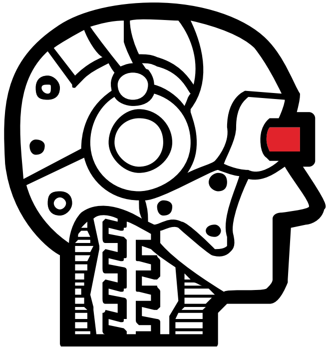
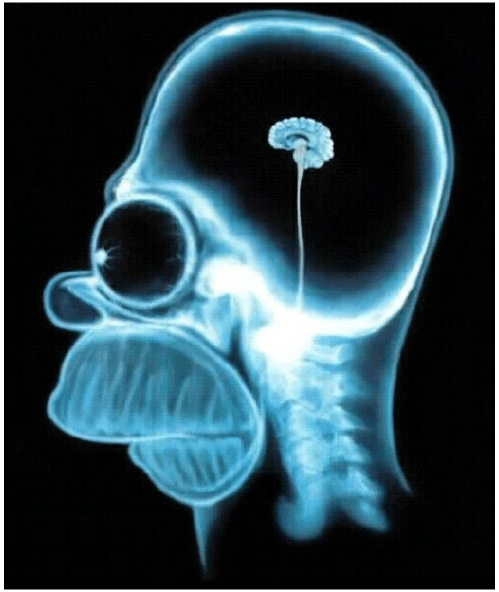


De la **controverse** au **succès**, comment les **neurones** ont changés la donne

Jean-Luc.Parouty@simap.grenoble-inp.fr

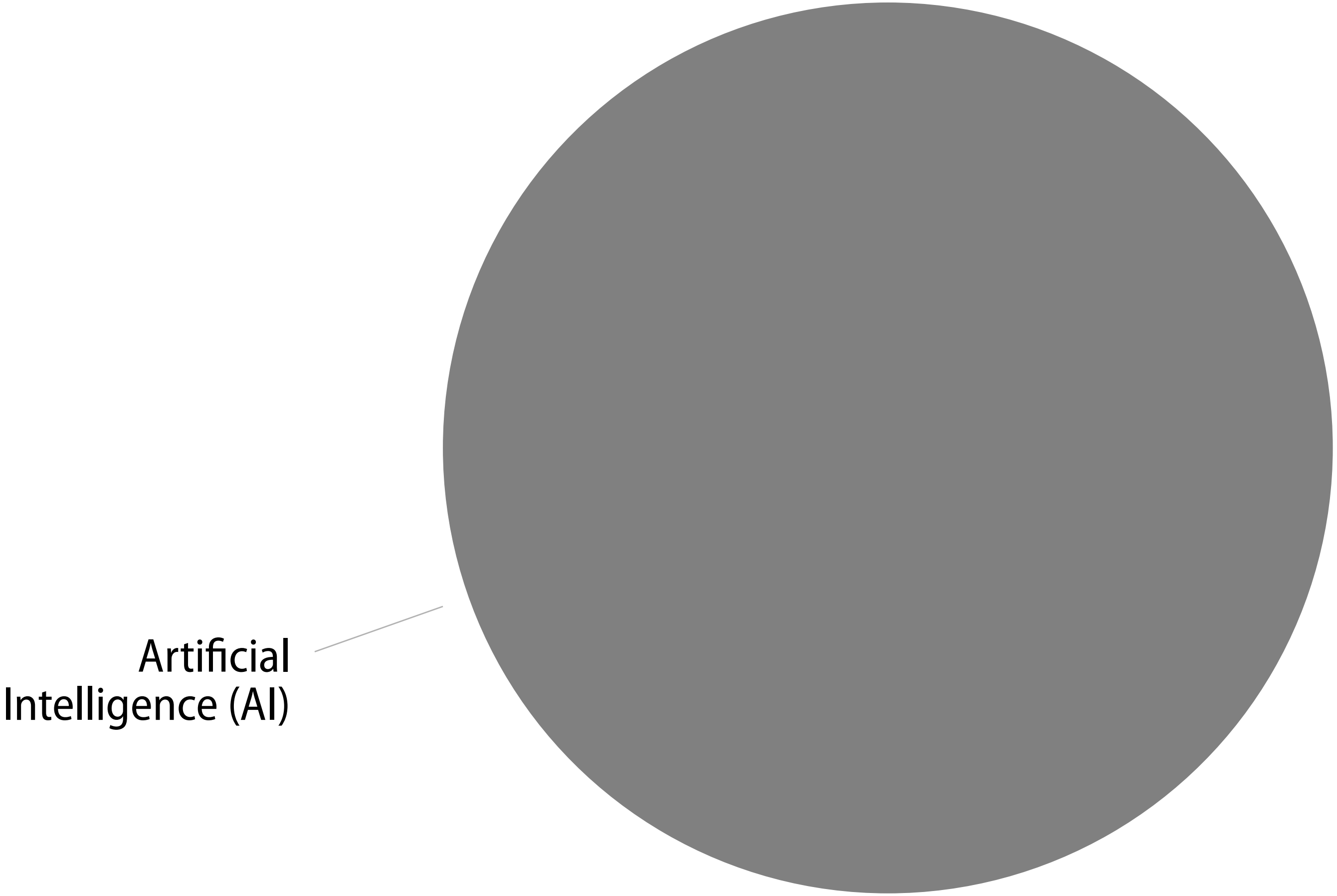
Journées prospectives
RFCT

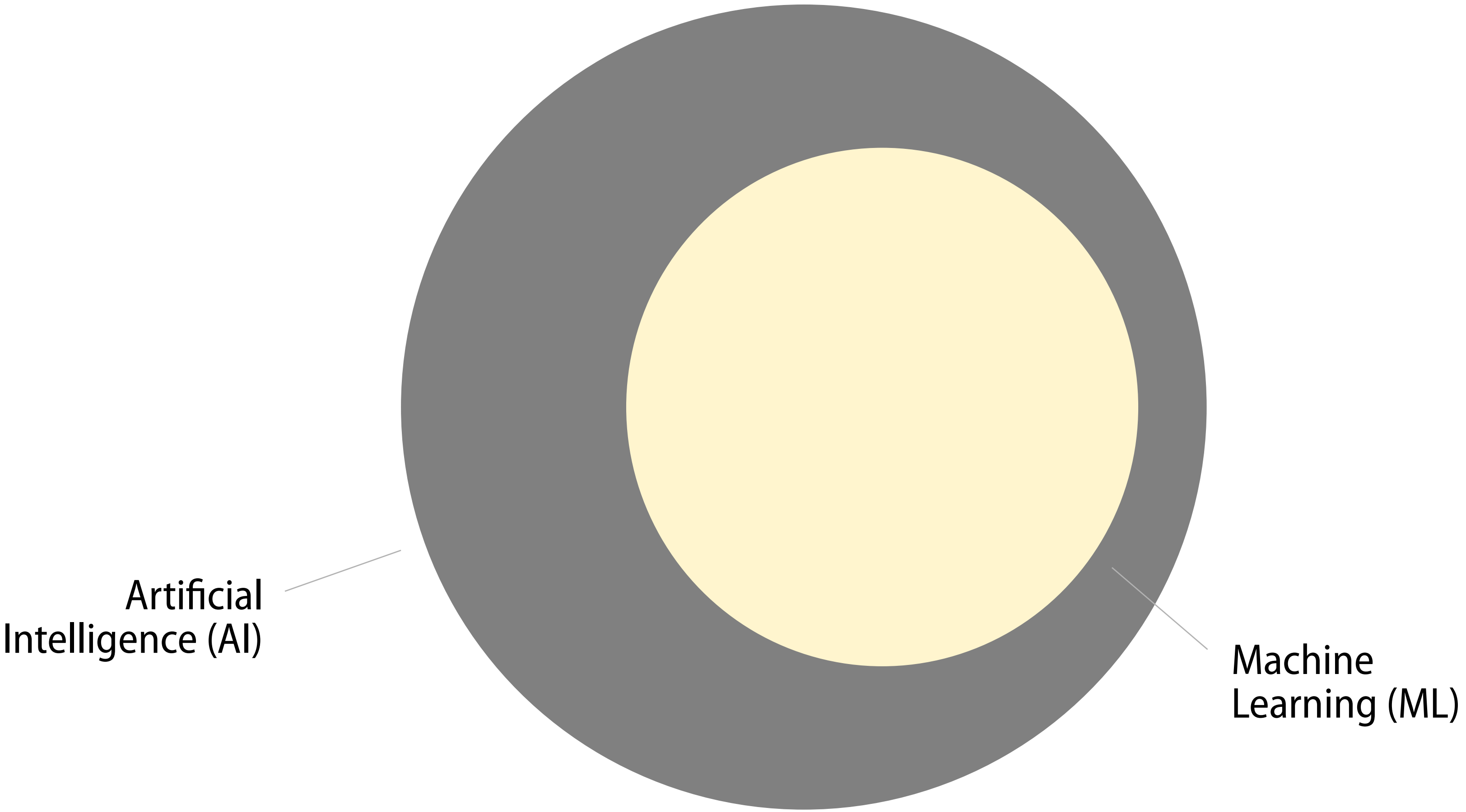


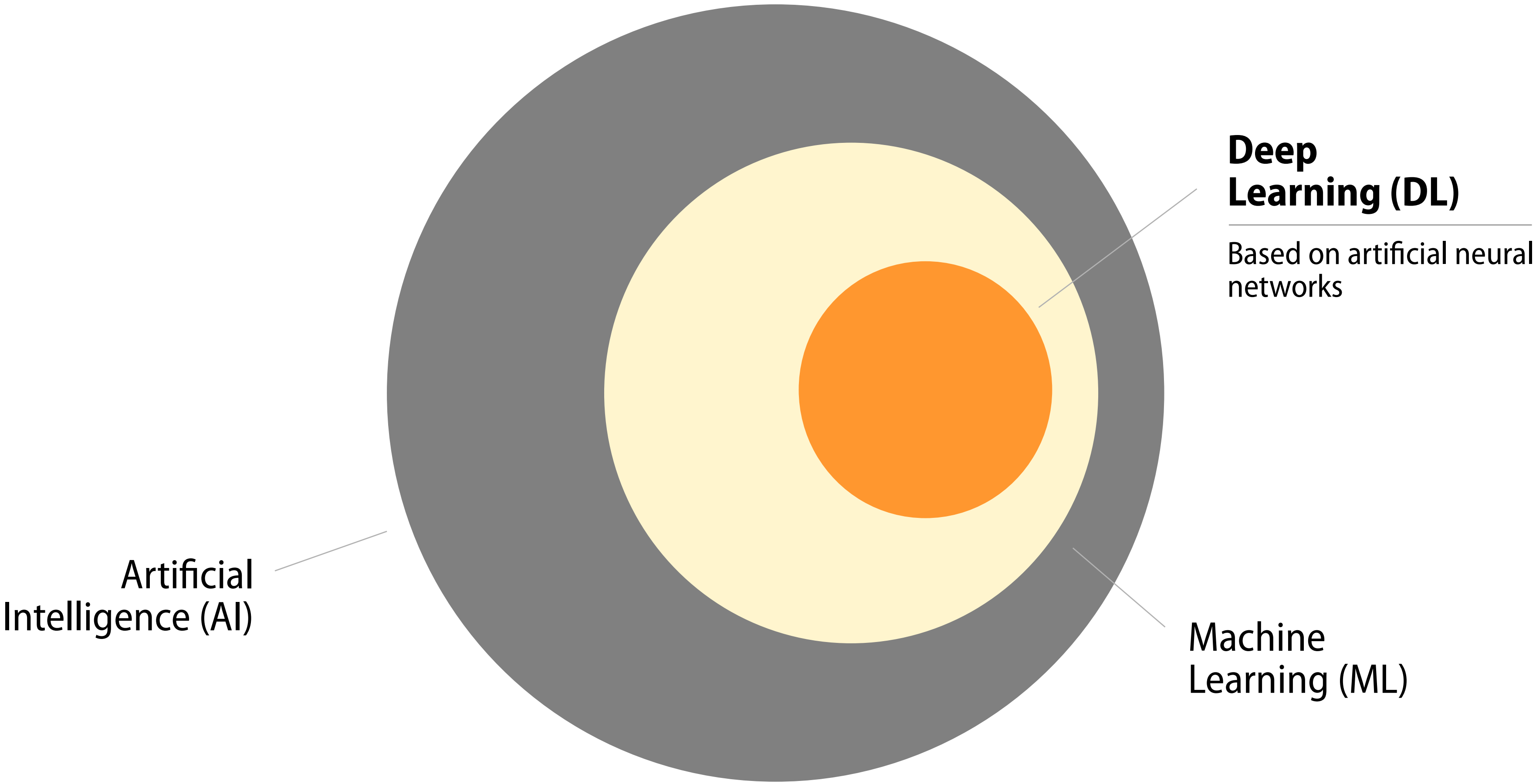
Why would we want to use artificial intelligence, when natural intelligence work so well?

What is an artificial neuron ?

How did they arrive ?

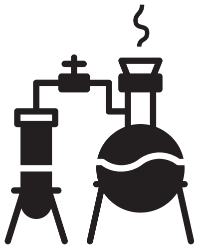

Artificial
Intelligence (AI)


Artificial
Intelligence (AI)
Machine
Learning (ML)


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

Scientific paradigms

1st paradigm



Experimental science

2nd paradigm

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

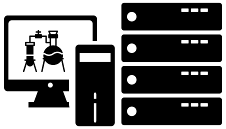
$$\nabla \times H = J + \frac{\partial D}{\partial t}$$

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

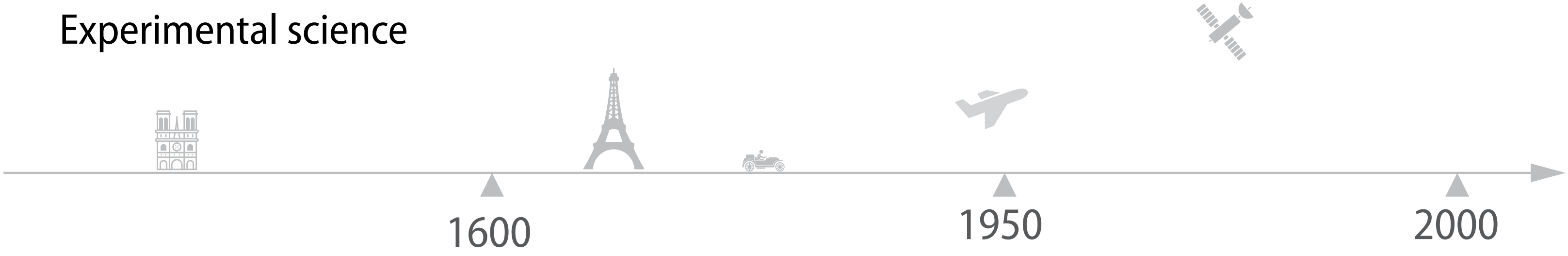
Theoretical science

3rd paradigm

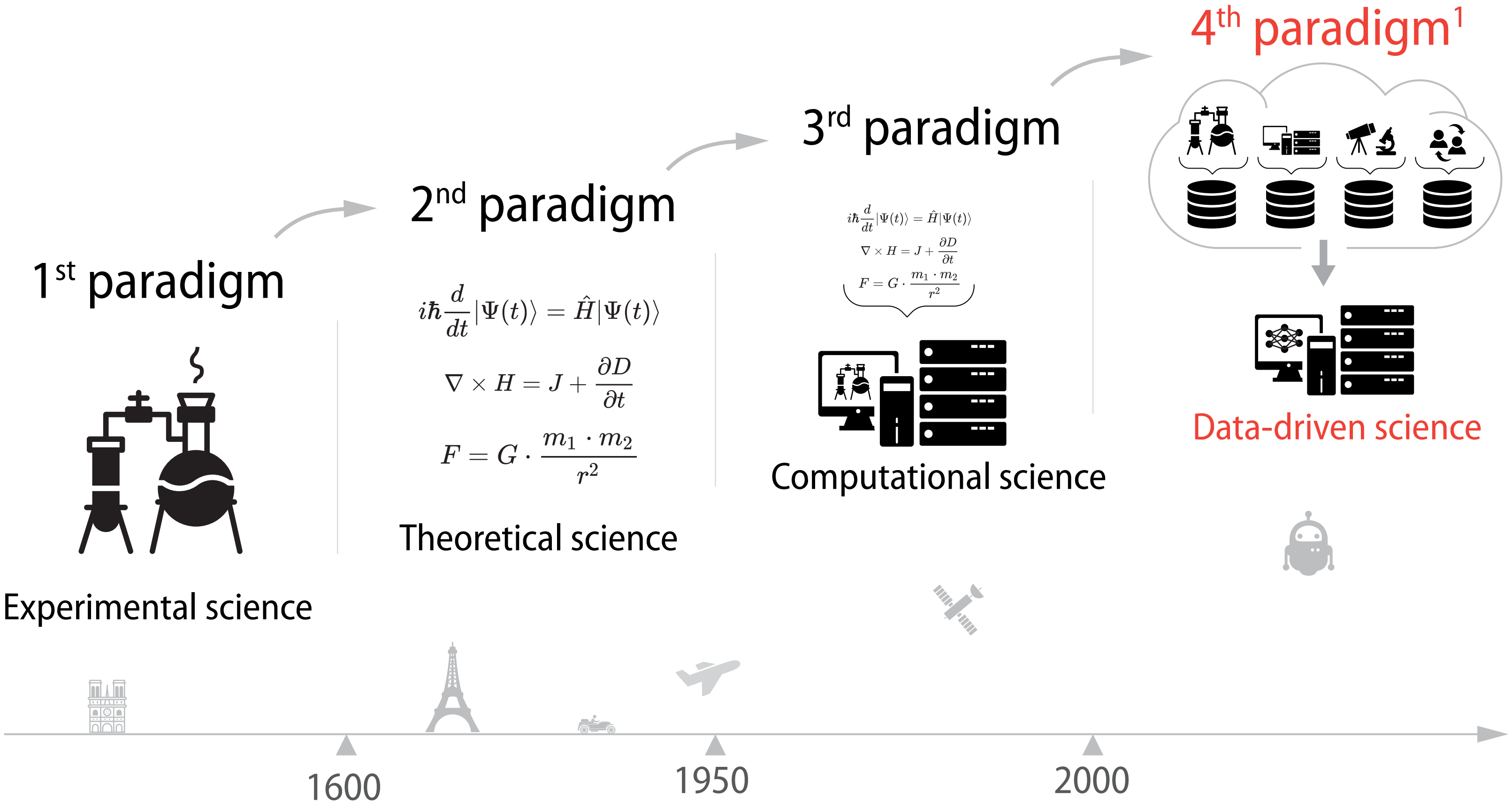
$$\underbrace{\begin{gathered} i\hbar\frac{d}{dt}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle \\ \nabla \times H = J + \frac{\partial D}{\partial t} \\ F = G \cdot \frac{m_1 \cdot m_2}{r^2} \end{gathered}}$$



