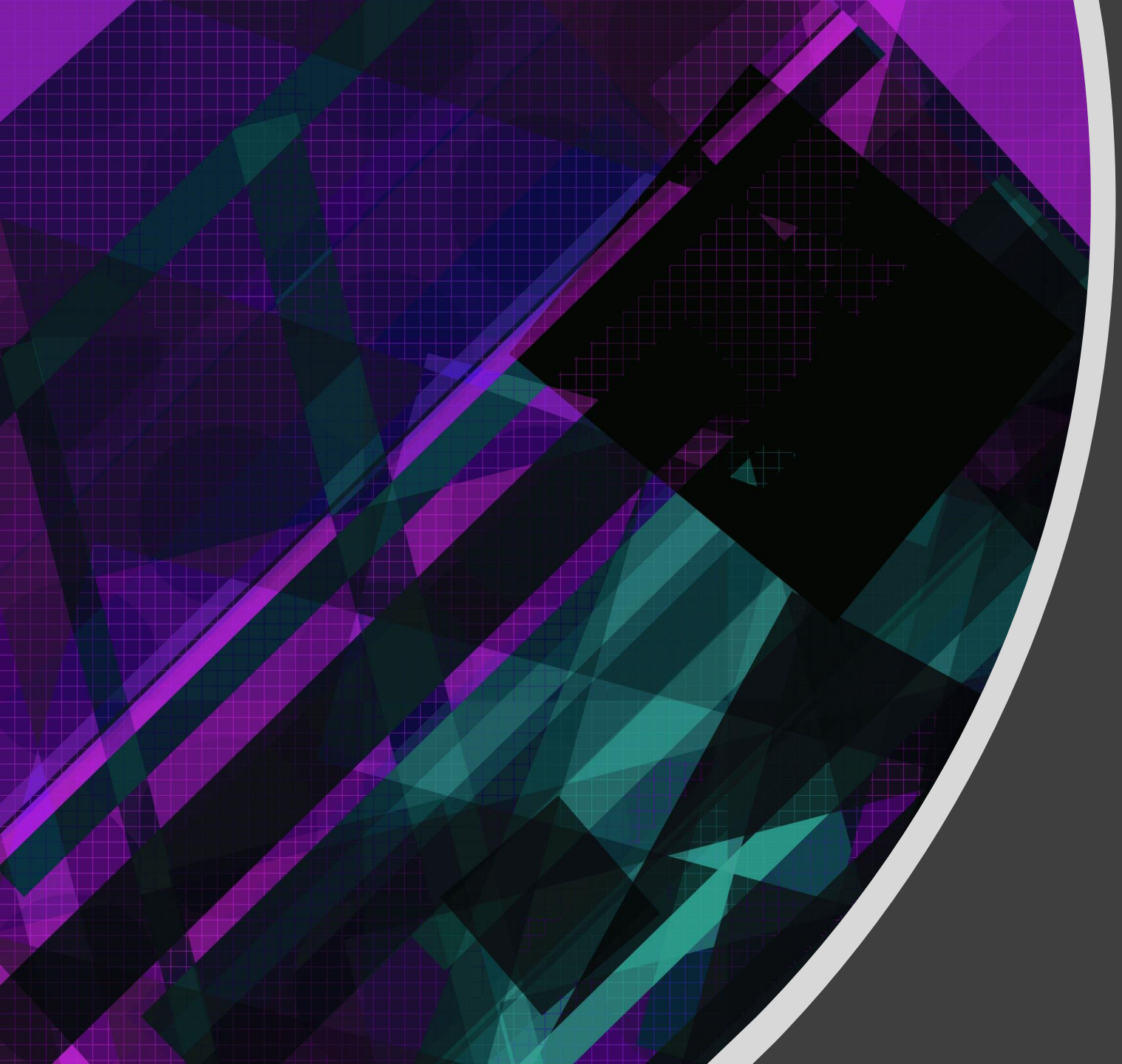
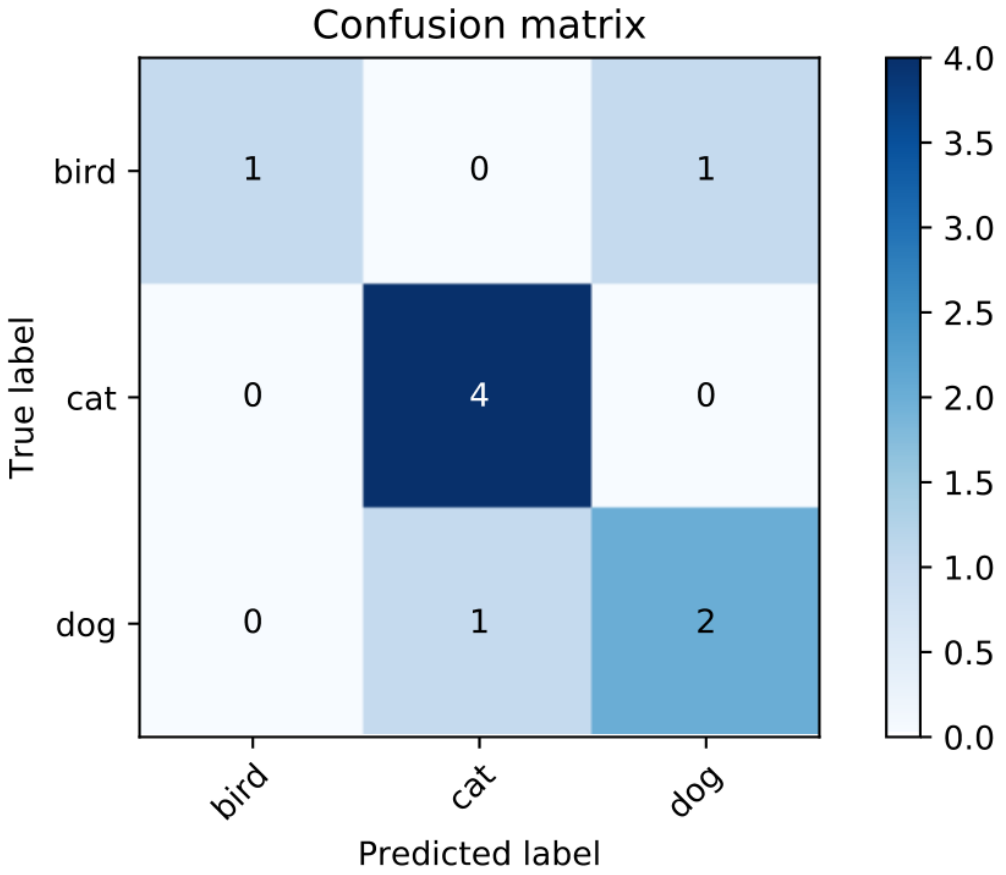
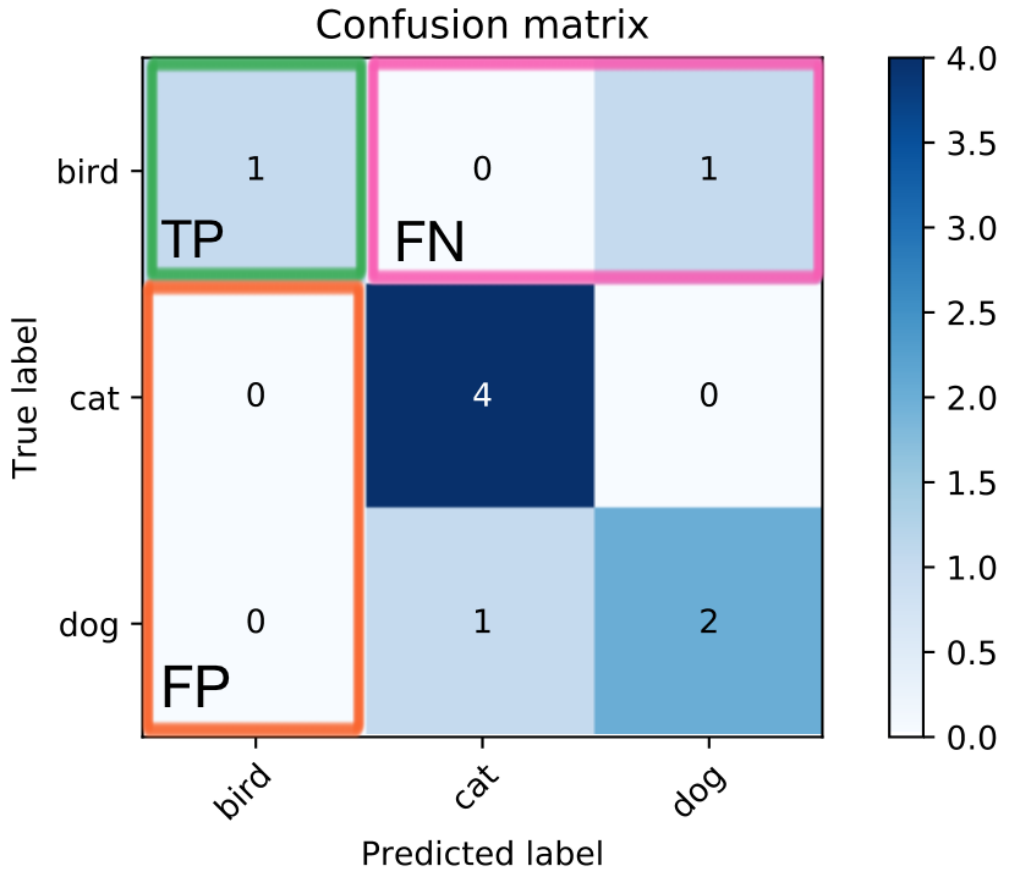


Concepts importants

Classification multilabel



Classification multilabel



Données Déséquilibrées

- une des deux modalités est fortement majoritaire

- Solutions:
  - utiliser des critères de performance adaptés au déséquilibre ;
  - ré-échantillonner les données pour se rapprocher d'une situation d'équilibre.

