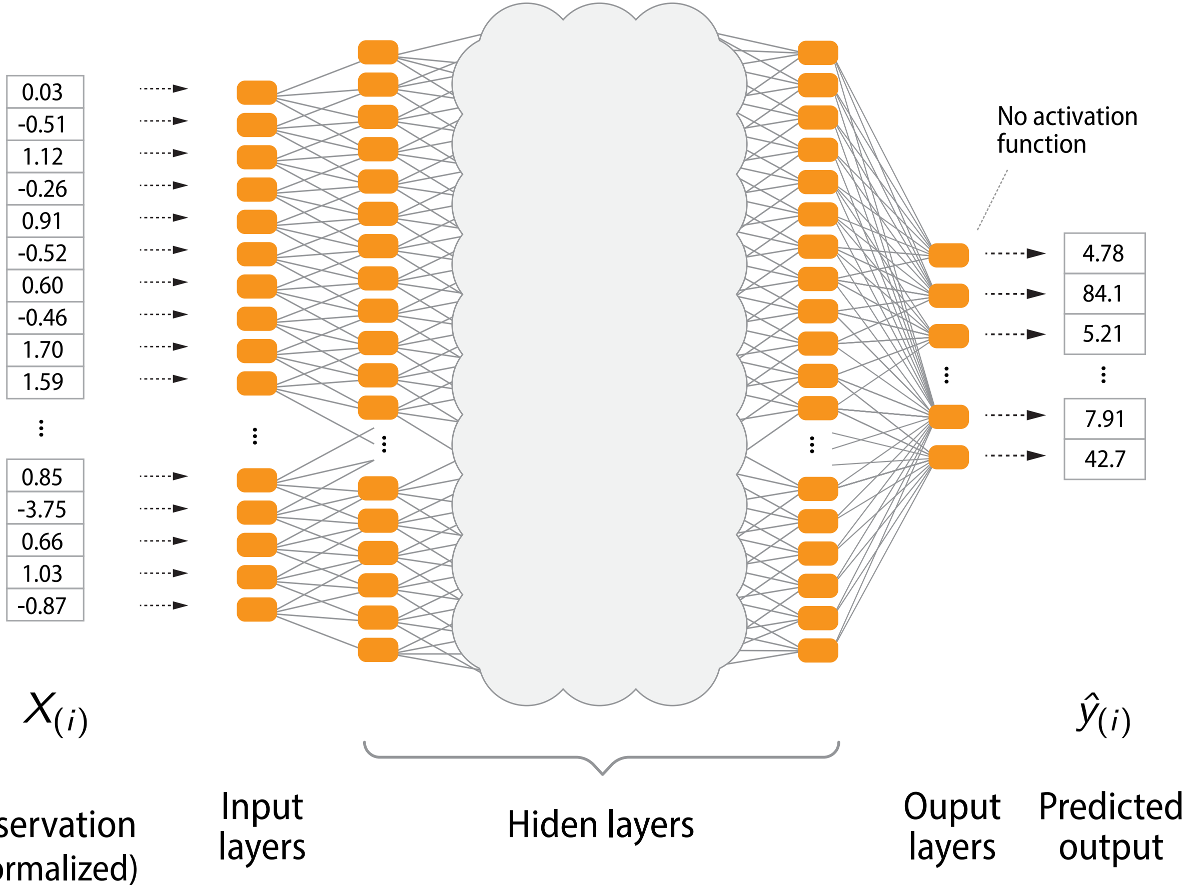
FIDLE

Regression with a DNN

Train set $\begin{cases} X_{(i)} \\ y_{(i)} \end{cases}$

$X_{(i)}$: Observations

$y_{(i)}$: Expected output

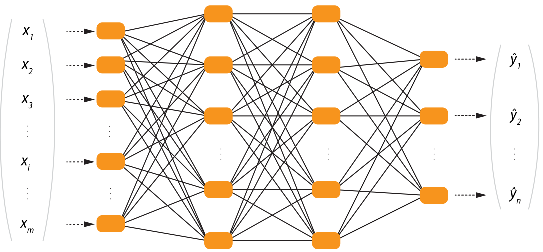


Training process - general

Observation

Feed forward $X^{(i)}$

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



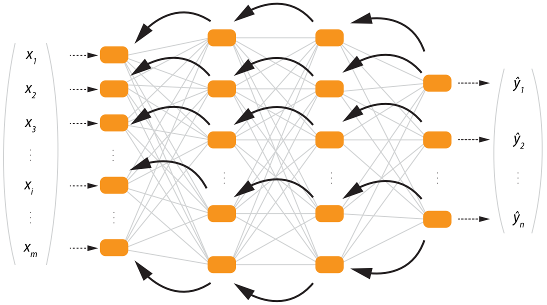
Predicted output $\hat{y}^{(i)}$