Computational science



Scientific paradigms



[1] Jim Gray, 2007 [GRAY]

Roadmap

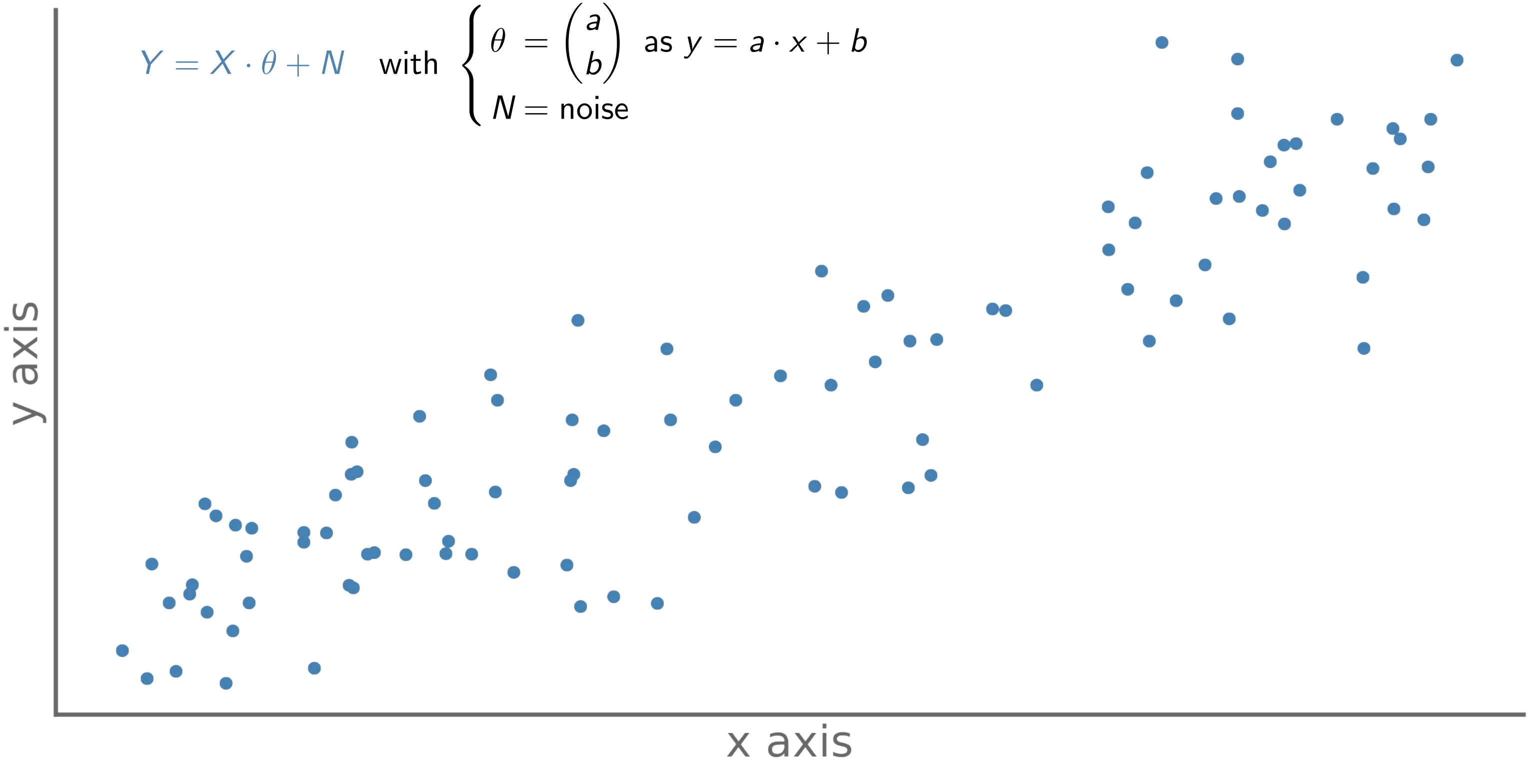
1. From the linear regression to the first neuron
2. Neurons in controversy
3. Data and neurons

From the linear regression to the first neuron

Linear regression

We have a phenomenon, for which we have observations

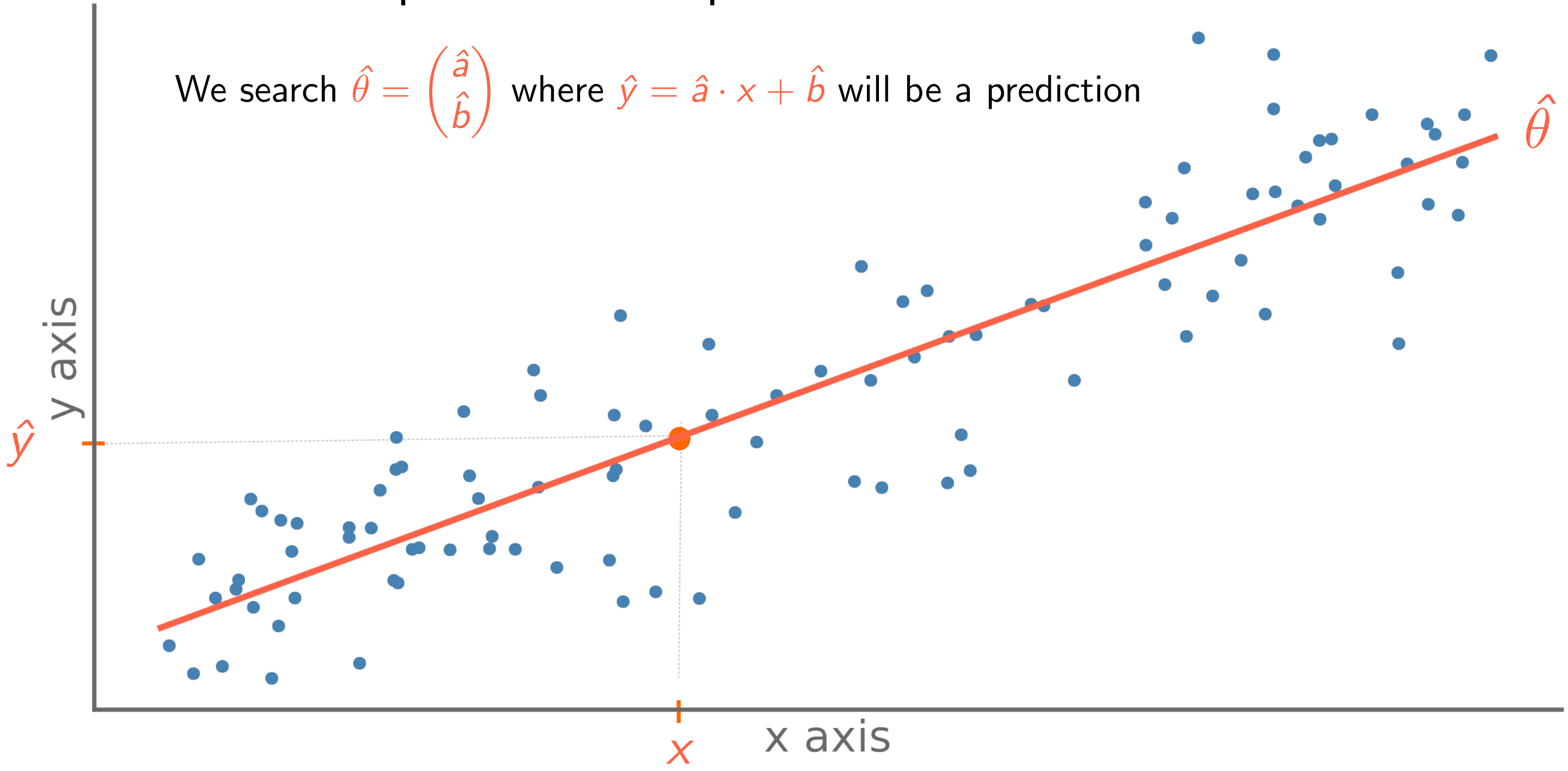
$$Y = X \cdot \theta + N \quad \text{with} \quad \begin{cases} \theta = \begin{pmatrix} a \\ b \end{pmatrix} & \text{as } y = a \cdot x + b \\ N = \text{noise} \end{cases}$$



Linear regression

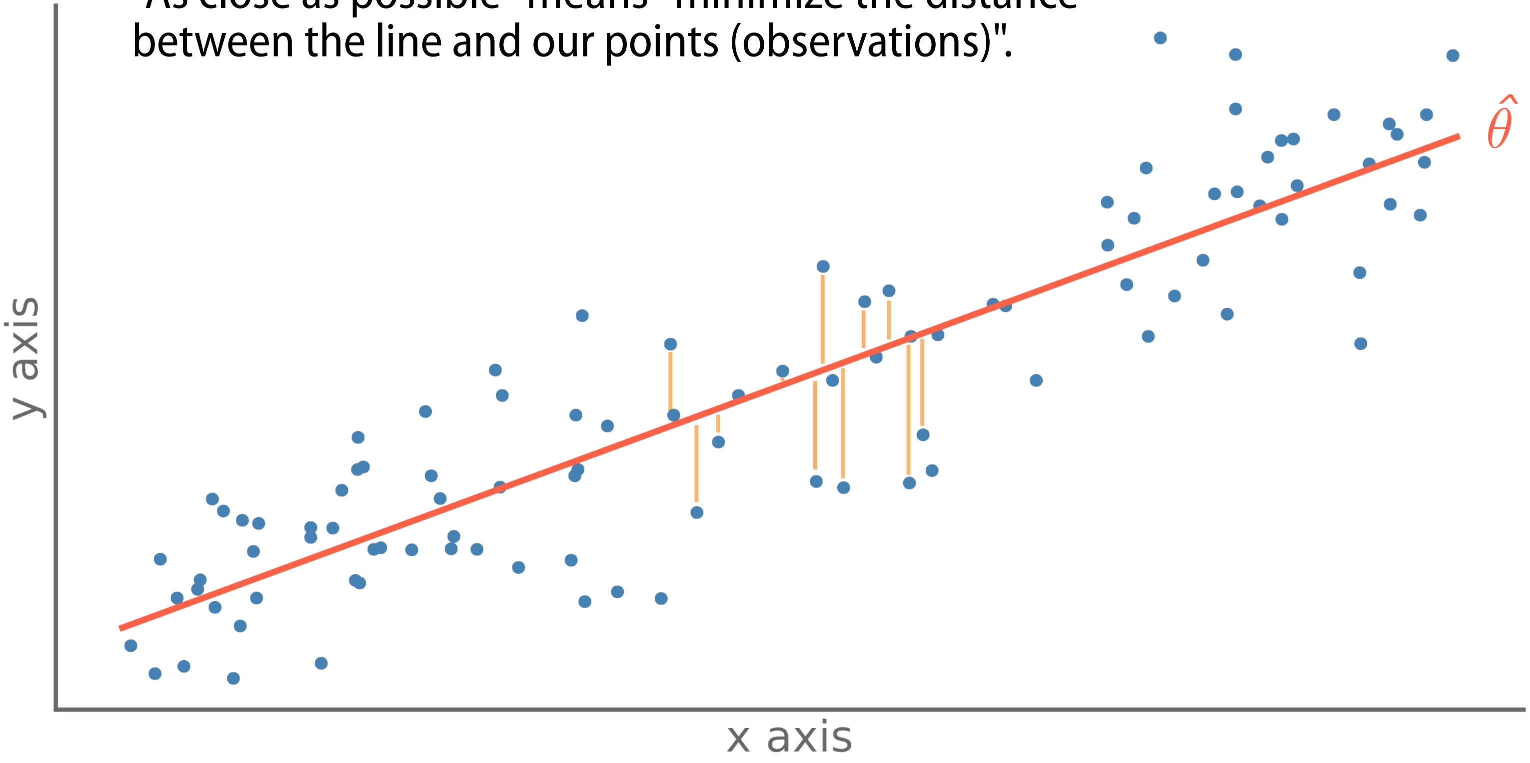
We are looking for a straight line that passes
« as close as possible » to our points.

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)".