Critères de performance pour données déséquilibrées

- **balanced accuracy**

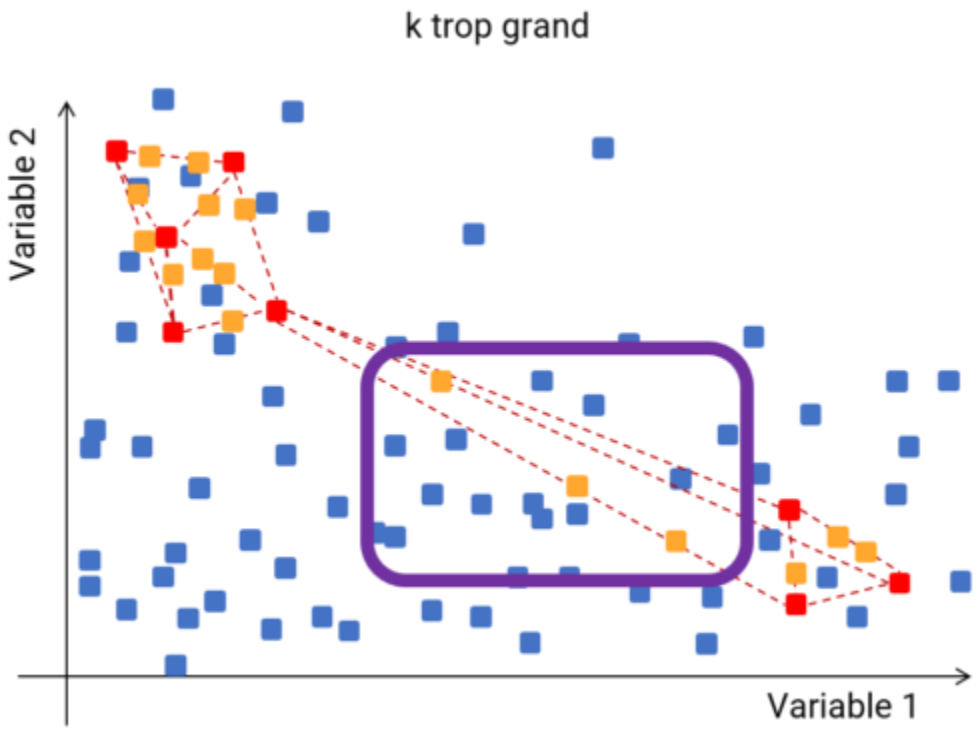
- **Micro-averaged**: all samples equally contribute to the final averaged metric
- **Macro-averaged**: all classes equally contribute to the final averaged metric
- **Weighted-averaged**: each classes's contribution to the average is weighted by its size

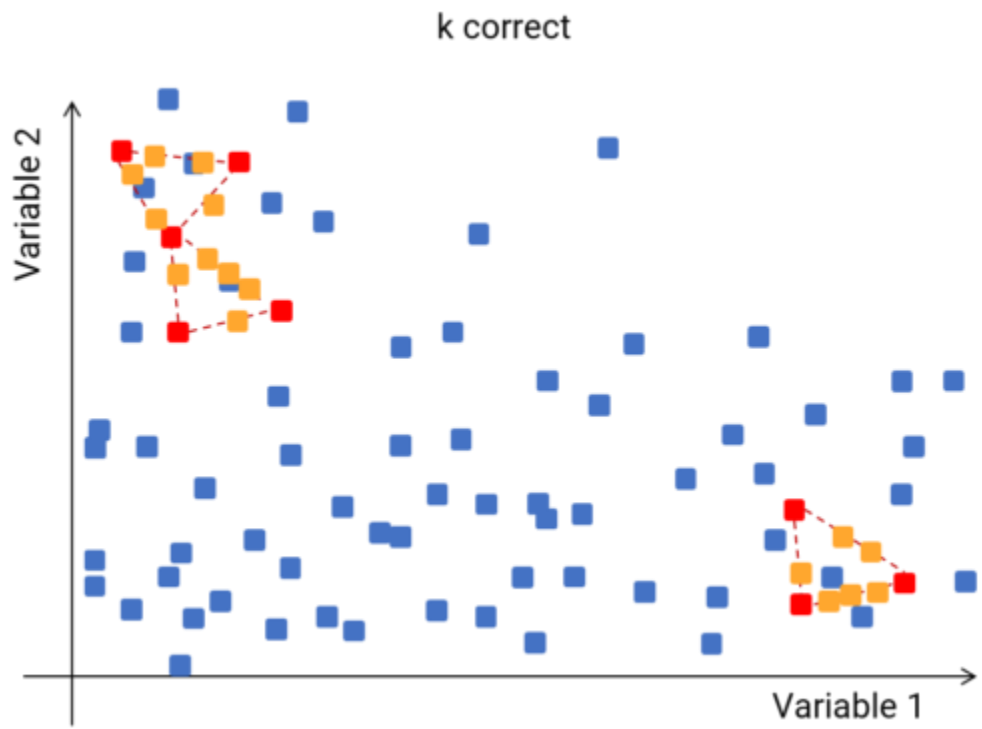
- **F1-score**

Ré-équilibrage

- **Oversampling**
- **Undersampling**

- **SMOTE :**
- **générer de nouveaux individus** minoritaires qui ressemblent aux autres, sans être strictement identiques


k trop grand
Variable 2
Variable 1


k correct
Variable 2
Variable 1

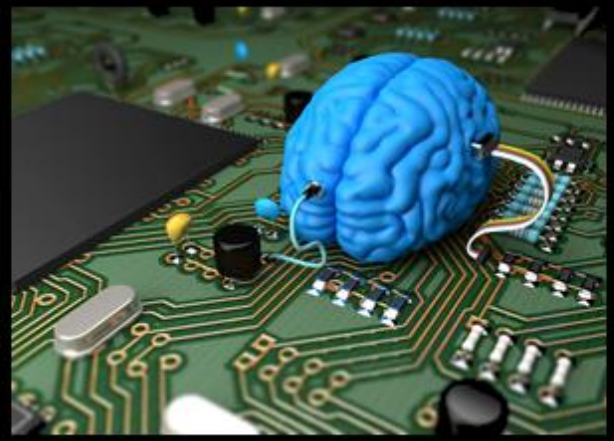
Importance des variables

- Certains modèle permet de récupérer l'importance des variables dans la prédiction
- Random Forest : Score d'impureté
- Importance par permutation: OOB de l'arbre VS OOB en permuttant les valeurs de la variables d'interêt

Deep Learning

Deep Learning

What society thinks I do



What my friends think I do

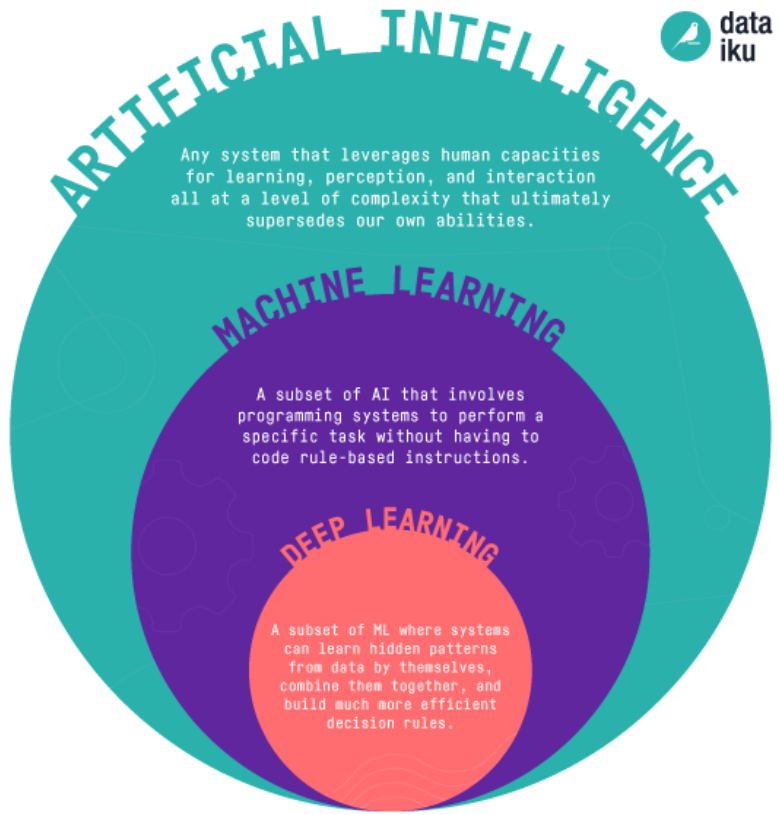
What other computer scientists think I do