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

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

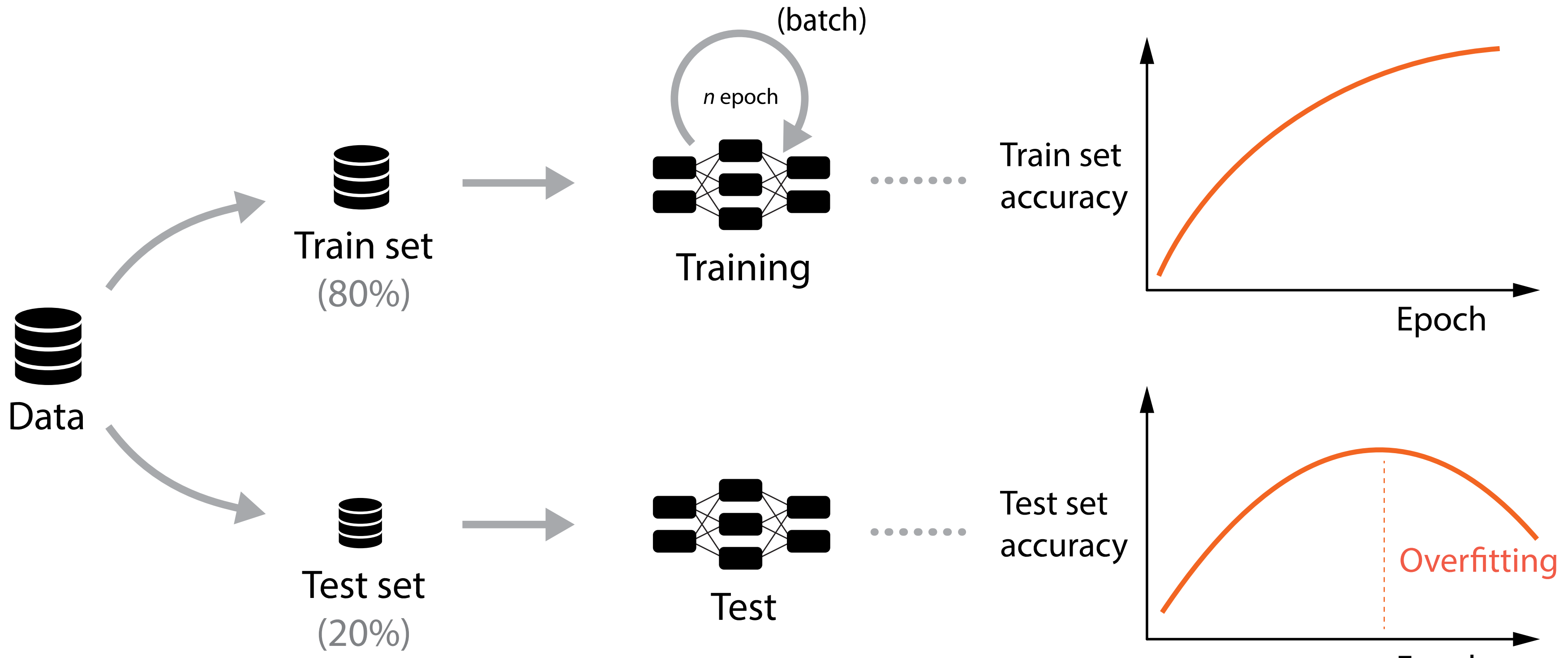
$$\begin{cases} m \leftarrow \beta m - \eta \cdot \dfrac{\delta E}{\delta \Theta} \\ \Theta \leftarrow \Theta + m \end{cases}$$

...

η : Learning rate
β : Momentum

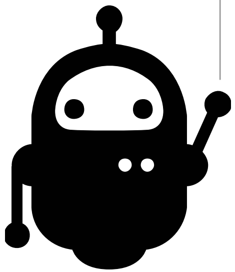
To define :
- Optimization
- Activation
- Loss
- Metrics
- ...