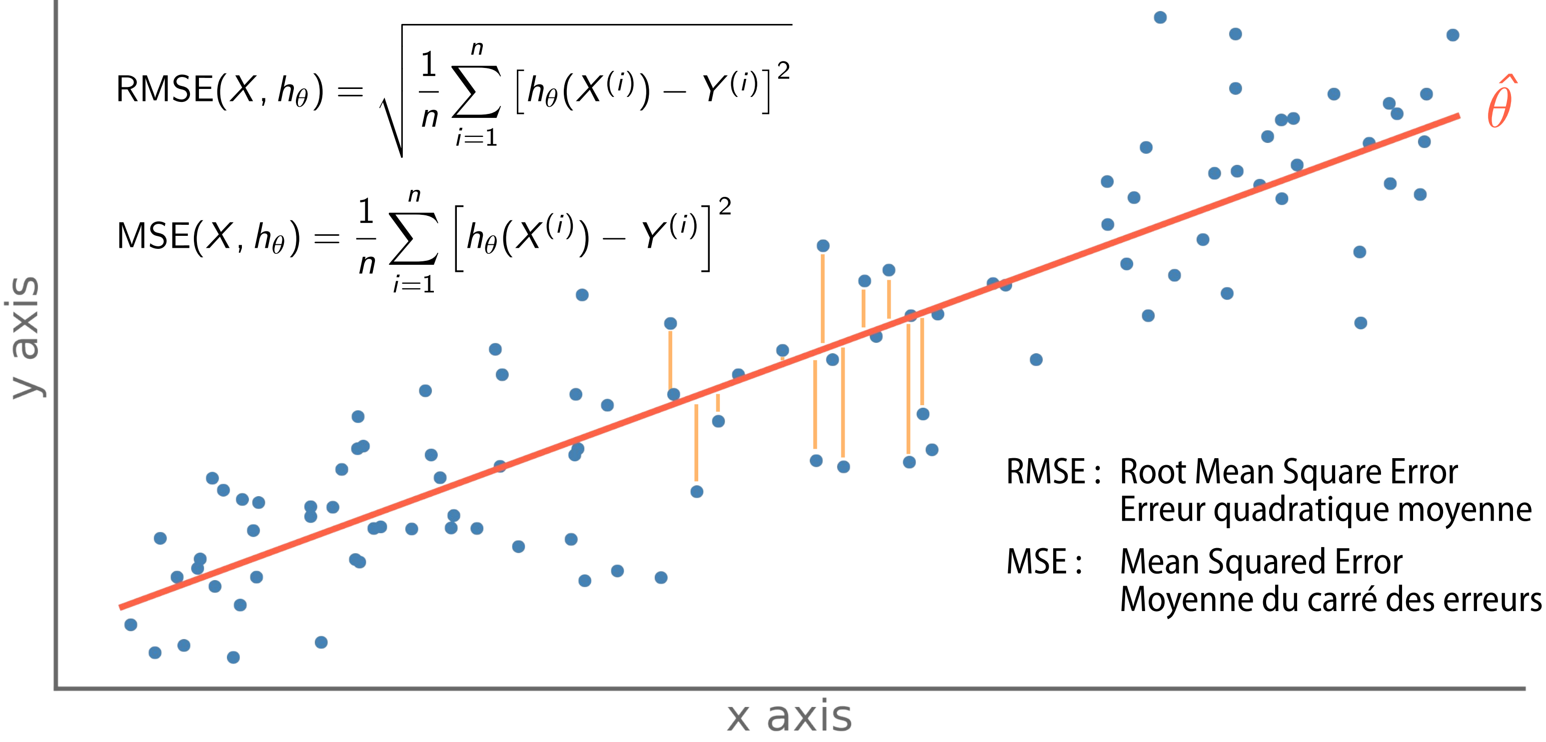


Linear regression

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

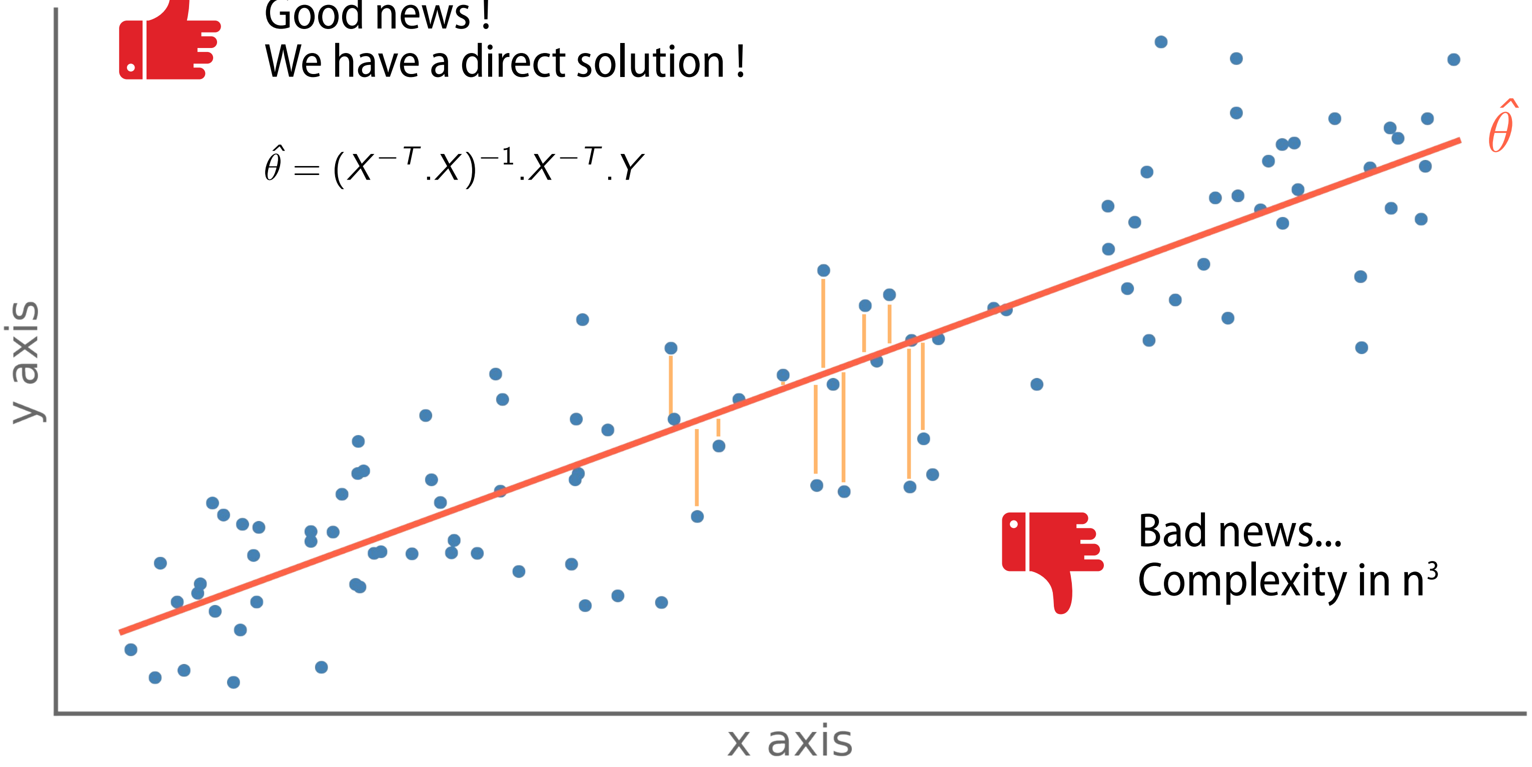
MSE : Mean Squared Error
Moyenne du carré des erreurs

Linear regression

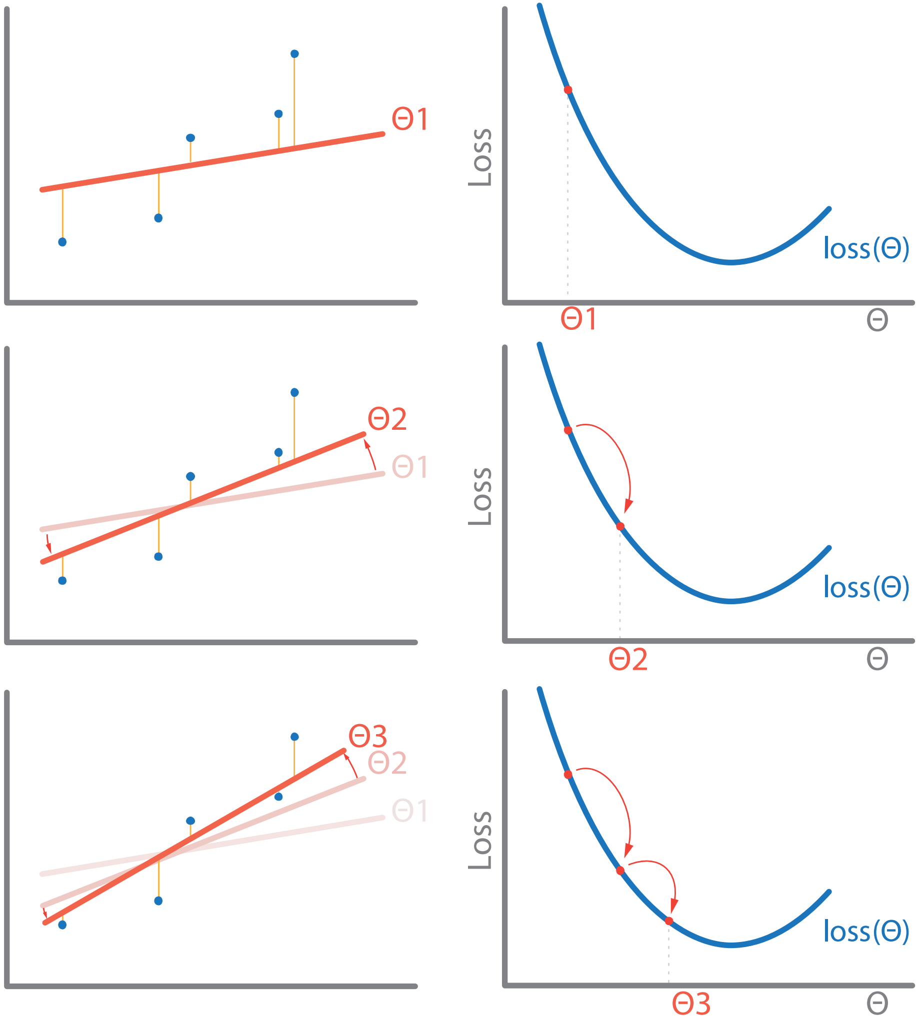
Good news !
We have a direct solution !

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

Bad news...
Complexity in n^3



Gradient descent



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

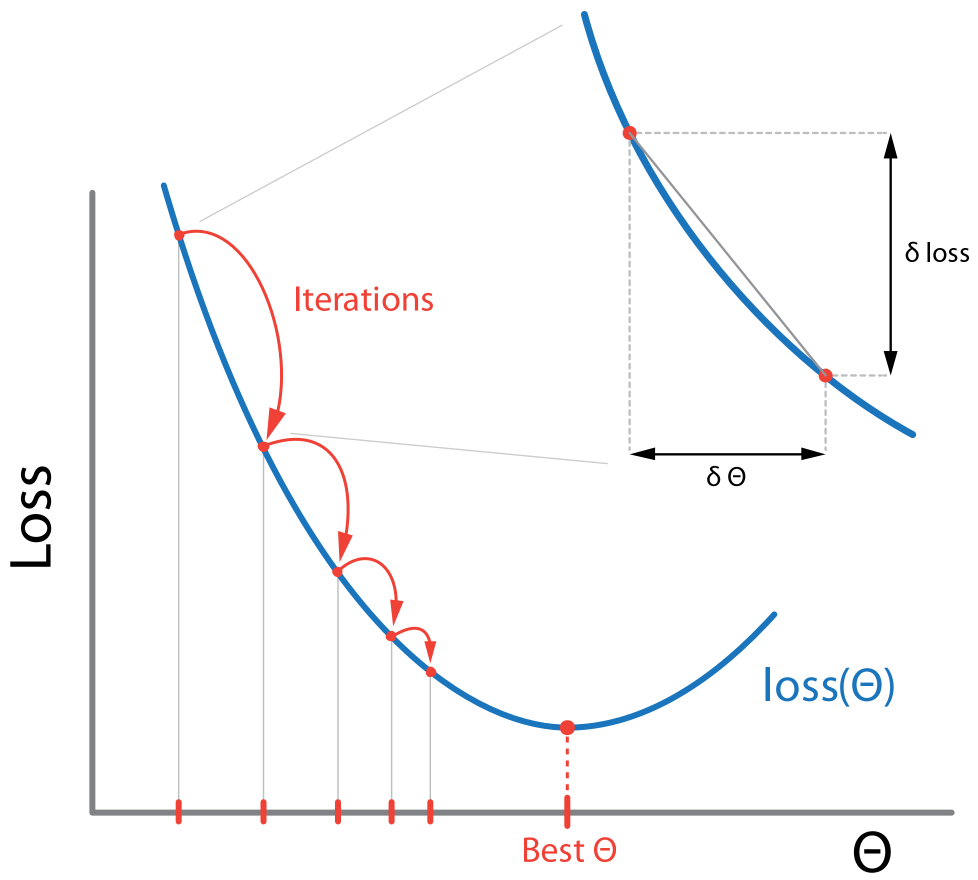
But how can we efficiently vary our parameters (Θ)?

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 : $\theta \leftarrow \theta - \eta \cdot \dfrac{\delta \text{loss}}{\delta \theta}$

where η is the learning rate

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

Gradient descent

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

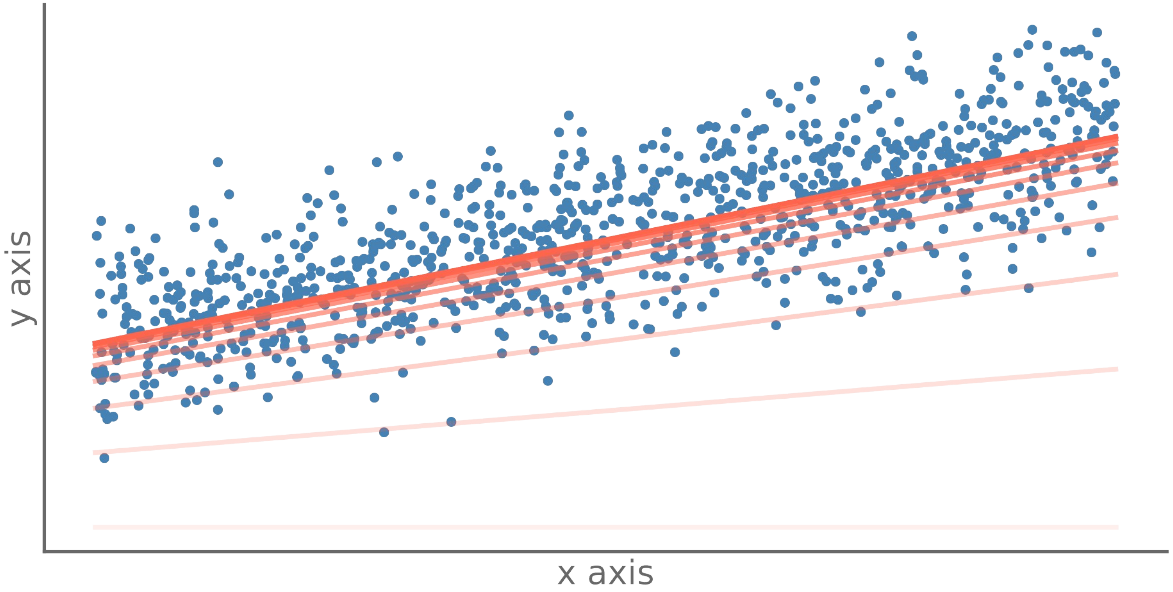
$$\nabla_\theta MSE(\Theta) = \begin{bmatrix} \frac{\partial}{\partial\theta_0} MSE(\Theta) \\ \frac{\partial}{\partial\theta_1} MSE(\Theta) \\ \vdots \\ \frac{\partial}{\partial\theta_m} MSE(\Theta) \end{bmatrix} = \frac{2}{n} X^T \cdot (X \cdot \Theta - Y)$$

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

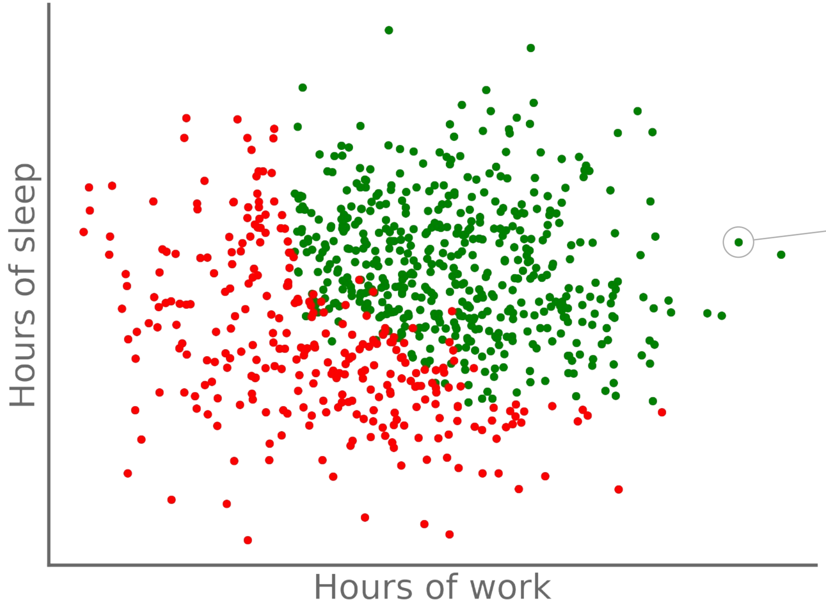


#i	Loss	Gradient		Theta	
0	+12.481	-6.777	-1.732	-3.388	+0.000
20	+4.653	-4.066	-1.039	-2.033	+0.346
40	+1.835	-2.440	-0.624	-1.220	+0.554
60	+0.821	-1.464	-0.374	-0.732	+0.679
80	+0.455	-0.878	-0.224	-0.439	+0.754
100	+0.324	-0.527	-0.135	-0.263	+0.799
120	+0.277	-0.316	-0.081	-0.158	+0.826
140	+0.260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