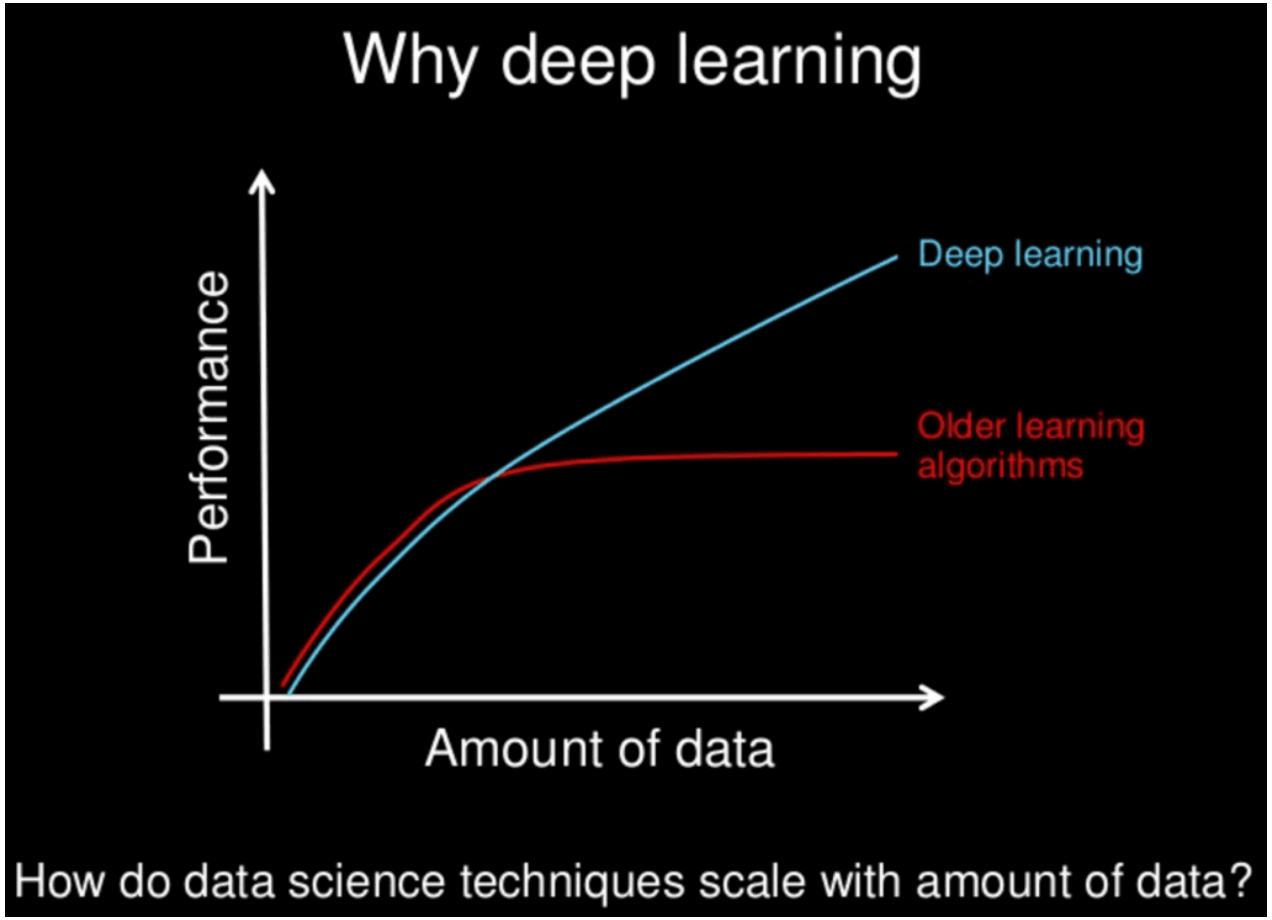
What mathematicians think I do

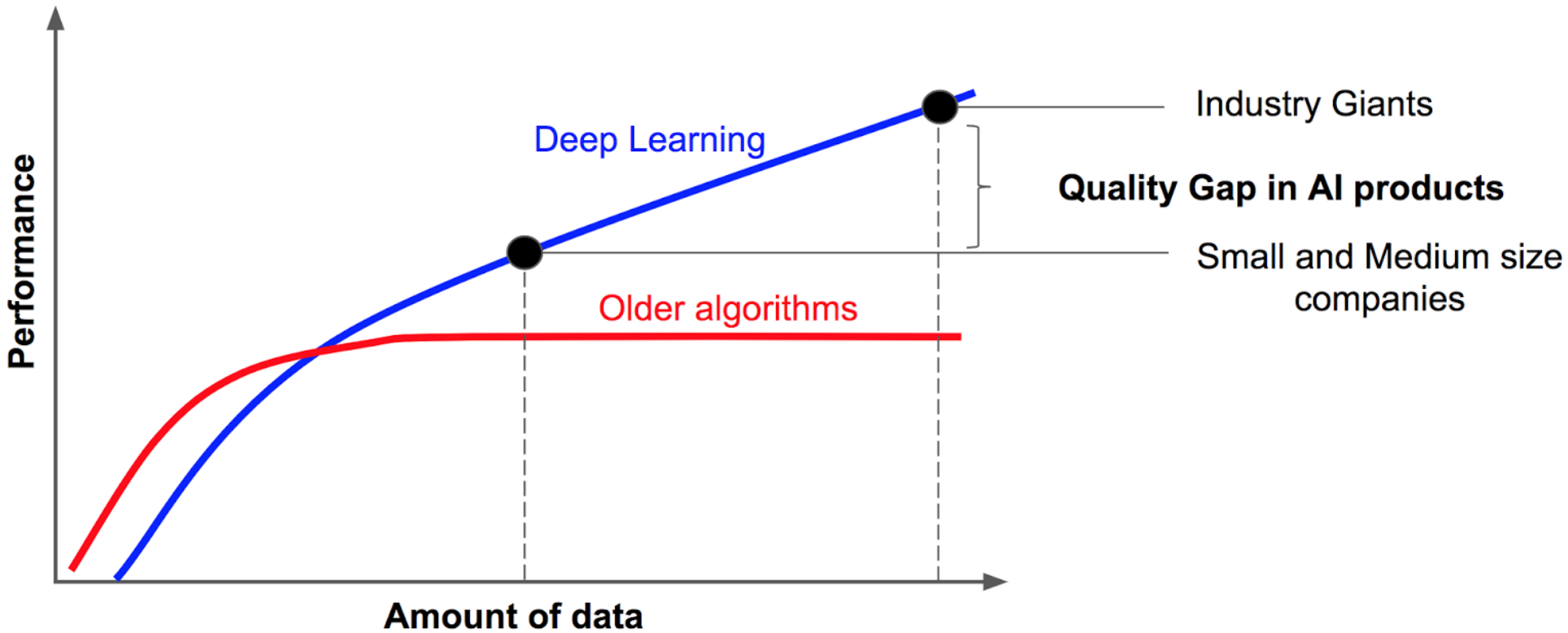
What I think I do

```
In [1]:
import keras
Using TensorFlow backend.
```

What I actually do


ARTIFICIAL INTELLIGENCE
Any system that leverages human capacities for learning, perception, and interaction all at a level of complexity that ultimately supersedes our own abilities.
MACHINE LEARNING
A subset of AI that involves programming systems to perform a specific task without having to code rule-based instructions.
DEEP LEARNING
A subset of ML where systems can learn hidden patterns from data by themselves, combine them together, and build much more efficient decision rules.


Why deep learning
Performance
Deep learning
Older learning algorithms
Amount of data
How do data science techniques scale with amount of data?

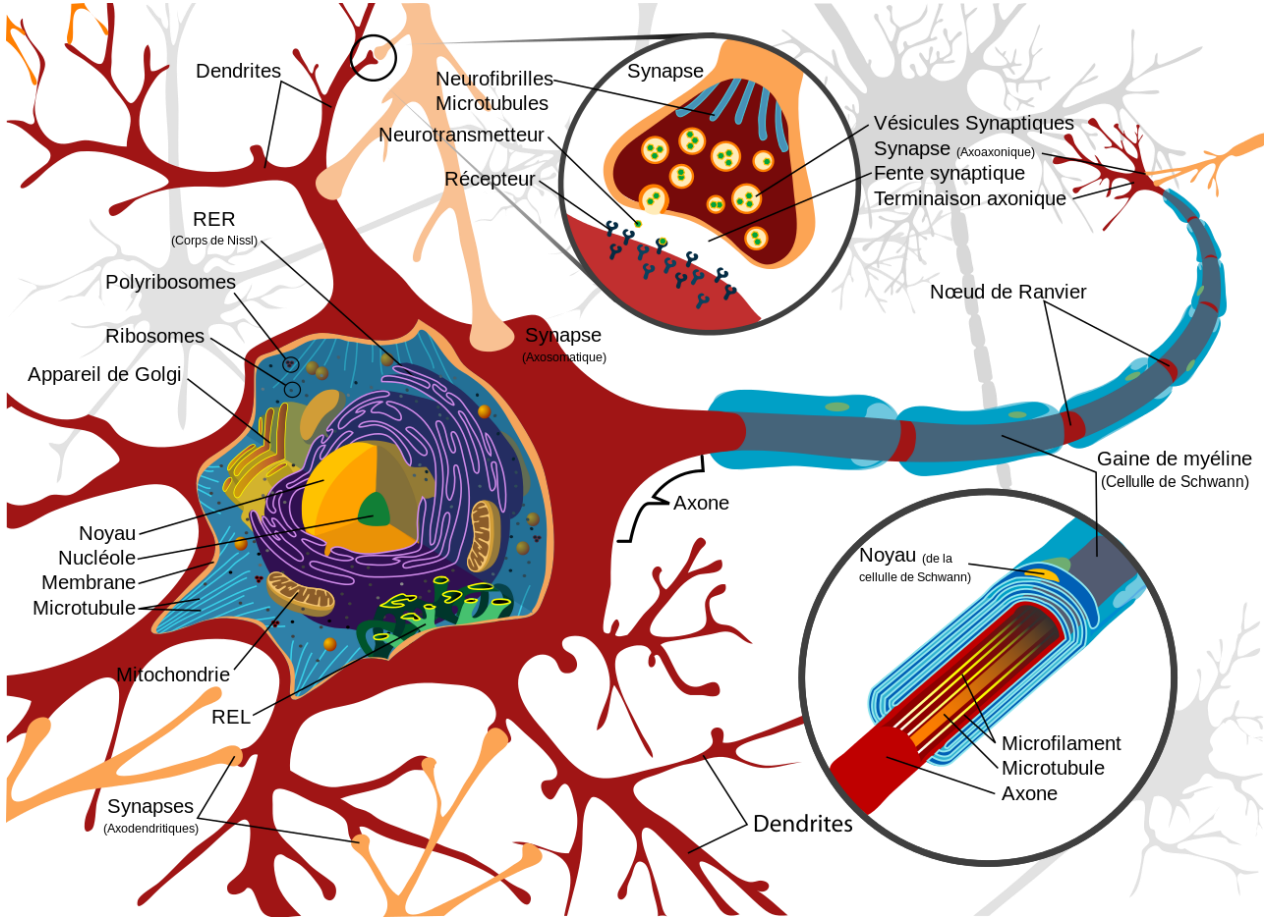

Performance
Deep Learning
Older algorithms
Industry Giants
Quality Gap in AI products
Small and Medium size companies
Amount of data

Classical Machine Learning:

- Interpretability and explainability are paramount
- Smaller amounts of relatively simple data
- Straightforward feature engineering
- Limited computational power
- Limited time, need for faster prototyping and operationalization
- Need for varied algorithm choices
- Accuracy of test dataset results is acceptable