Training process - general



Regression with a Dense Neural 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)

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 :

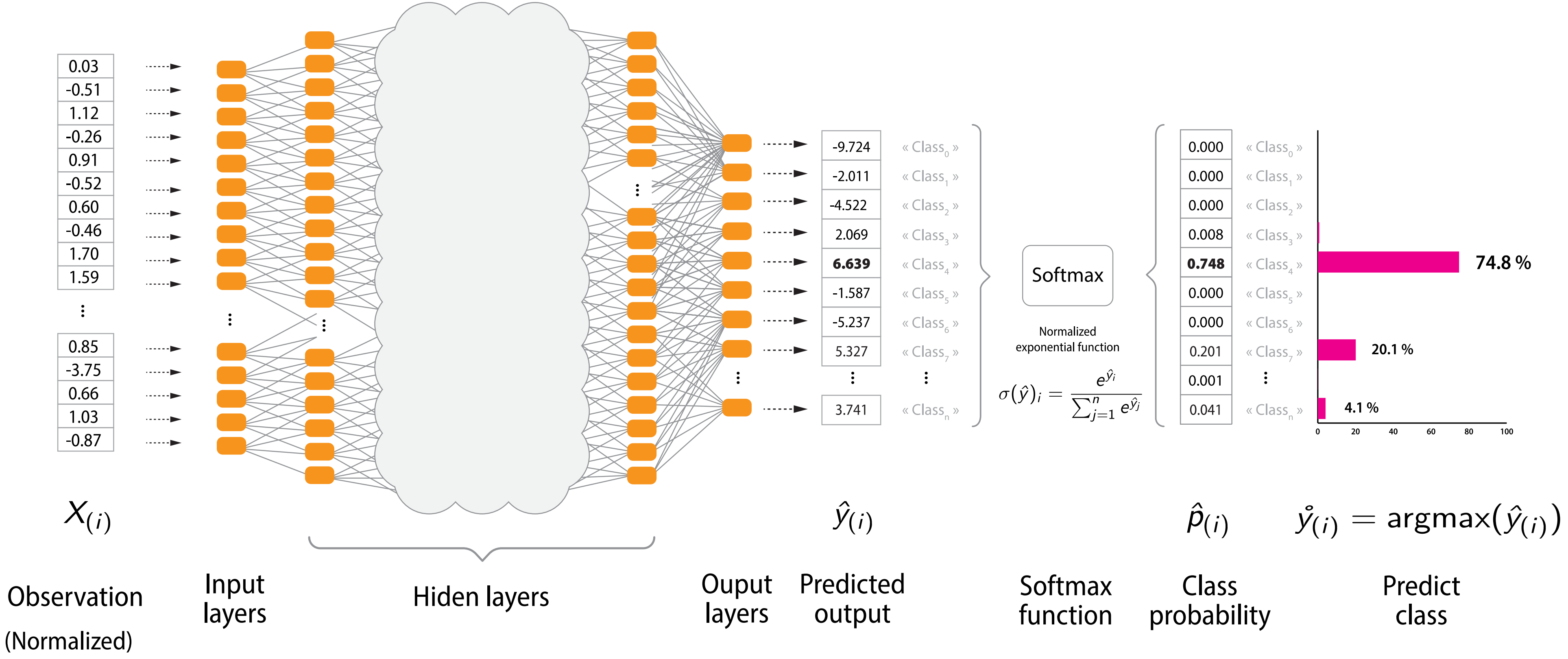
Make a first regression via a DNN network

Illustration #2

Data classification : Classification

Using a Dense Neural Network (DNN)

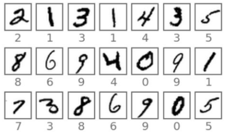
Classification with a DNN



Simple classification with DNN

Notebook : [MNIST1]

Objective :
Recognizing handwritten numbers



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

MNIST1

50s
(5 epochs)

Test set
X_{test}
y_{test}
Train set
X_{train}
y_{train}

START

Step 1 — Import and init

Step 2 — Load dataset

dataset

Step 3 — Preparing the data

Step 4 — Create model

Step 5 — Train the model

Step 6 — Evaluate

END

Objectives :

Make a first classification via a DNN network

FIDLE

https://youtube.com/@CNRS-FIDLE
https://fidle.cnrs.fr

Jeudi
24
Novembre
à 14h00

Next, on Fidle :

Bases, concepts et enjeux,

2

Data, models and representation's hell

Data and models

Réalisation, technique et mise en scène :
Agathe, Baptiste et Yanis - UGA/DAPI
Thibaut, Kamel - IDRIS