Logistic regression

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

Dataset : X Observations
y Classe



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

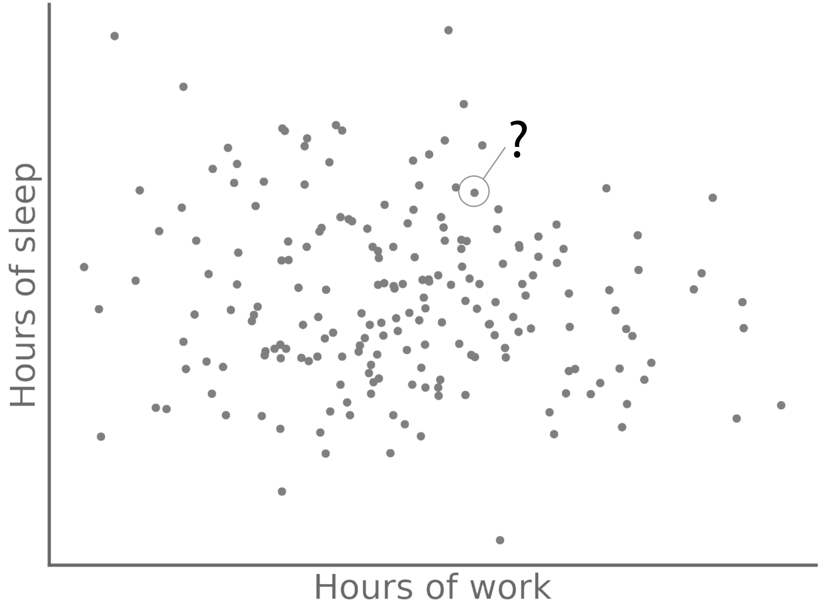
Notebook [LOGR1]

Logistic regression

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

Dataset : X Observations
y Classe

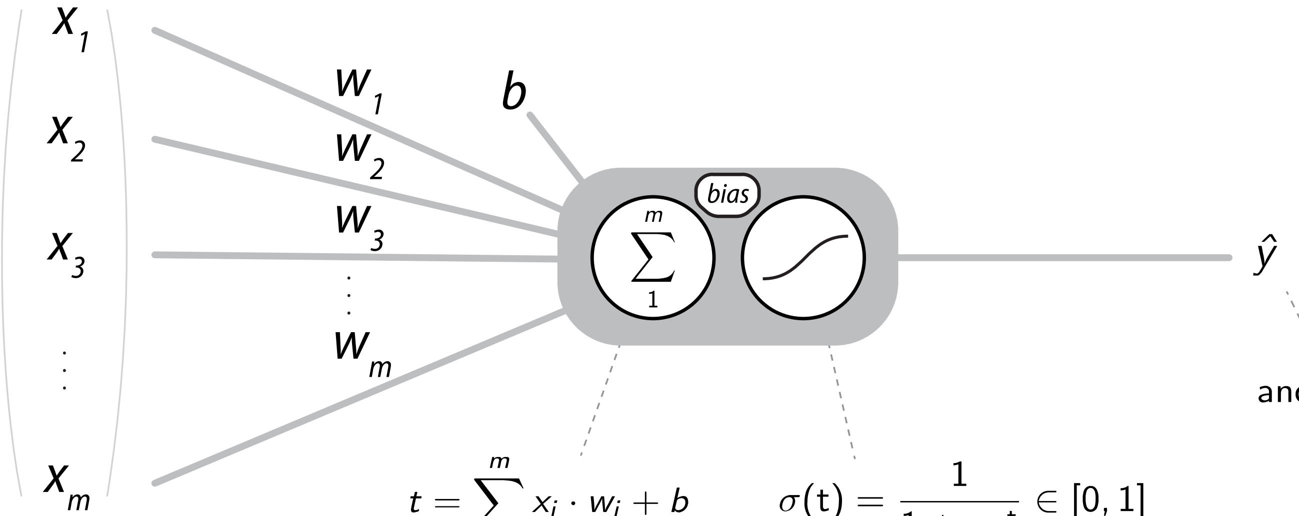
Objective : Predict the class
x given, we want to predict y
$\mathbf{y_{pred} = f(x)}$
where f is a linear function



Notebook [LOGR1]

Logistic regression

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



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

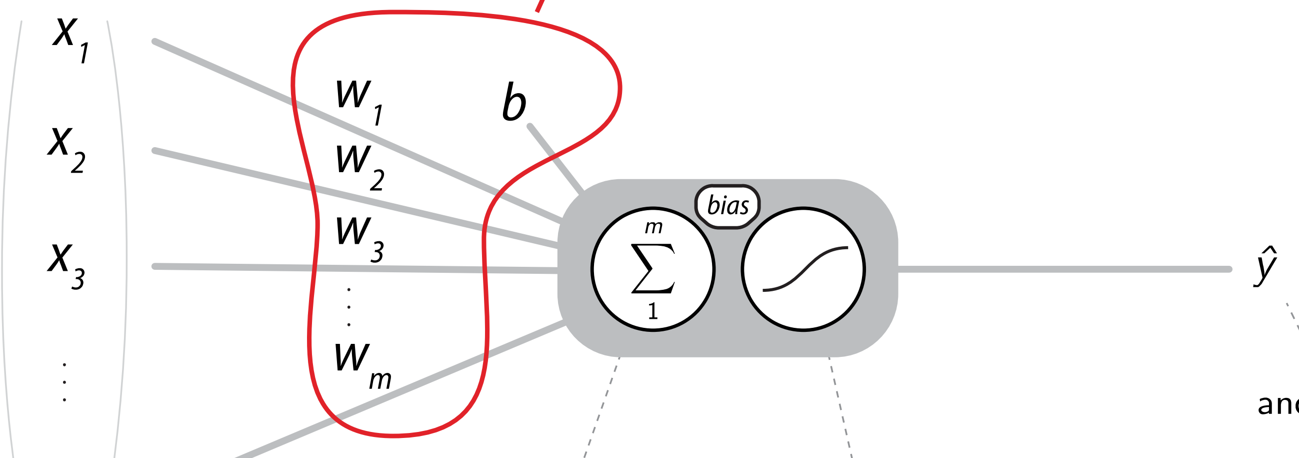
$$t = \sum_1^m x_i \cdot w_i + b \qquad \sigma(\mathrm{t}) = \frac{1}{1 + \mathrm{e}^{-\mathrm{t}}} \in [0, 1]$$

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

Logistic regression

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

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



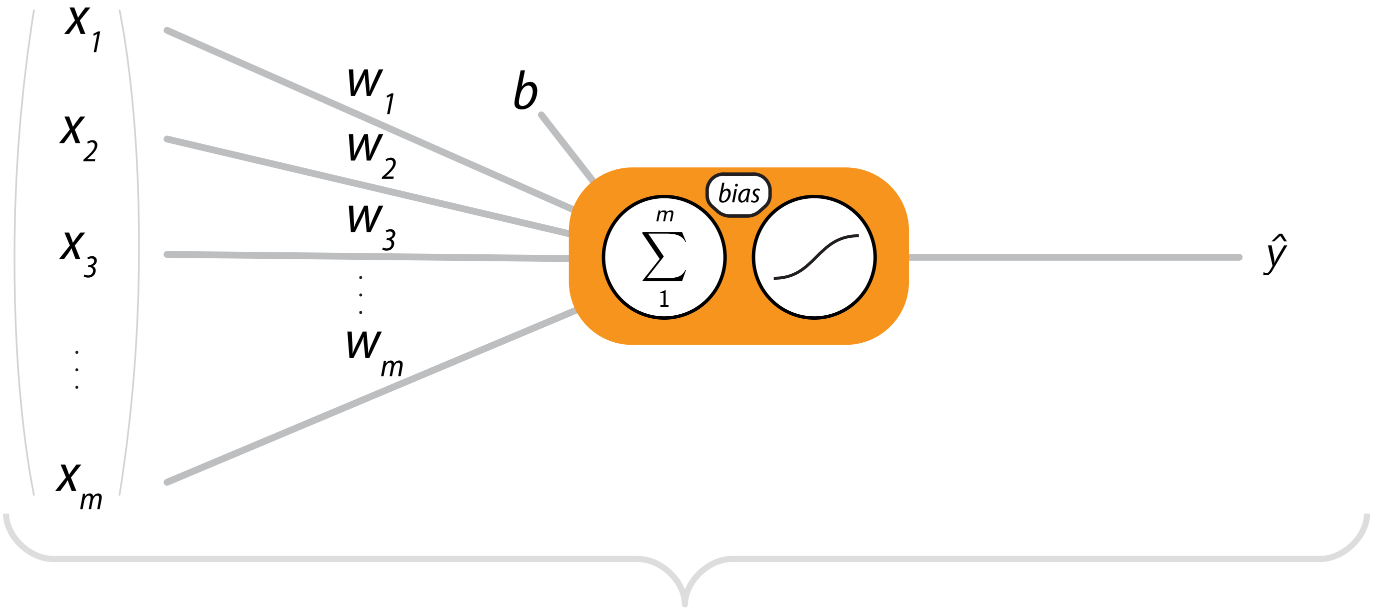
$$\text{and } \bar{y} = \begin{cases} 0 & \text{if } \hat{y} < 0.5 \\ 1 & \text{if } \hat{y} \geq 0.5 \end{cases}$$

$$t = \sum_1^m x_i \cdot w_i + b \qquad \sigma(t) = \frac{1}{1 + e^{-t}} \in [0, 1]$$

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

Logistic regression

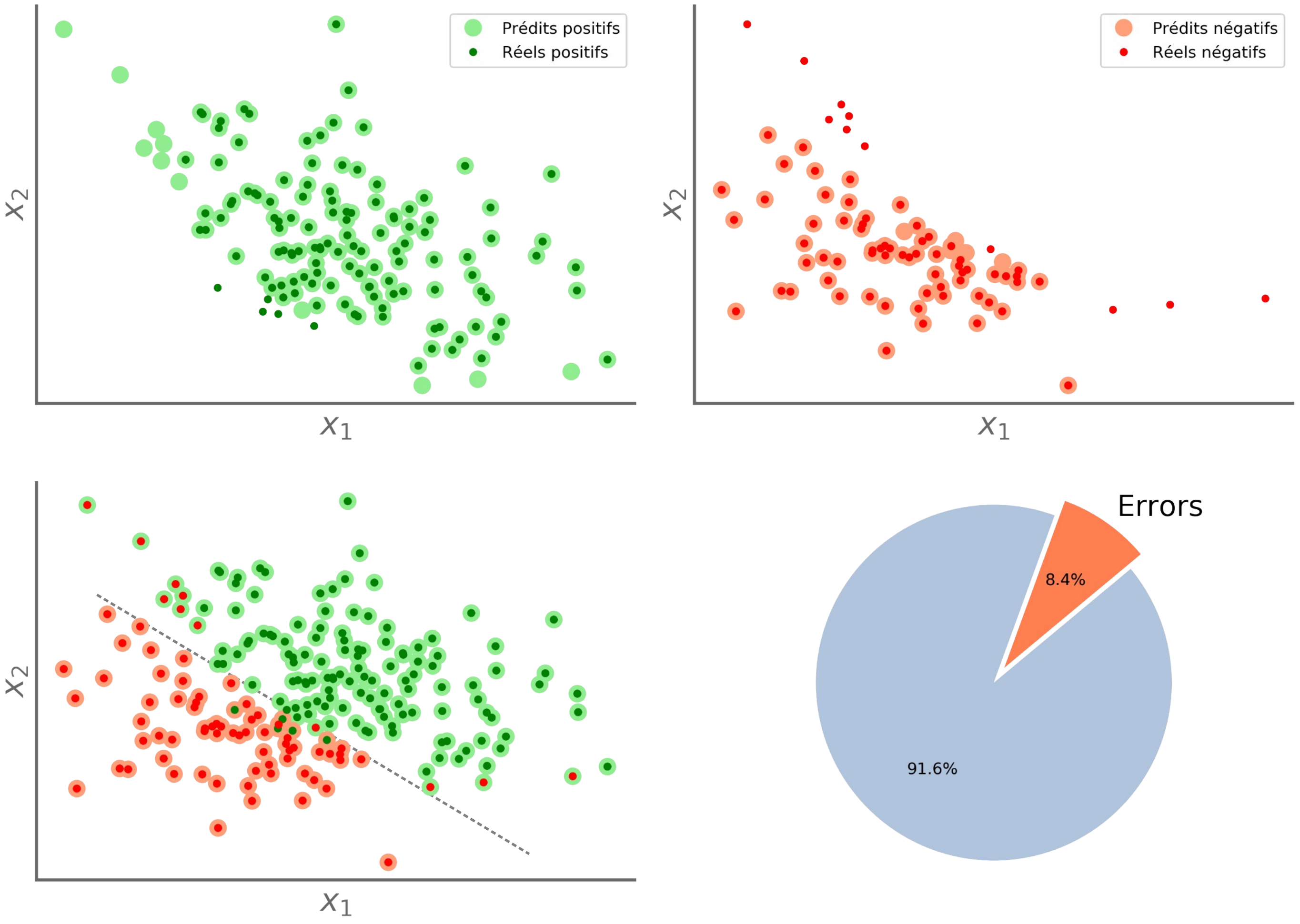
$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$



That's an « artificial neuron » !

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

Logistic regression



Notebook [LOGR1]

Neurons in controversy[1]

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

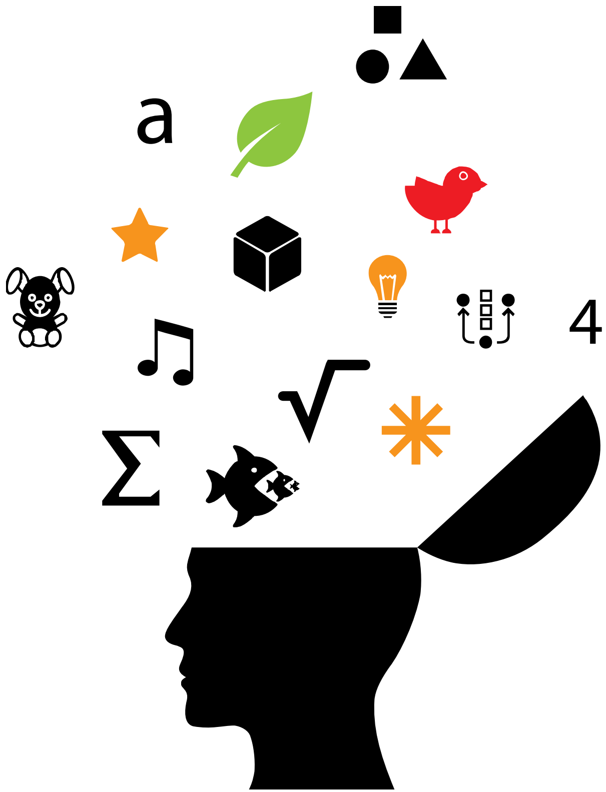
[intelligence]

[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 :

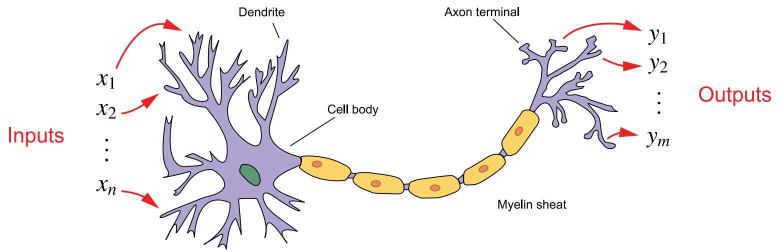
« Penser, c'est calculer des **symboles** qui
ont à la fois une réalité matérielle et une
valeur sémantique de représentation »[1]