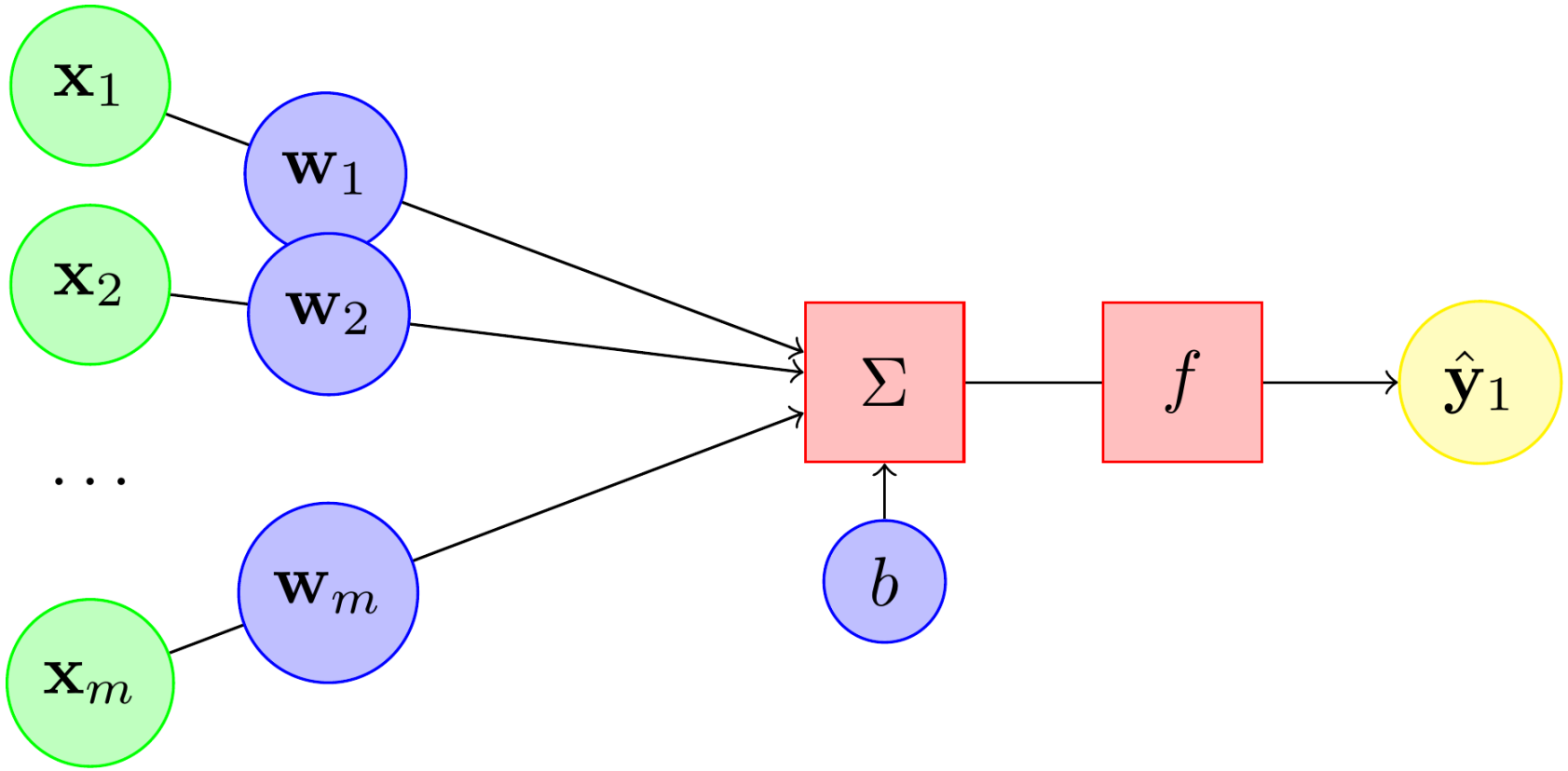
Deep Learning:

- Very high accuracy is a priority (and primes over straightforward interpretability and explainability)
- Large amounts of precisely labeled data
- Complex feature engineering
- Powerful compute resources available (GPU acceleration)
- Augmentation and other transformations of the initial dataset will be necessary


Dendrites
Neurofibrilles
Microtubules
Neurotransmetteur
Récepteur
Synapse
Vésicules Synaptiques
Synapse (Axoaxonique)
Fente synaptique
Terminaison axonique
RER
(Corps de Nissl)
Polyribosomes
Ribosomes
Appareil de Golgi
Synapse
(Axosomatique)
Nœud de Ranvier
Gaine de myéline
(Cellulle de Schwann)
Axone
Noyau
Nucléole
Membrane
Microtubule
Mitochondrie
REL
Noyau (de la cellulle de Schwann)
Microfilament
Microtubule
Axone
Synapses
(Axodendritiques)
Dendrites

Neurone formel (perceptron)

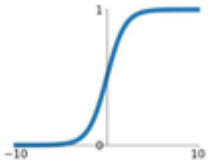
- Entrée **x**
- Sortie **y**
- Paramètres **w** et **b**
- $\hat{y} = f(\sum(w, x) + b)$



Fonctions d’activation

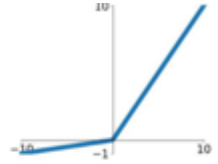
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



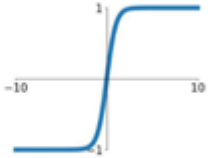
Leaky ReLU

$\max(0.1x, x)$



tanh

$\tanh(x)$

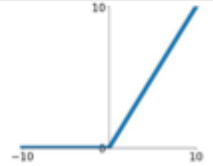


Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

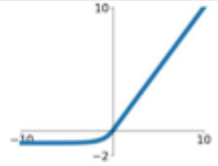
ReLU

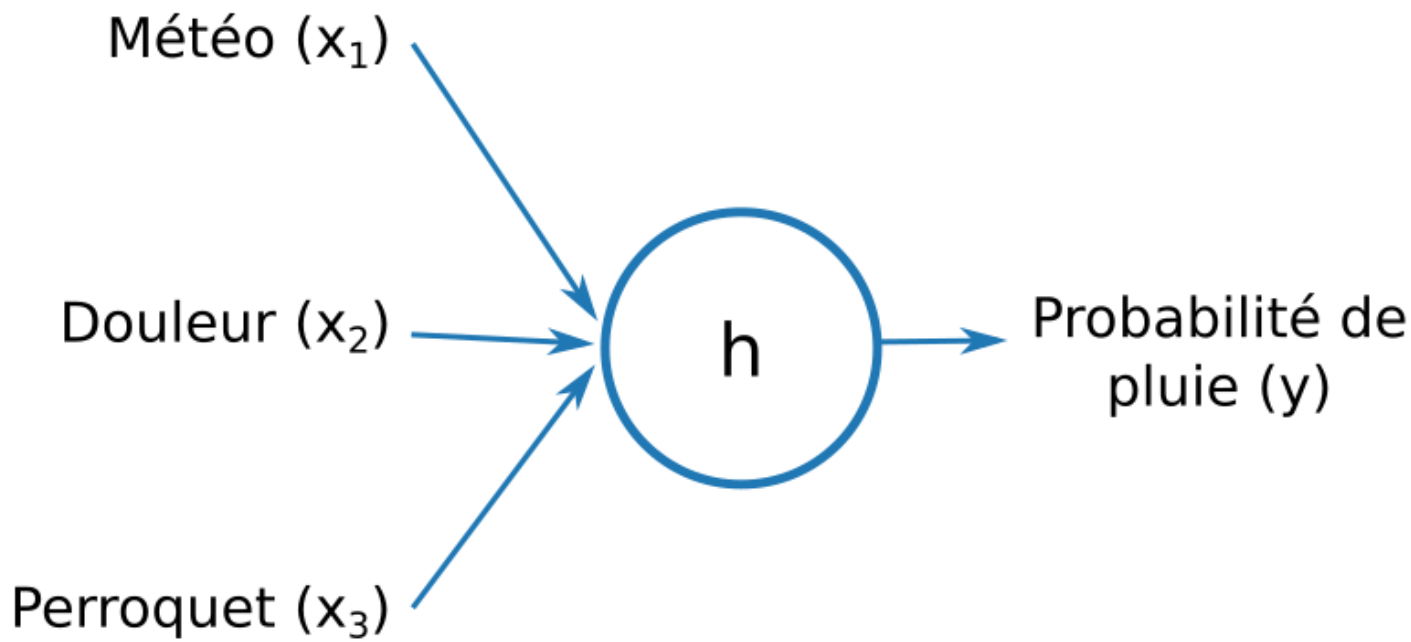
$\max(0, x)$



ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

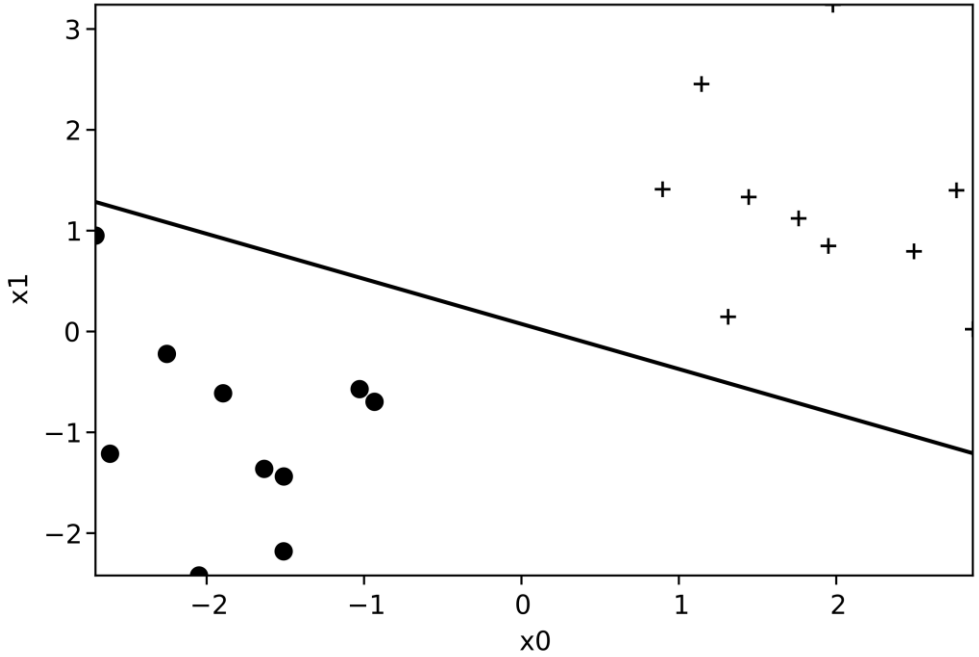



Météo (x_1)
Douleur (x_2)
Perroquet (x_3)
h
Probabilité de pluie (y)

Exemple pour une séparation linéaire

- W : 0,985|2,186
- B = -0,522

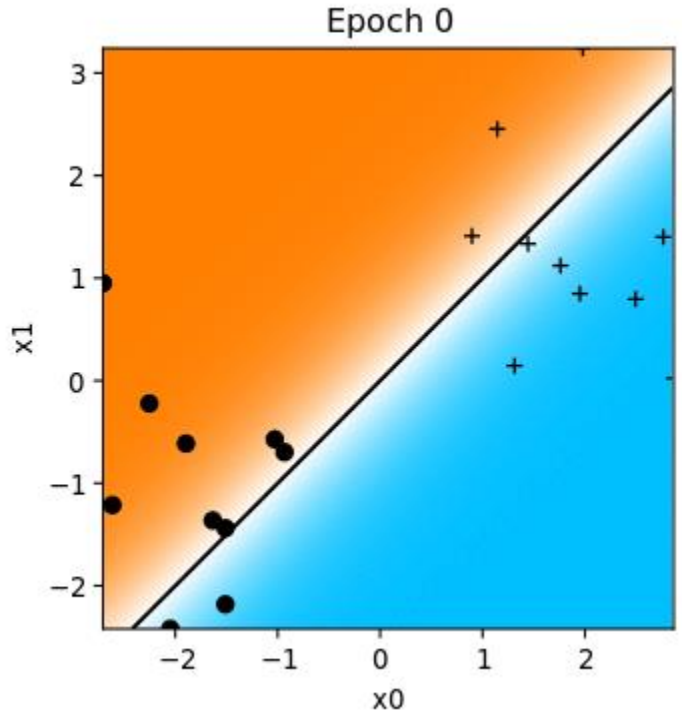
x0	x1	$\hat{y}$
0.896	1.410	3.445
-1.509	-1.438	-5.155
1.443	1.333	3.816
-1.895	-0.613	-3.731
-2.048	-2.420	-7.833



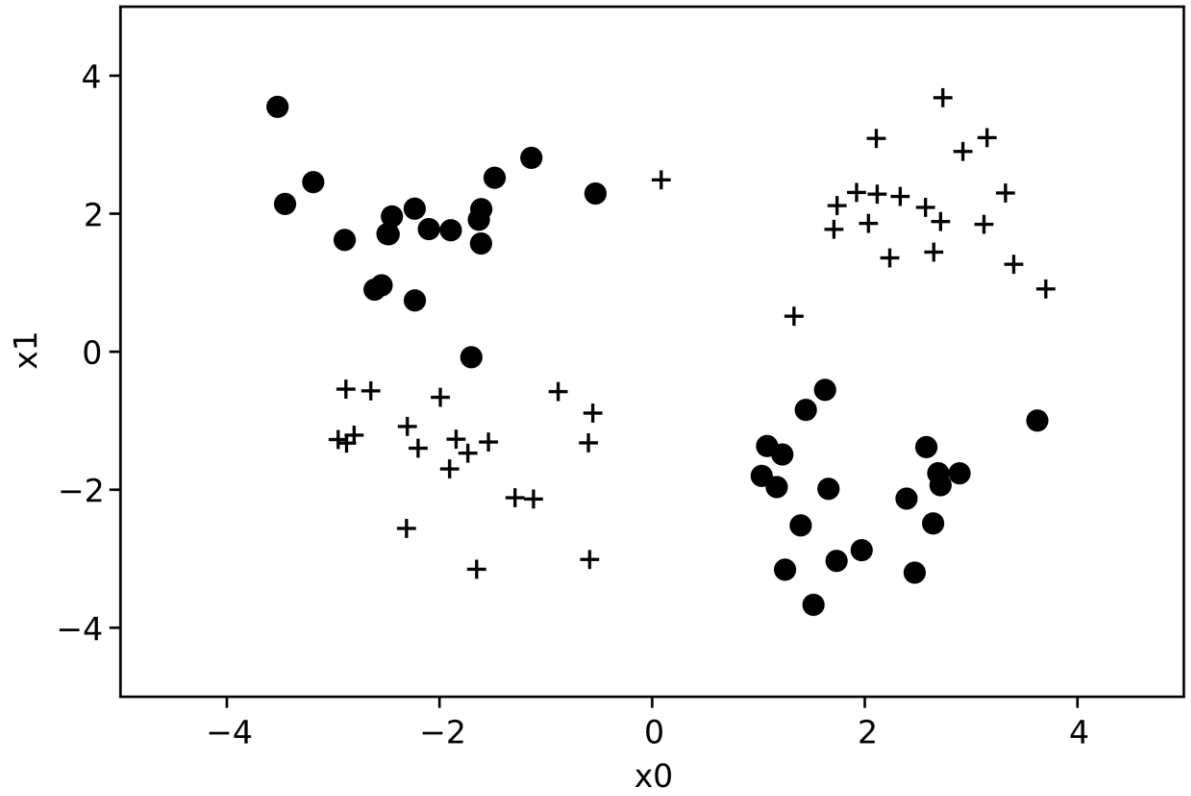
Apprentissage par descente de gradient

- Neuron apprend par optimisation d'une fonction de perte
- Exemple log-vraisemblance négative :

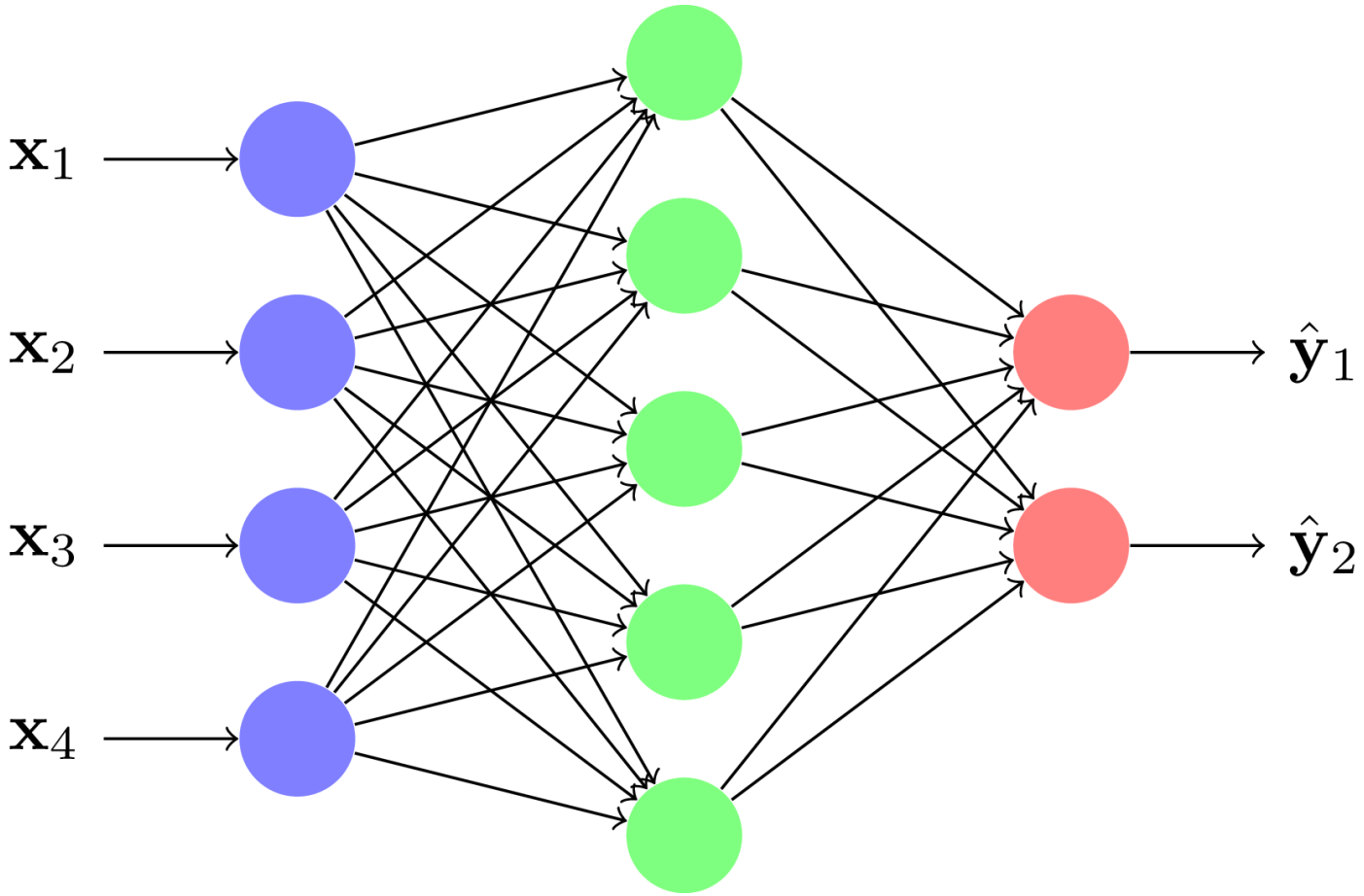
$$L(\hat{y}, y) = -(y.log(\hat{y}) + (1-y).log(1-\hat{y}))$$