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

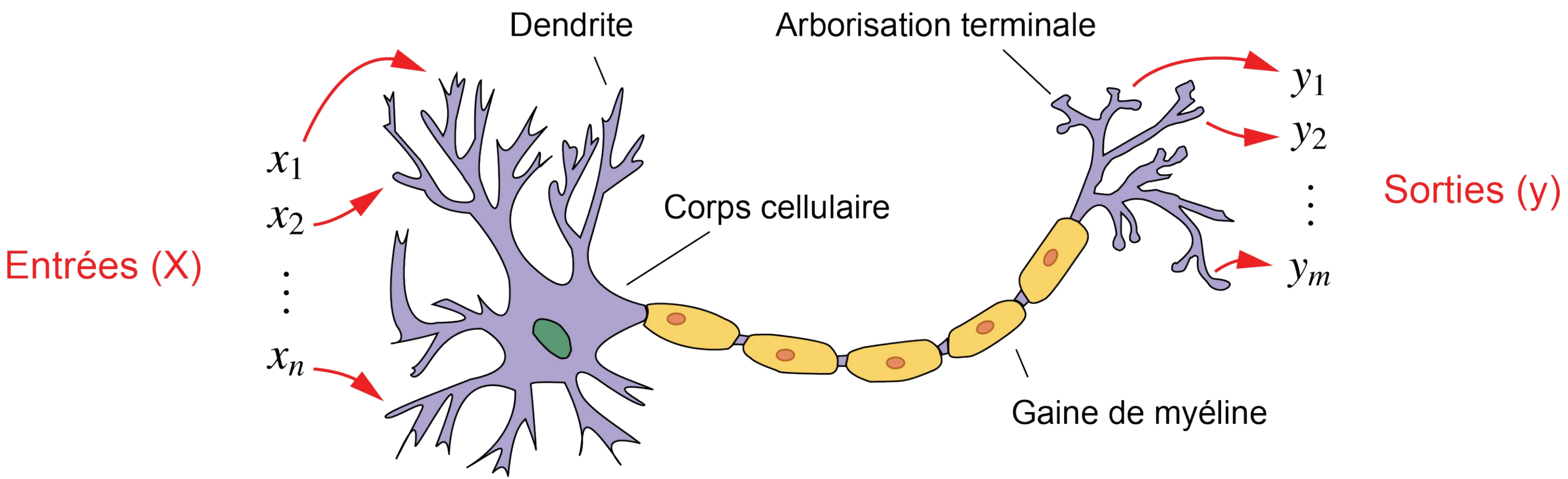
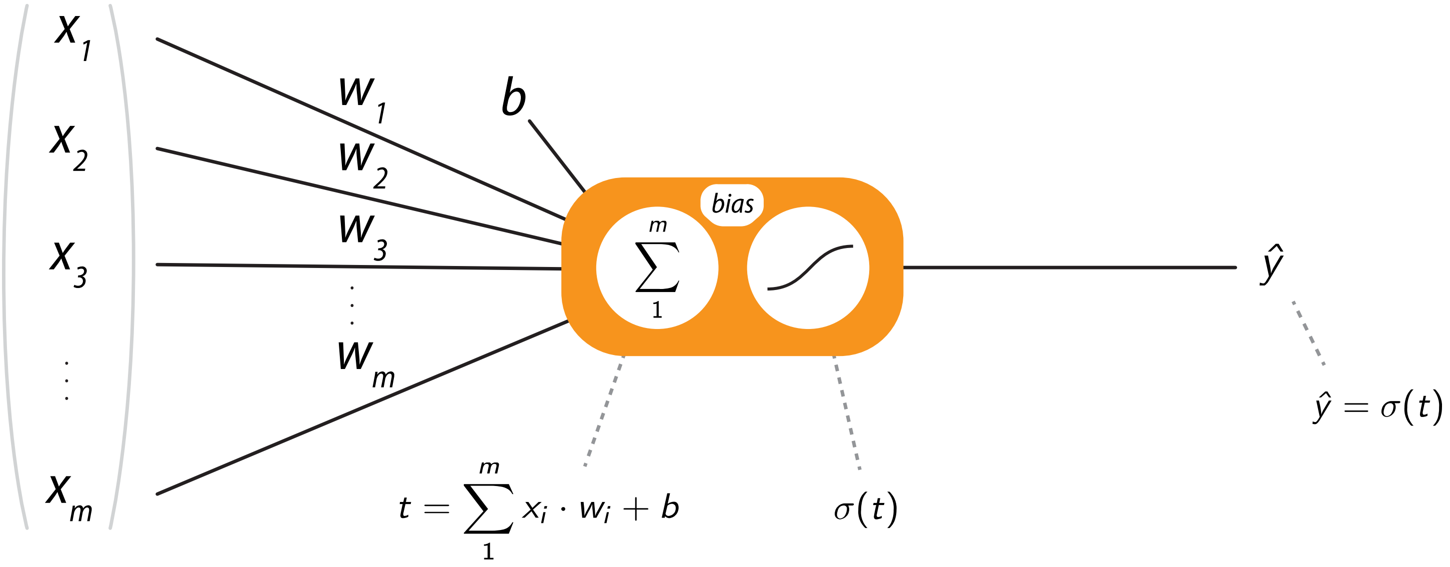


Illustration : [NEURON]

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



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**.

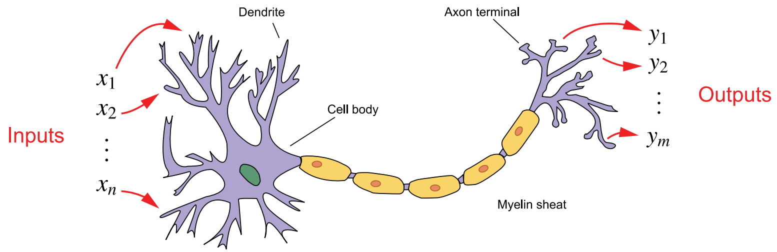
Connectionnism

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

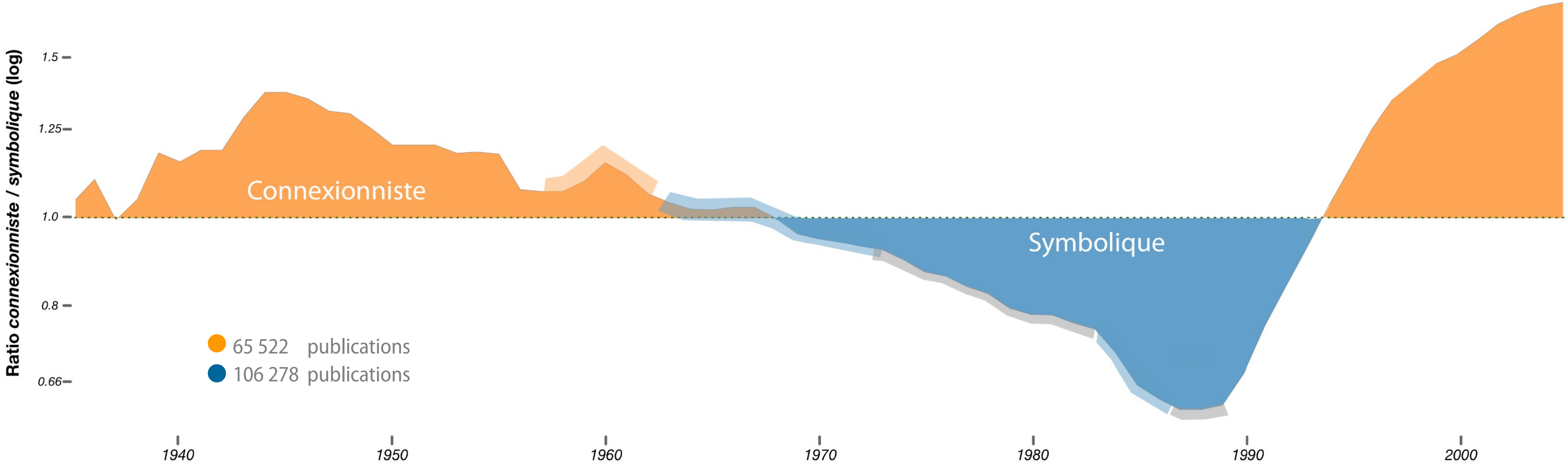
Expert

Rules and laws ▶ Special case

The big Controversy

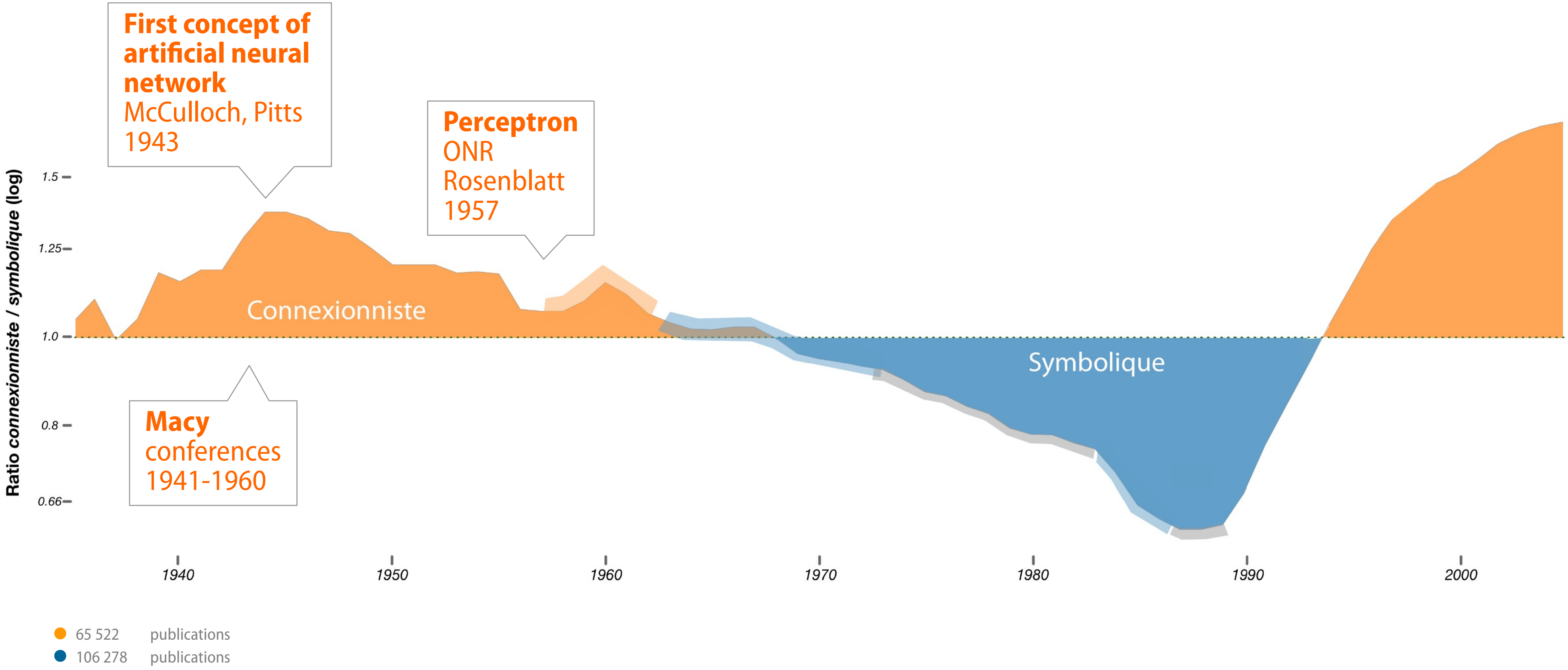
Evolution of the academic influence of connexionist and symbolic approaches[1]

Ration of publications between connexionists and symbolists



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

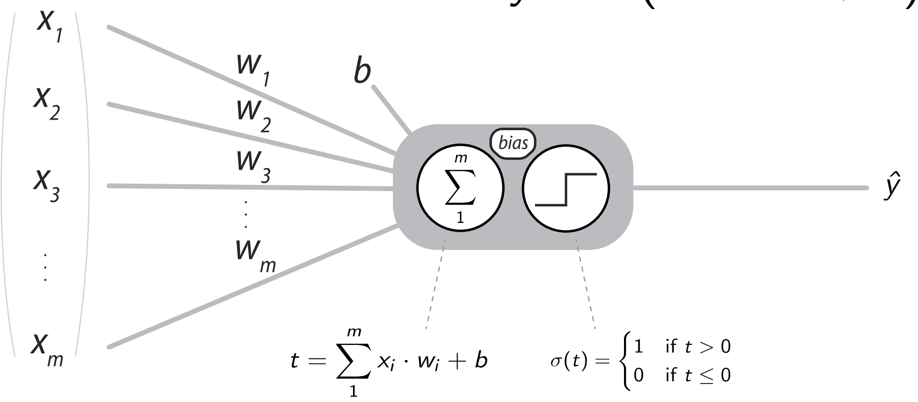
Evolution of the academic influence of connexionist and symbolic approaches[1]



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

Perceptron

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



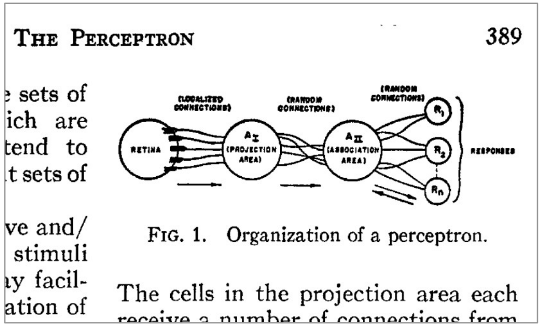

THE PERCEPTRON 389

Fig. 1. Organization of a perceptron.

The cells in the projection area each

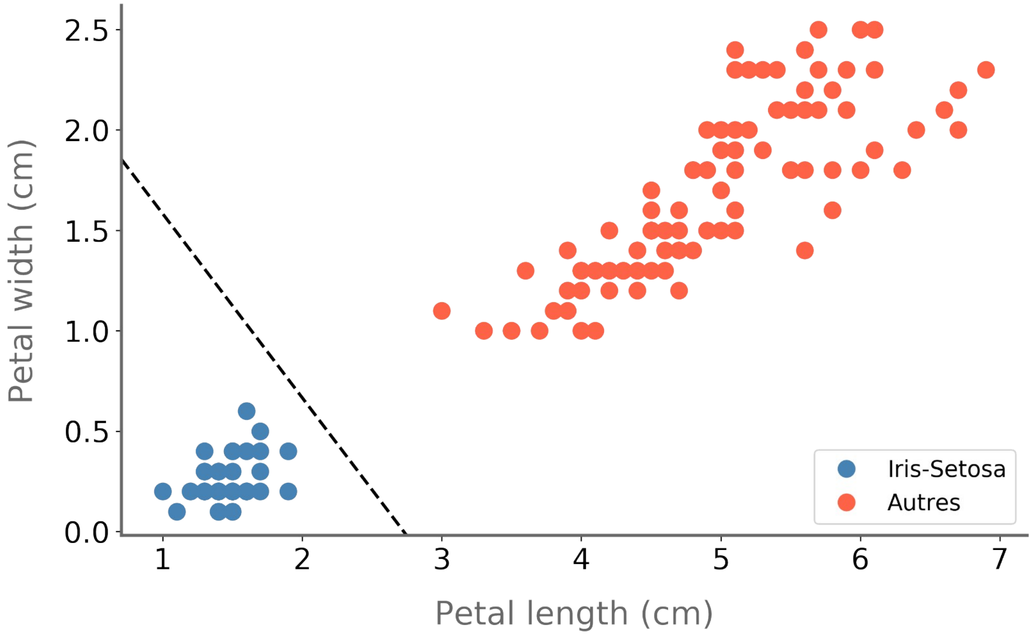
Perceptron
Frank Rosenblatt
1958

Linear and binary classifier

F. Rosenblatt, 1958 [FROS]

Iris plants dataset

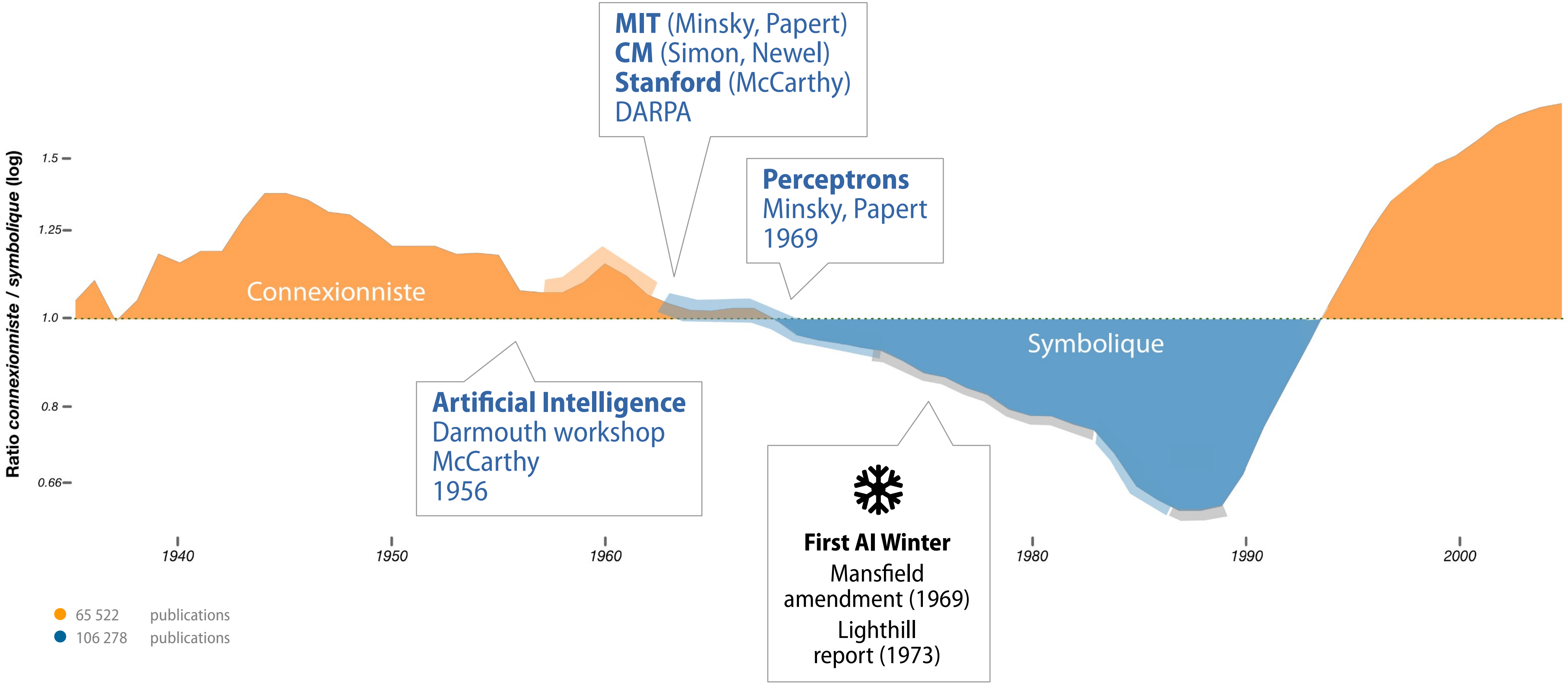
Dataset from : Fisher, R.A. "The use of multiple measurements in taxonomic problems" Annual Eugenics, 7, Part II, 179-188 (1936)