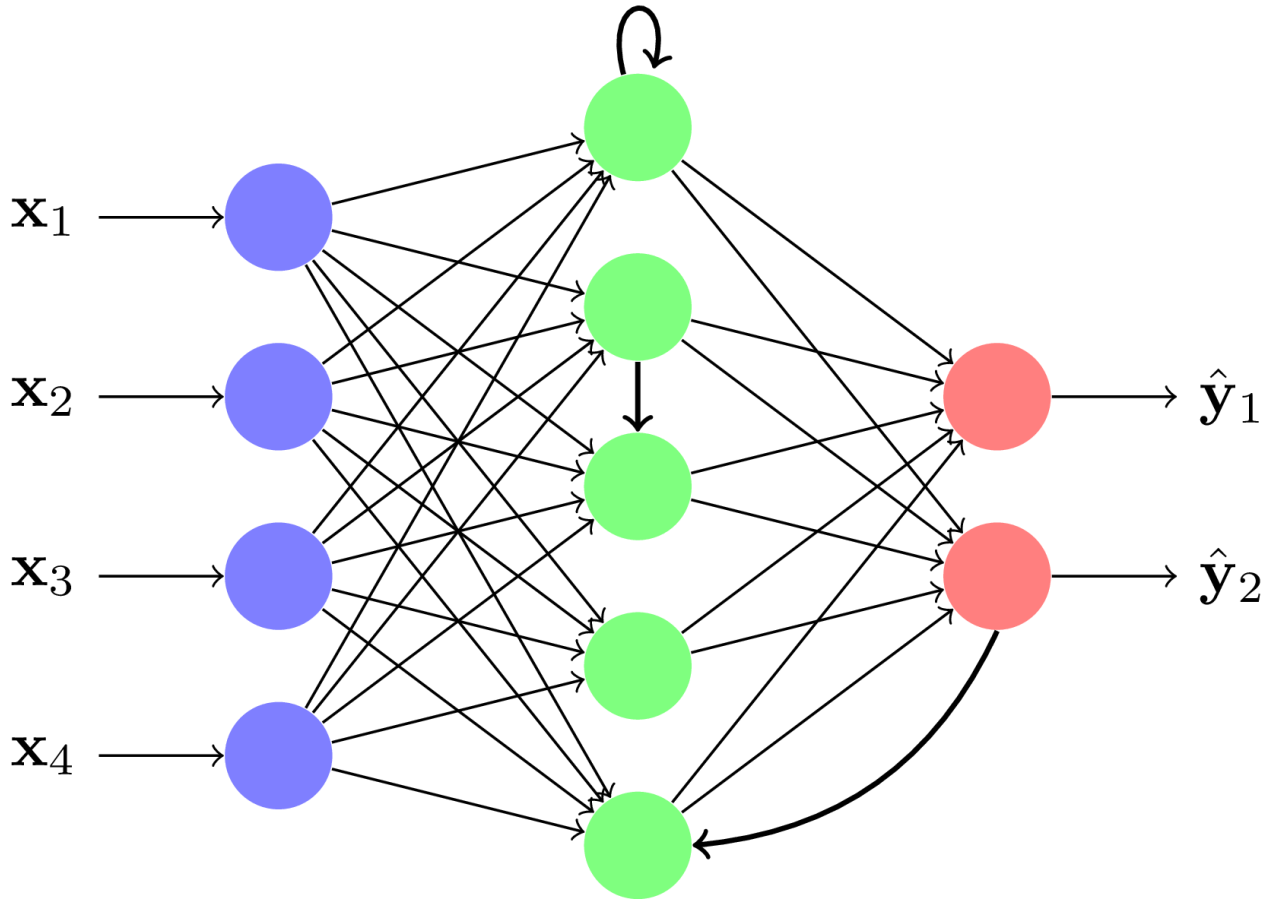
Perceptron ne suffit pas

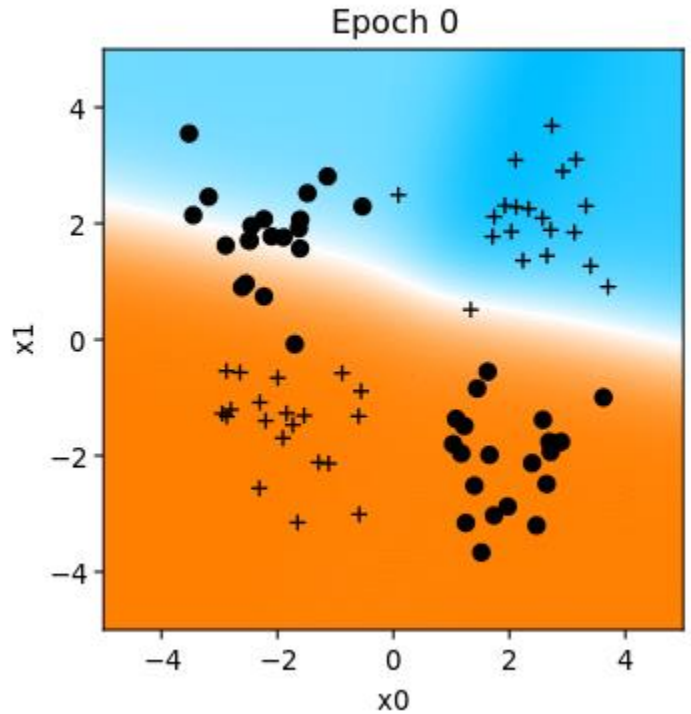


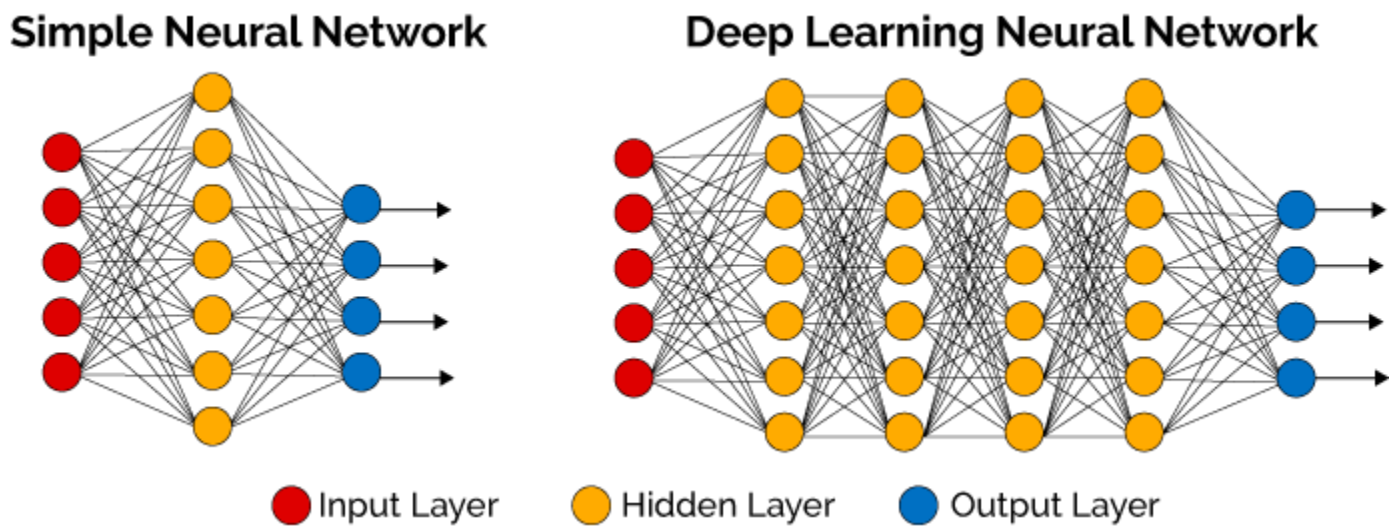
Réseau de neurone

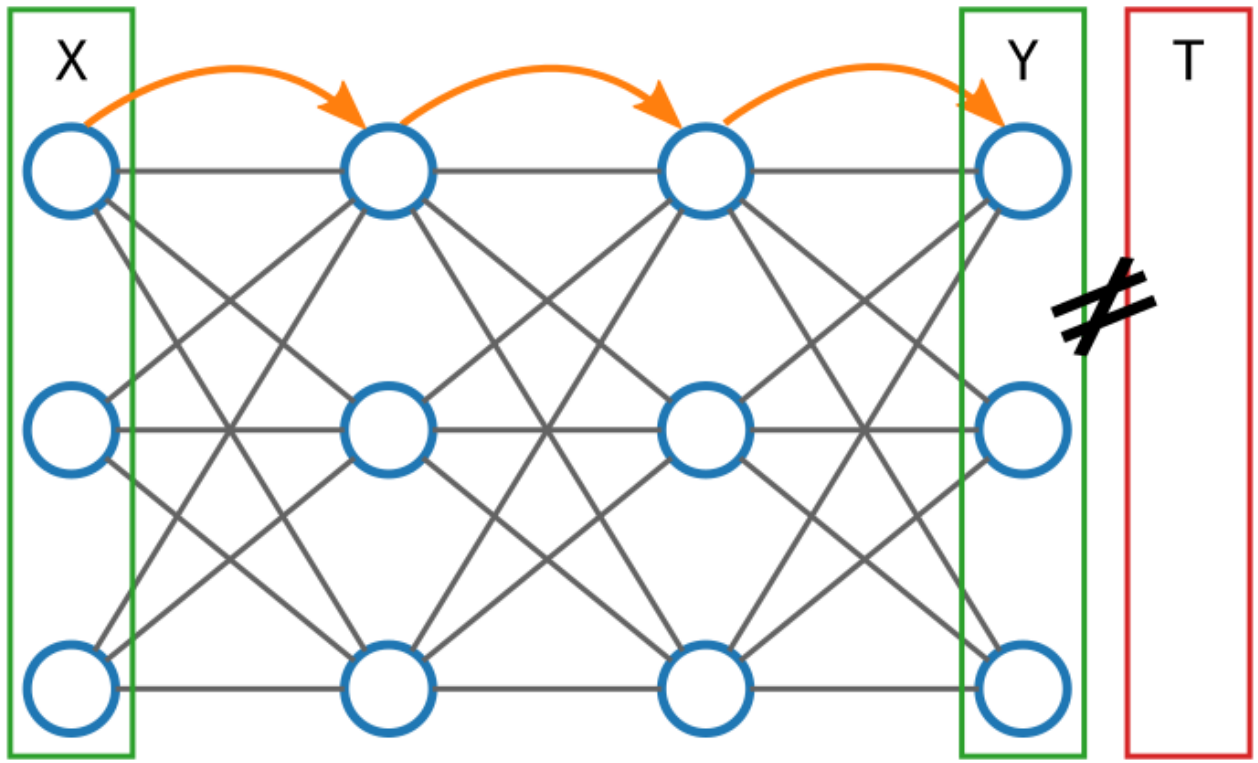


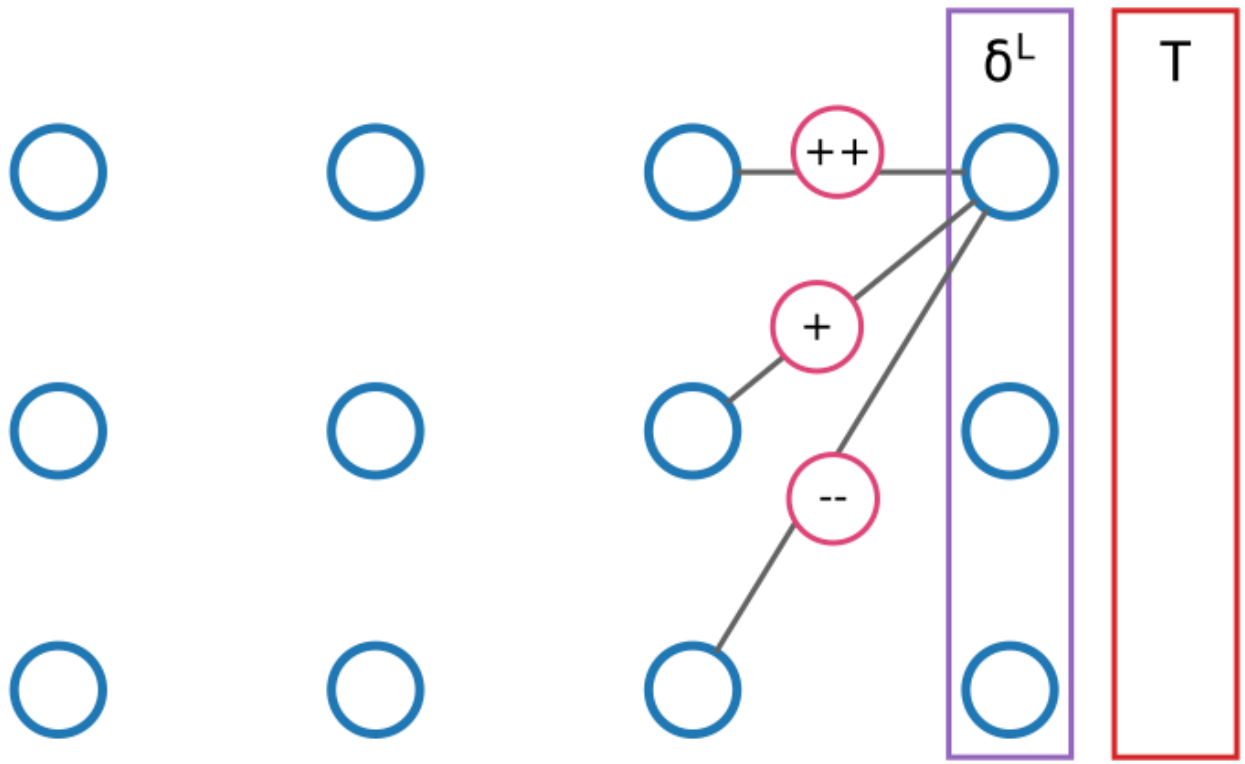