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]

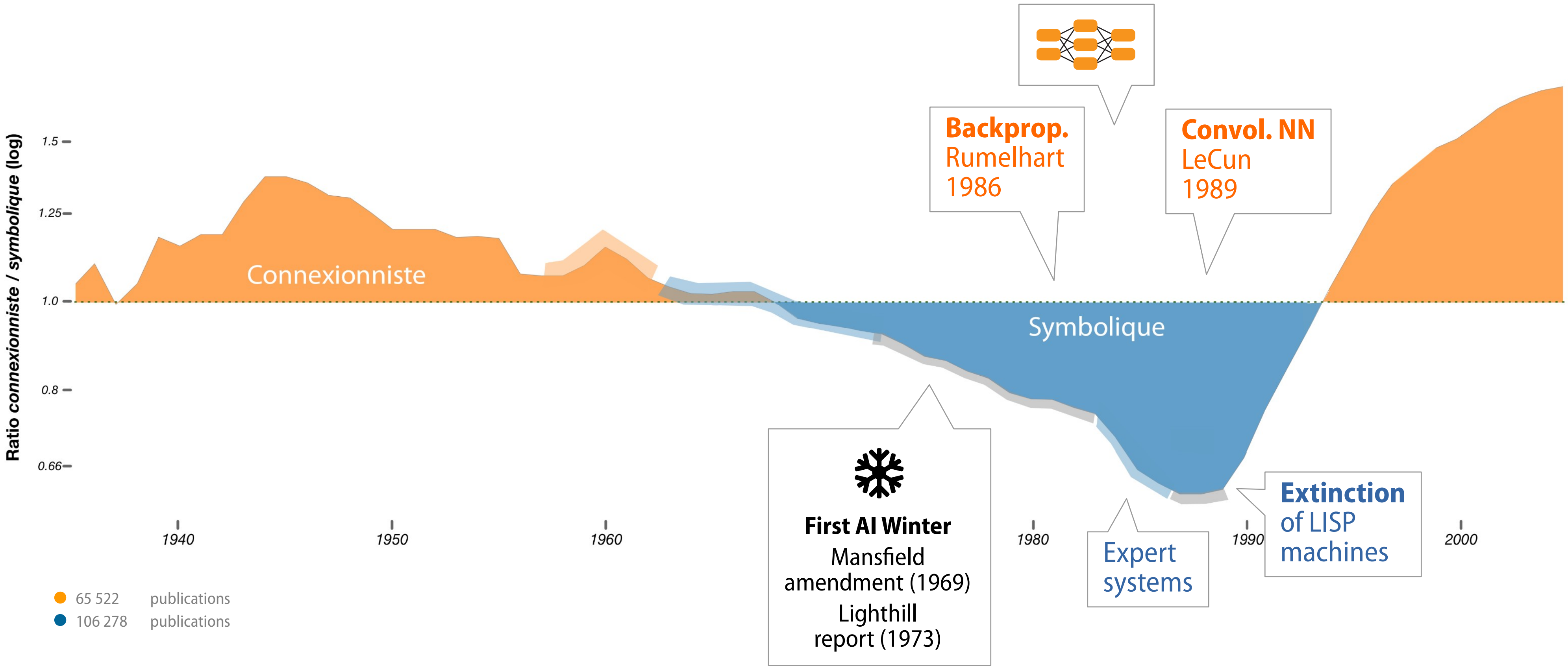
Evolution of the academic influence of connexionist and symbolic approaches[1]



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

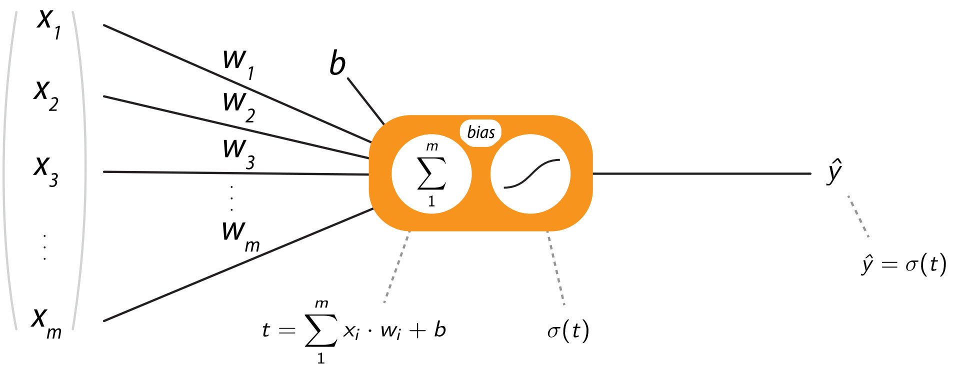
The big Controversy

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

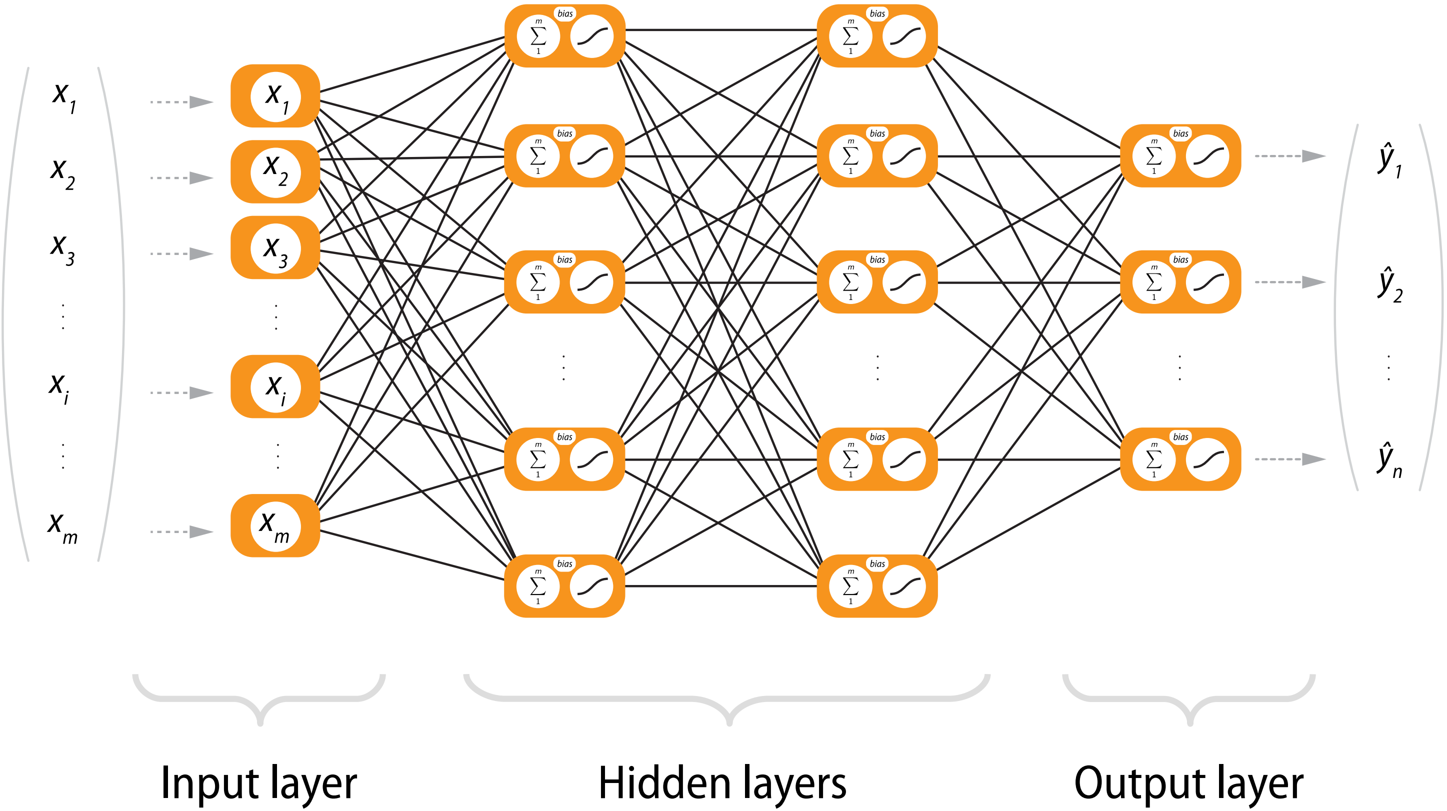


[1] D Cardon, JP Cointet, A Mazieres, 2018 [LRDN]

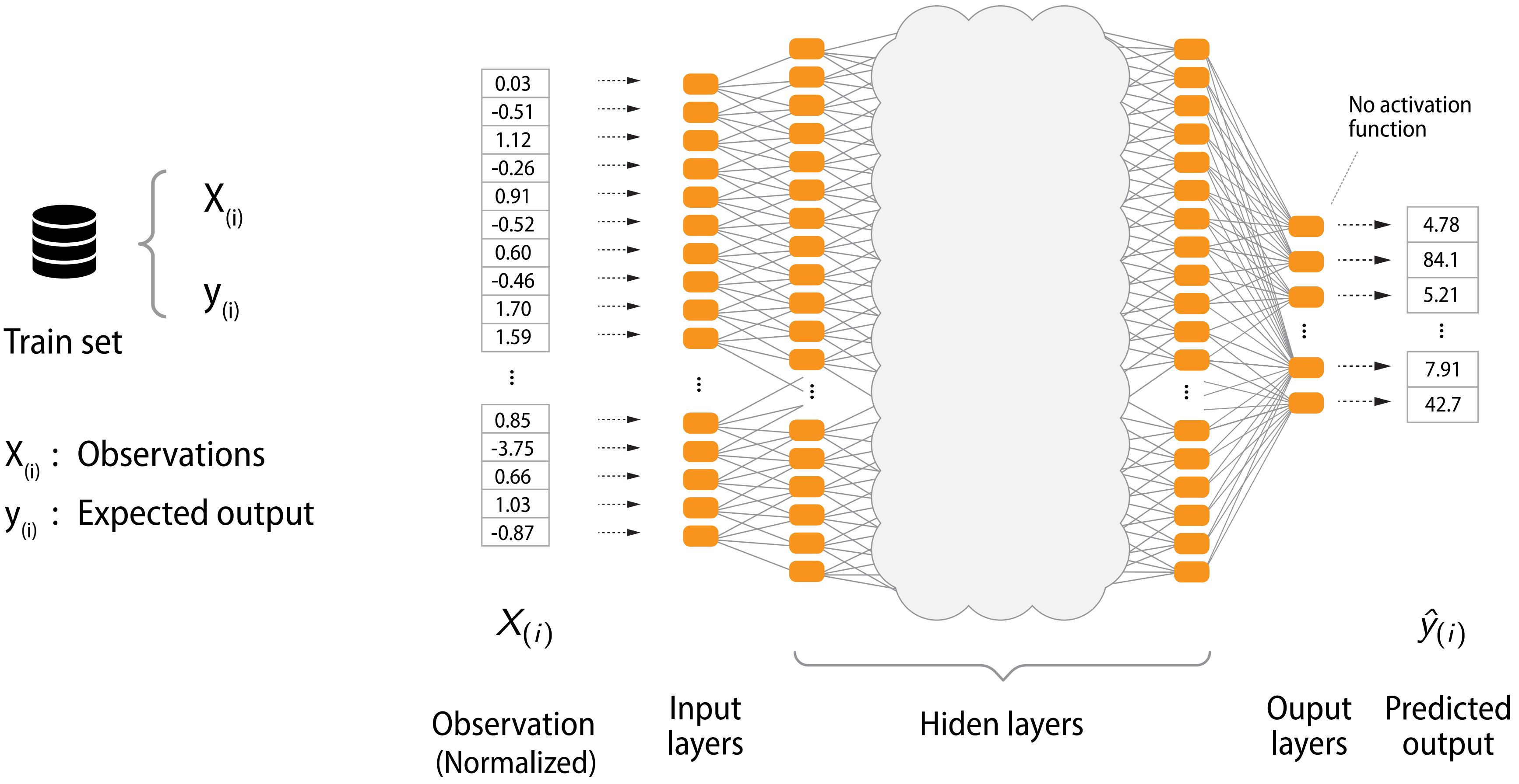
Deep Neural Networks



Deep Neural Networks



Deep Neural Networks



$X_{(i)}$: Observations

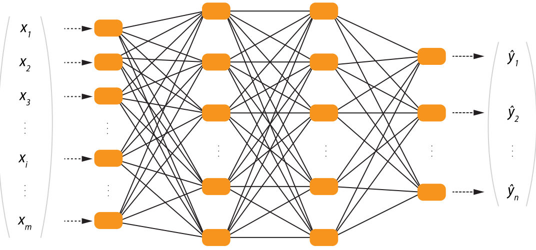
$y_{(i)}$: Expected output

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

$y^{(i)}$

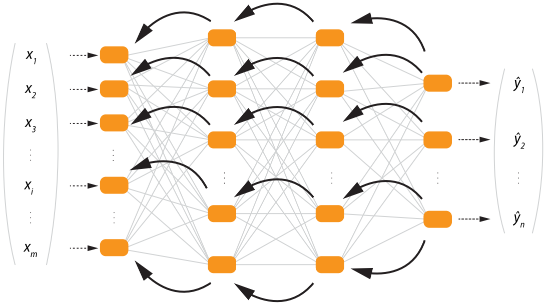
loss function

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

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

...

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 :

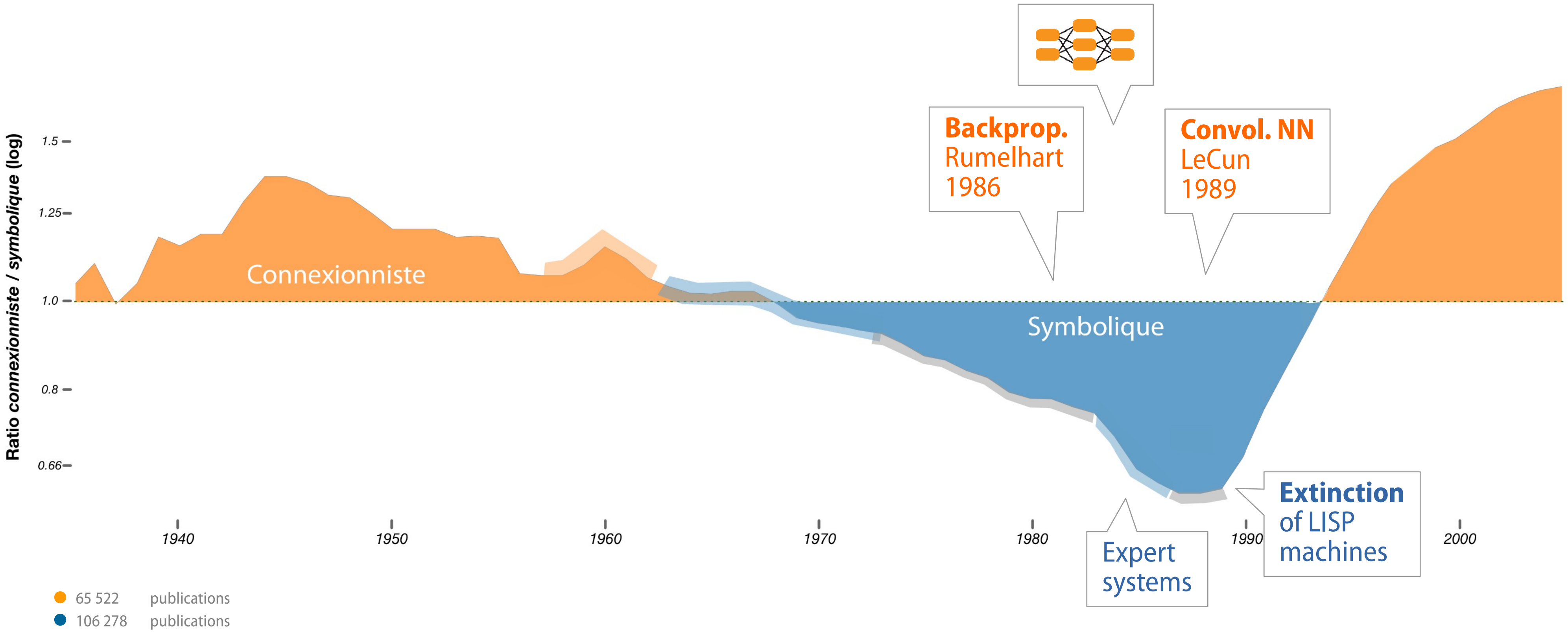
- Optimization
- Activation
- Loss
- Metrics
- ...

Back-propagation

Learning process

The big Controversy

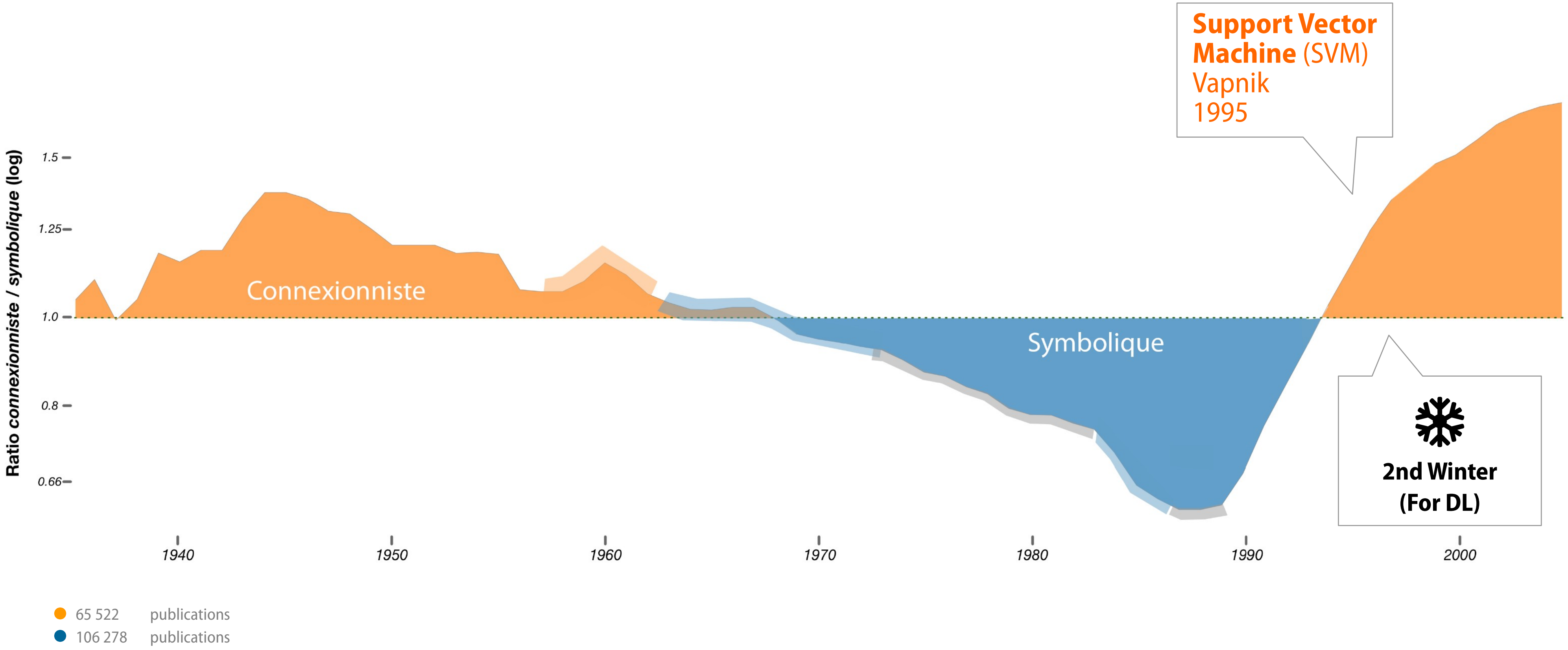
Evolution of the academic influence of connexionist and symbolic approaches[1]



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

The big Controversy

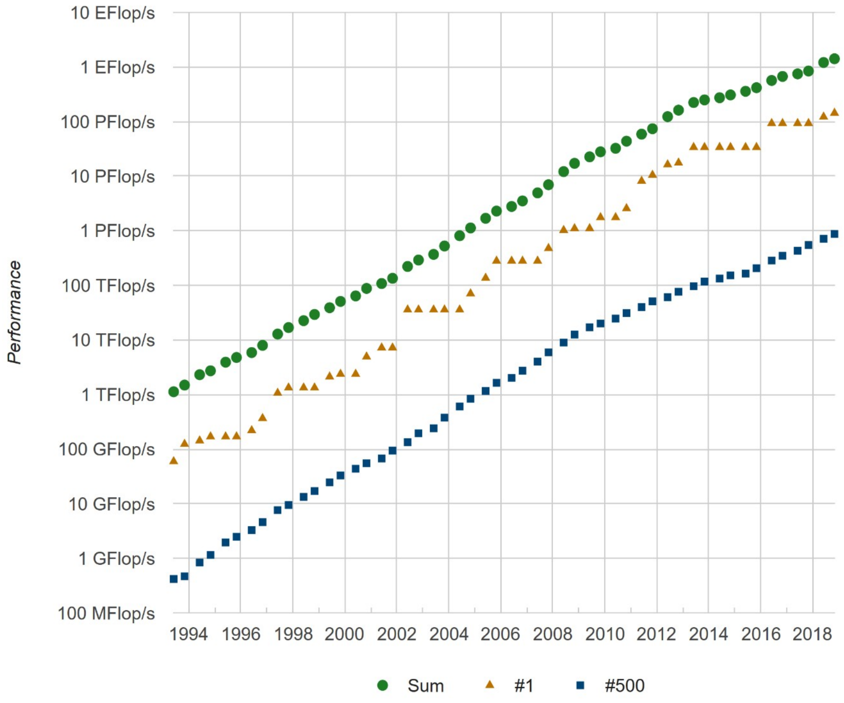
Evolution of the academic influence of connexionist and symbolic approaches[1]



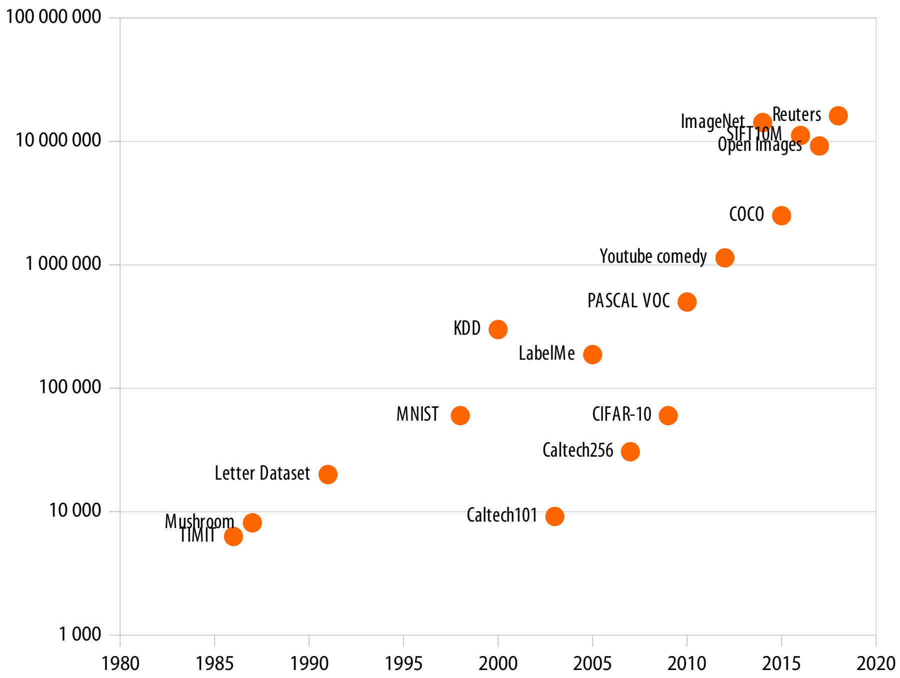
[1] D Cardon, JP Cointet, A Mazieres, 2018 [LRDN]

The big Controversy

Performance Development[1]



Datasets for machine-learning[2]



Laboratoire
Cas particulier

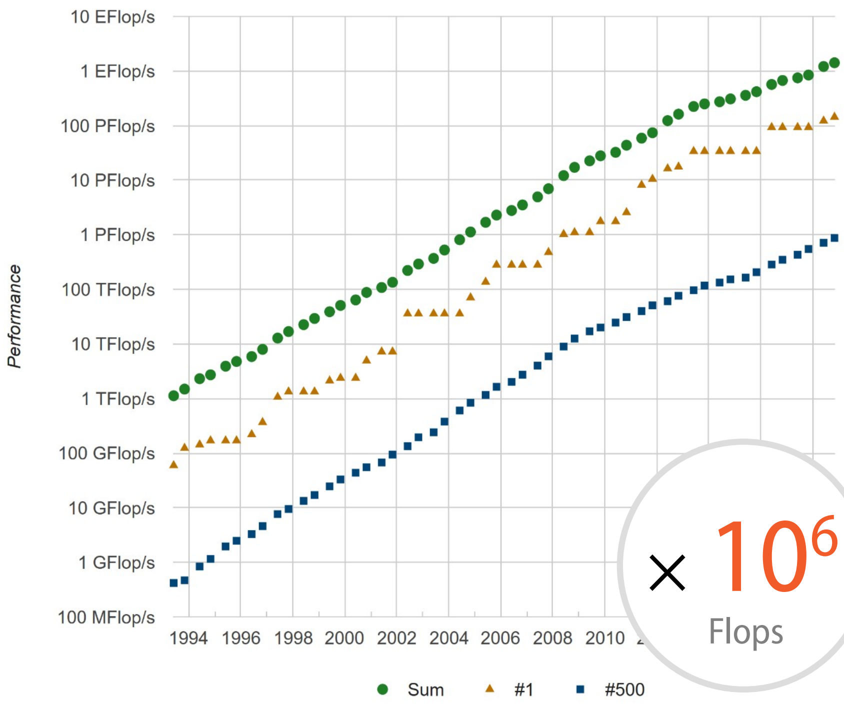
Monde réel

[1] TOP500 List [TOP500]

[2] Wikipedia [WKP1]

The big Controversy

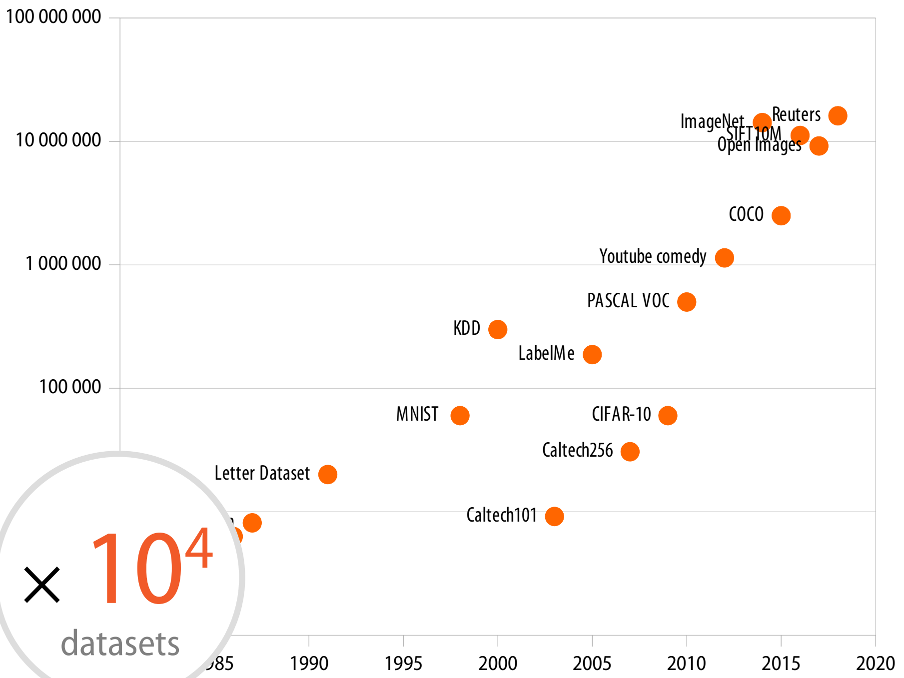
Performance Development[1]



$\times 10^6$ Flops

25 ans

Datasets for machine-learning[2]



$\times 10^4$ datasets

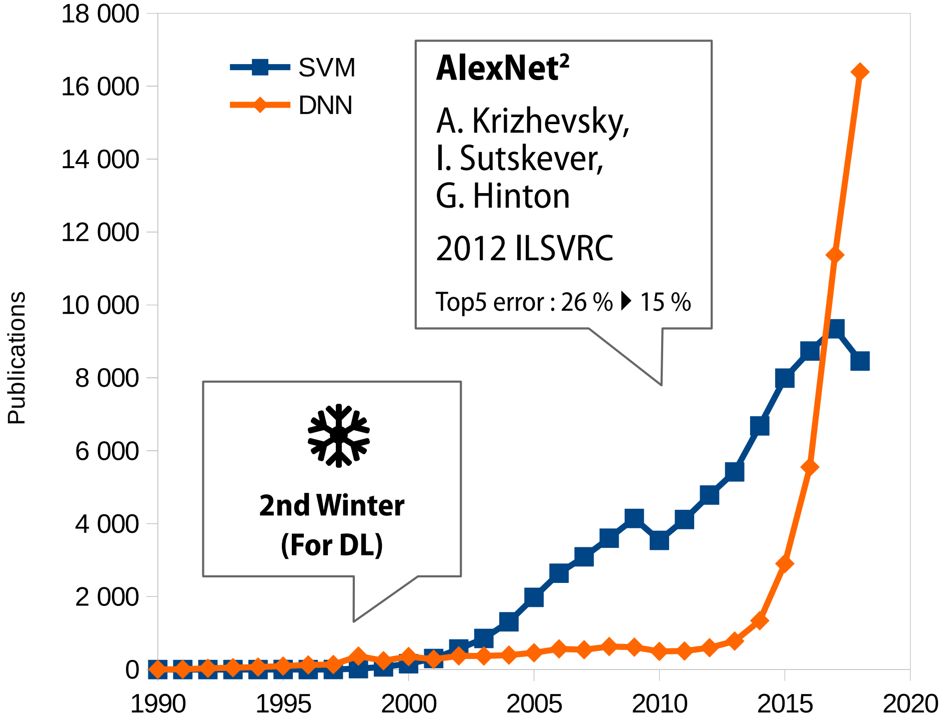
Laboratoire
Cas particulier $\not\Rightarrow$ Monde réel

[1] TOP500 List [TOP500]

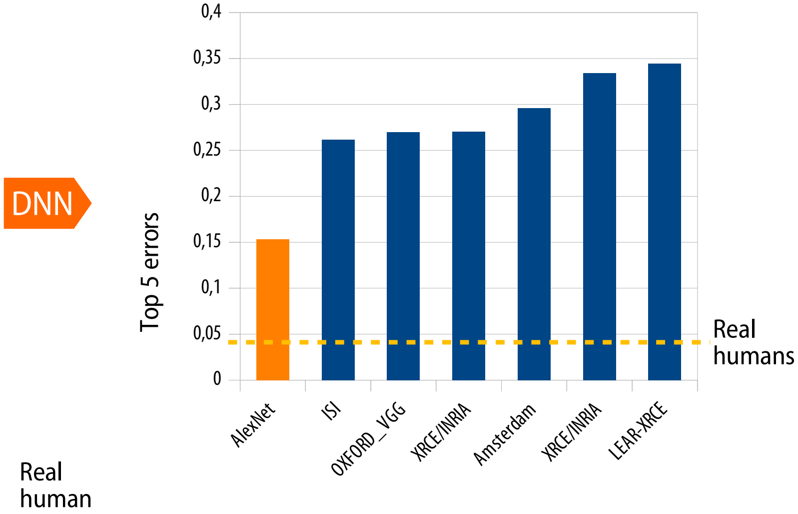
[2] Wikipedia [WKP1]

The big Controversy

Publications SVM vs DNN[1]



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



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, ...