Réseau récurrent

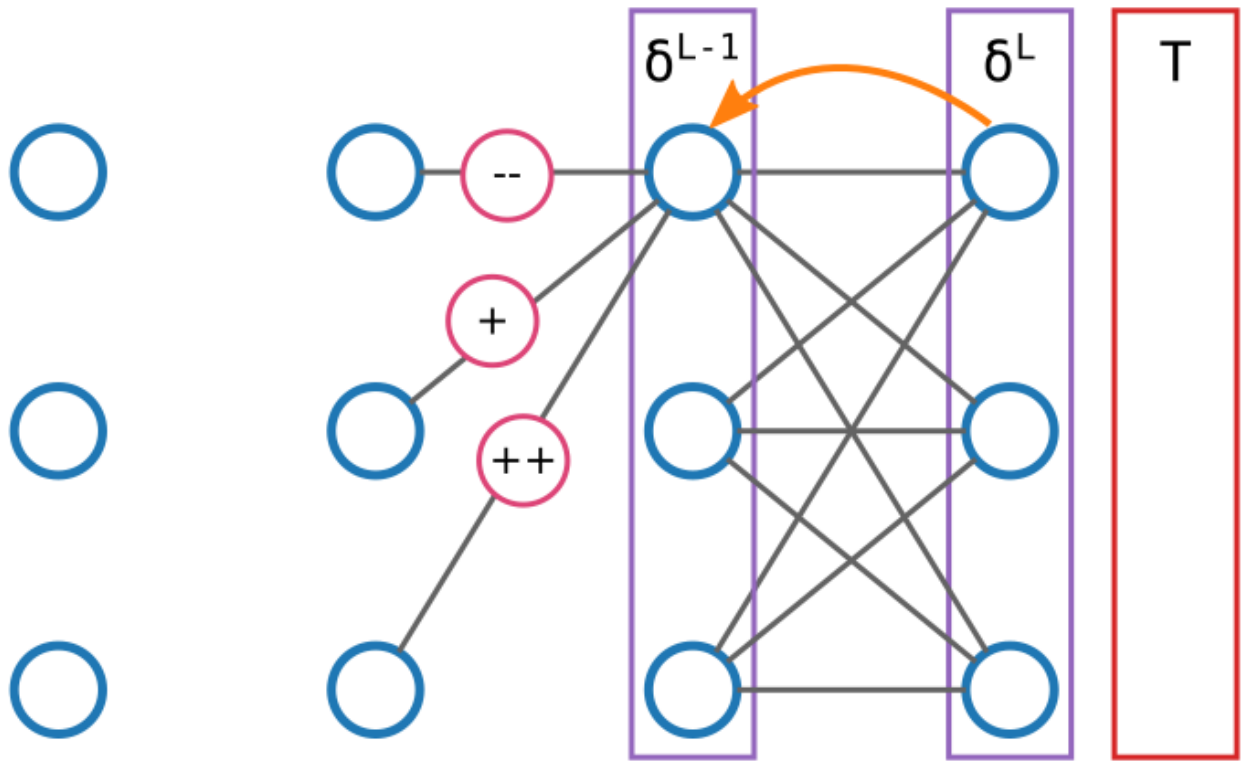


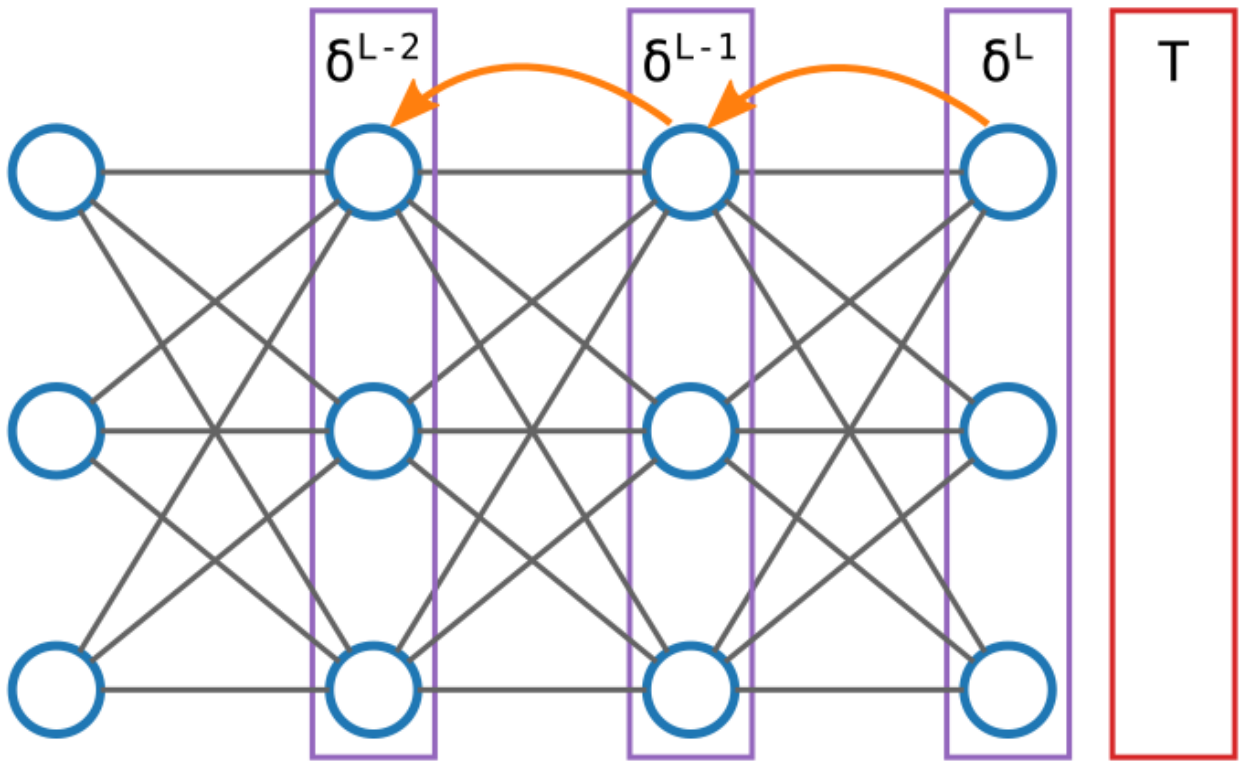

Epoch 0
4
2
0
−2
−4
x1
−4
−2
0
2
4
x0


Simple Neural Network
Deep Learning Neural Network
Input Layer
Hidden Layer
Output Layer