Data and neurons

Regression with a DNN

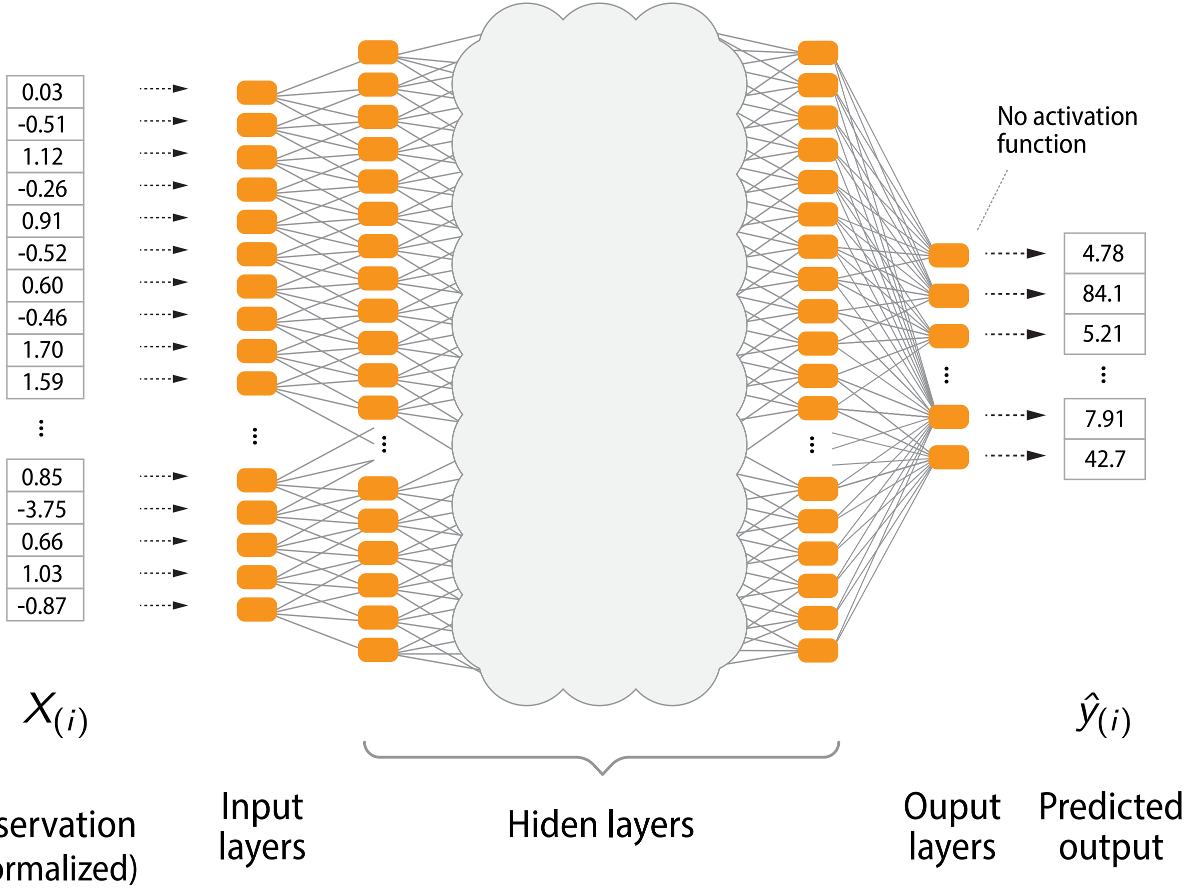
Train set

$X_{(i)}$

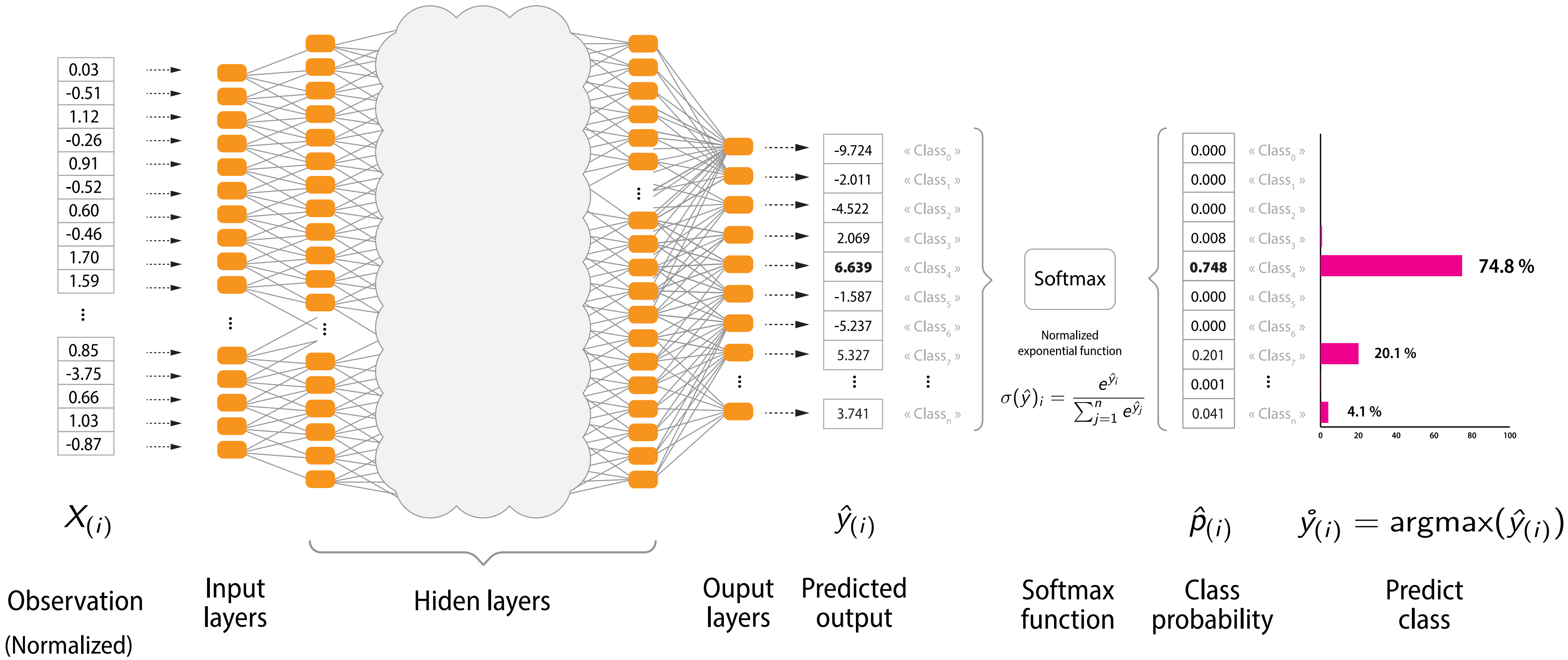
$y_{(i)}$

$X_{(i)}$: Observations

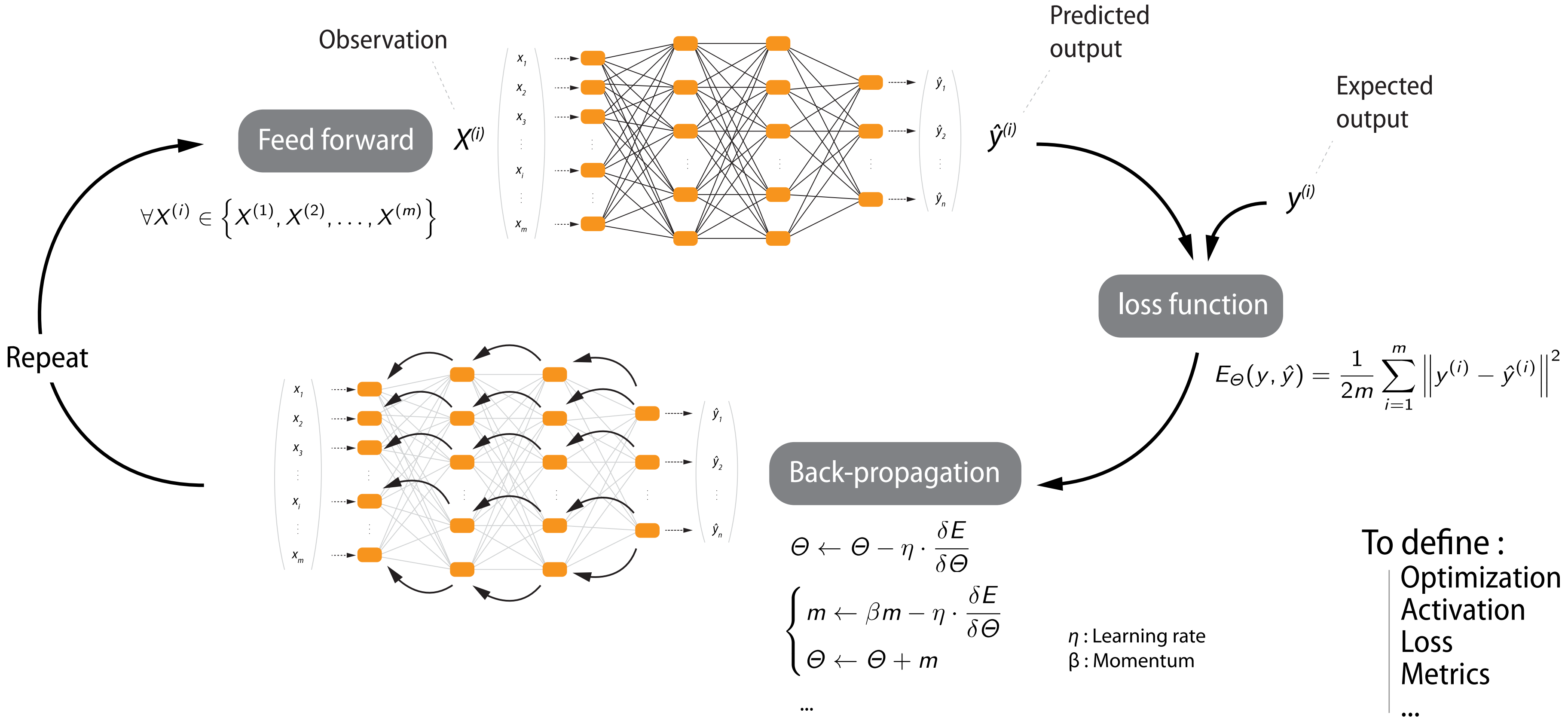
$y_{(i)}$: Expected output



Classification with a DNN



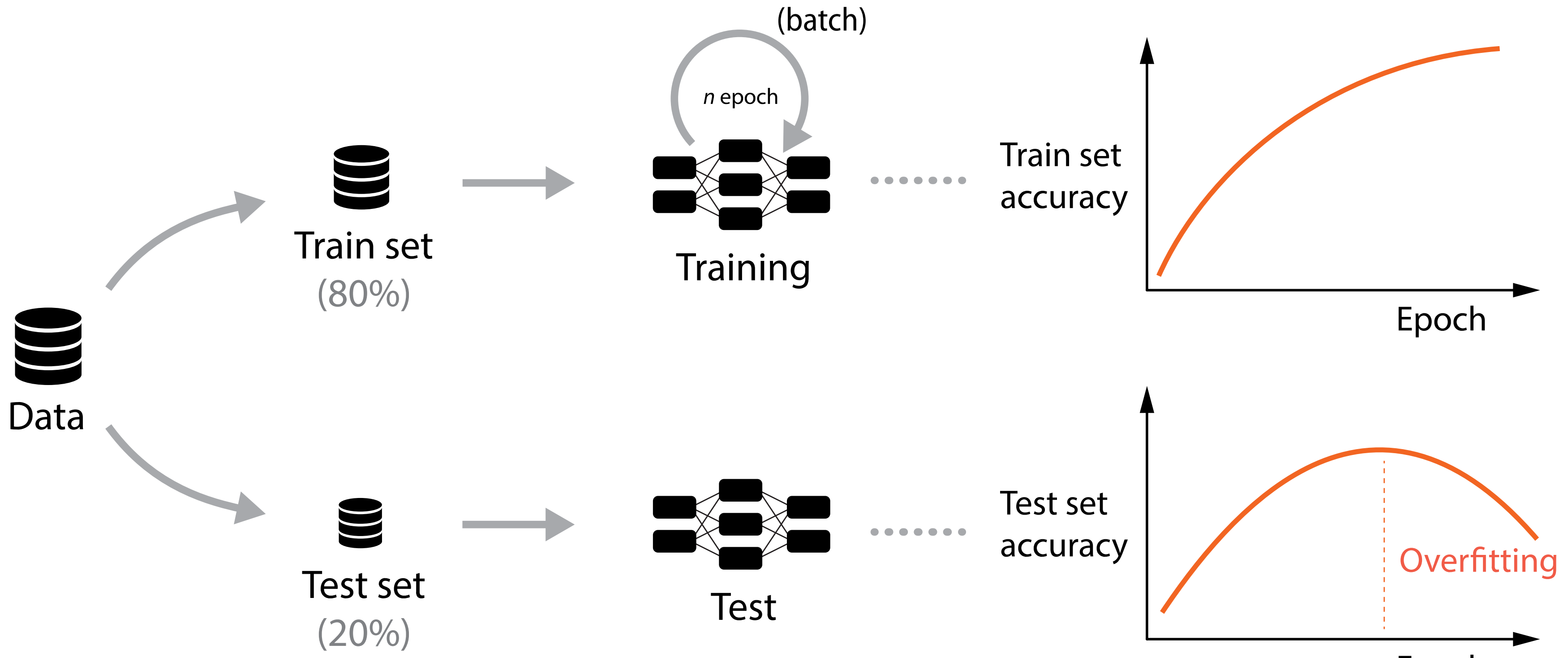
Training process - general



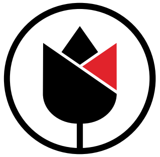
To define :

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

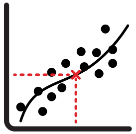
Training process - general



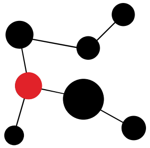
Neurons and data



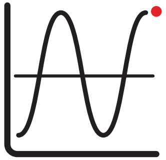
Hight Dimensionnal Data
CNN



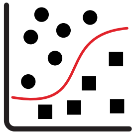
Basic Regression
DNN



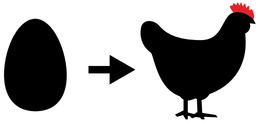
Graph Neural Network
GNN



Sequences data
RNN



Basic Classification
DNN



«Attention is All You Need»
Transformers

...

Neurons and data

Basic Regression
DNN

Basic Classification
DNN

Hight Dimensionnal Data
CNN

Sequences data
RNN

«Attention is All You Need»
Transformers

Graph Neural Network
GNN

...

Variational Antoencoder
VAE

Object Detection
R-CNN

Autoencoder networks
AE

Deep Reinforcement Learning
RL

...

Generative Adversarial Networks
GAN

Object Segmentation
Mask R-CNN

Neurons and data

Basic Regression
DNN

Hight Dimensionnal Data
CNN

Graph Neural Network
GNN

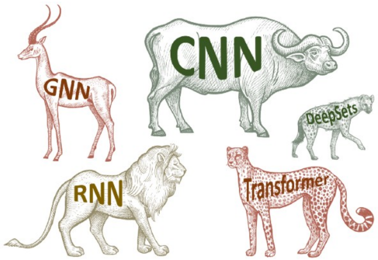
Sequences data
RNN

Basic Classification
DNN

«Attention is All You Need»
Transformers

...

«Zoo of neural network architecturesfor different kinds of data»



Geometric Deep Learning ?

BRONSTEIN, Michael M., BRUNA, Joan, COHEN, Taco, et al. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. arXiv preprint arXiv:2104.13478, 2021.a

Variational Antoencoder
VAE

Object Detection
R-CNN

Autoencoder networks
AE

Deep Reinforcement Learning
RL

...

Generative Adversarial Networks
GAN

Object Segmentation
Mask R-CNN

Pipelines

- For a slightly longer version -

FIDLE

Formation

Introduction au
Deep Learning

1er épisode

Jeudi 25 nov. 14h

https://fidle.cnrs.fr

En distanciel – Sans inscription – Accès libre

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

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