X
Y
T
≠


δ^L
T
++
+
--

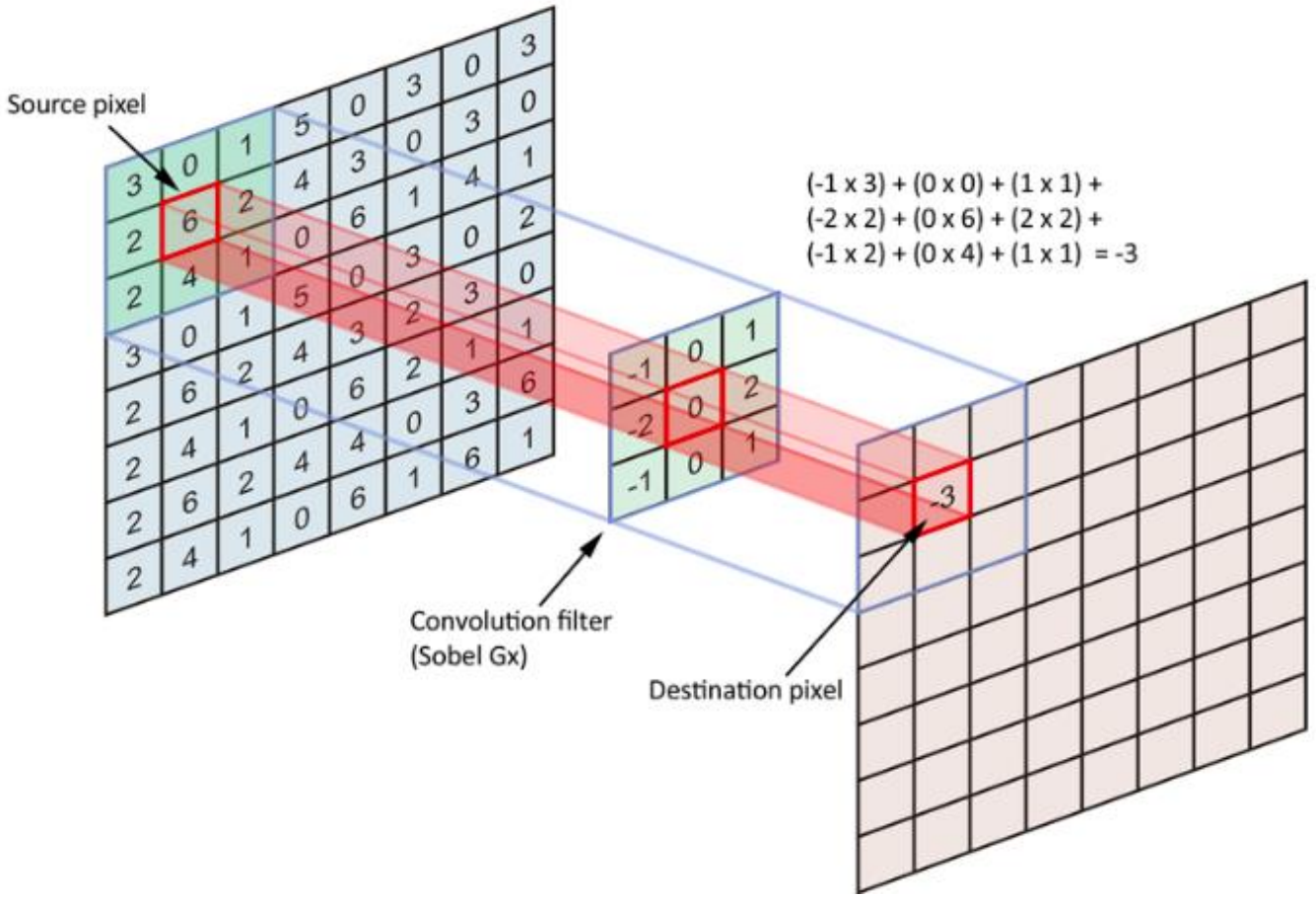

δ^L-1
δ^L
T
--
+
++

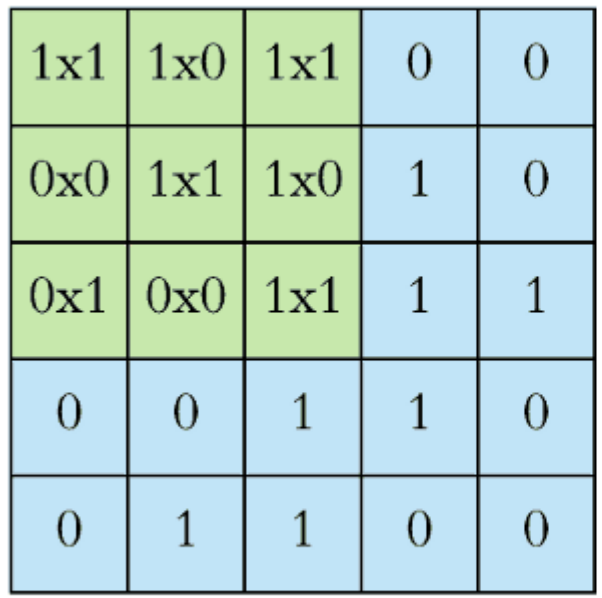

δ^{L-2}
δ^{L-1}
δ^{L}
T

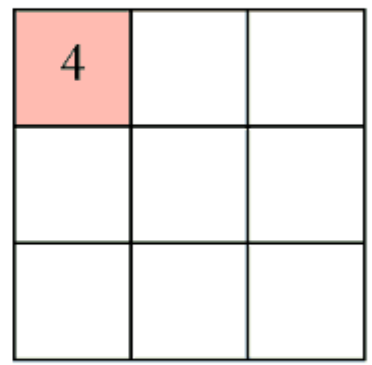
Interêt du deeplearning

- Reconnaissance d’image
- Traitement automatique de la langue
- Reconnaissance vocal

Convolution




1x1 1x0 1x1 0 0
0x0 1x1 1x0 1 0
0x1 0x0 1x1 1 1
0 0 1 1 0
0 1 1 0 0


4

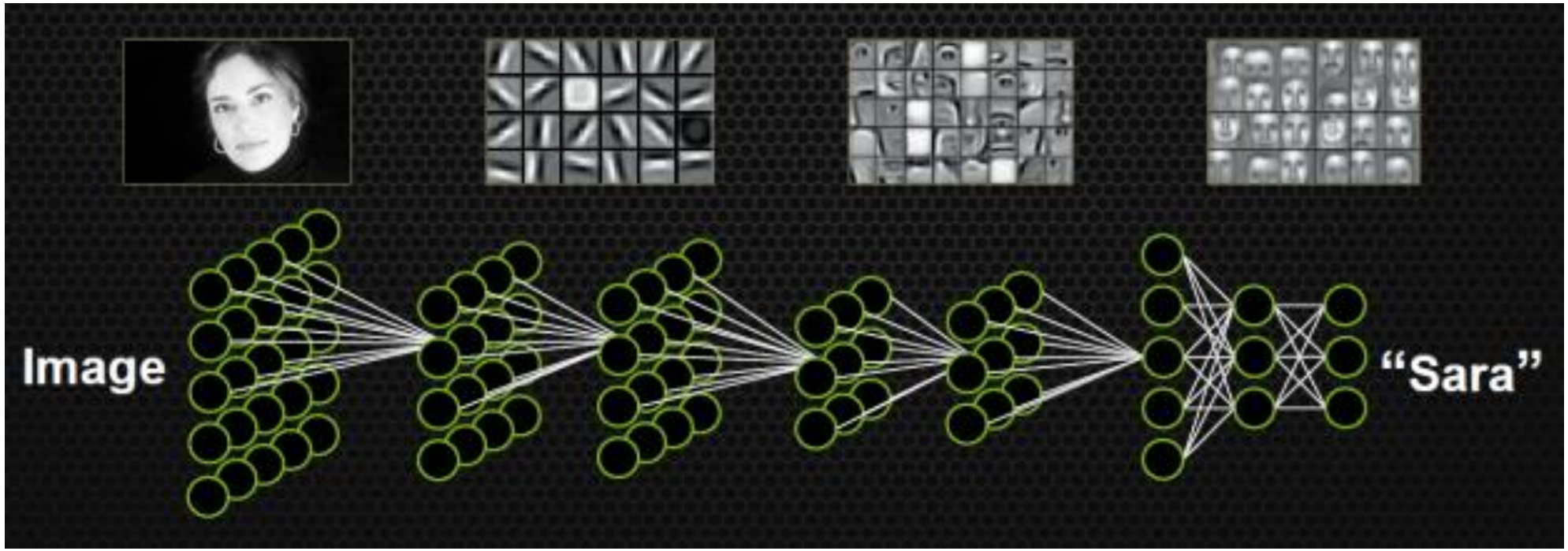
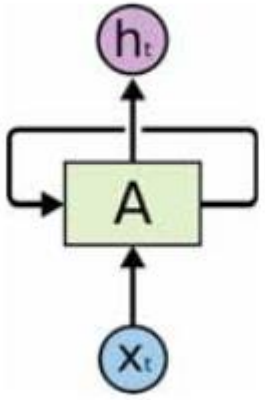
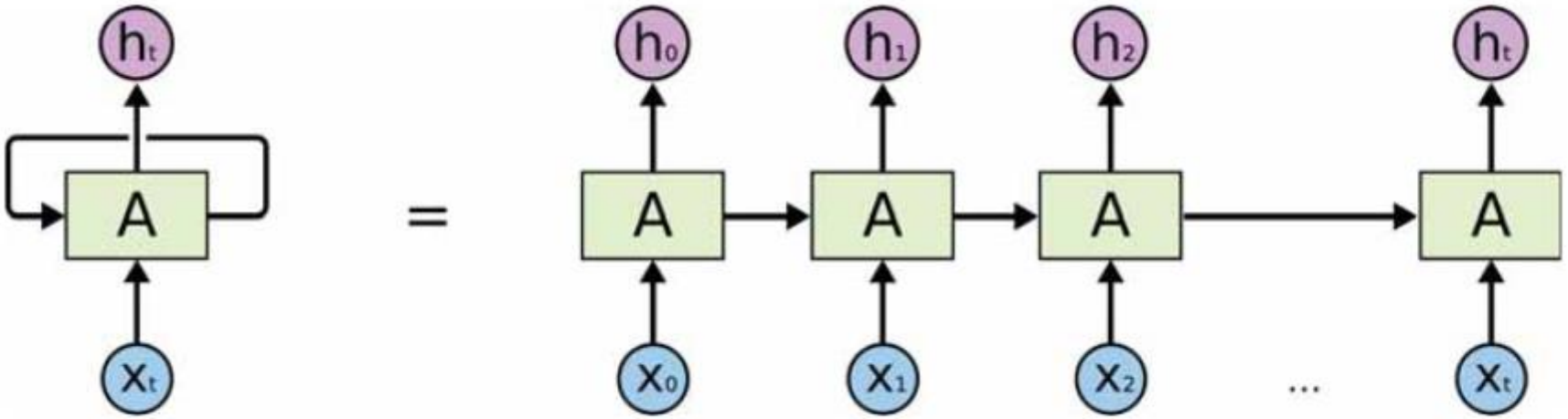


Image
"Sara"

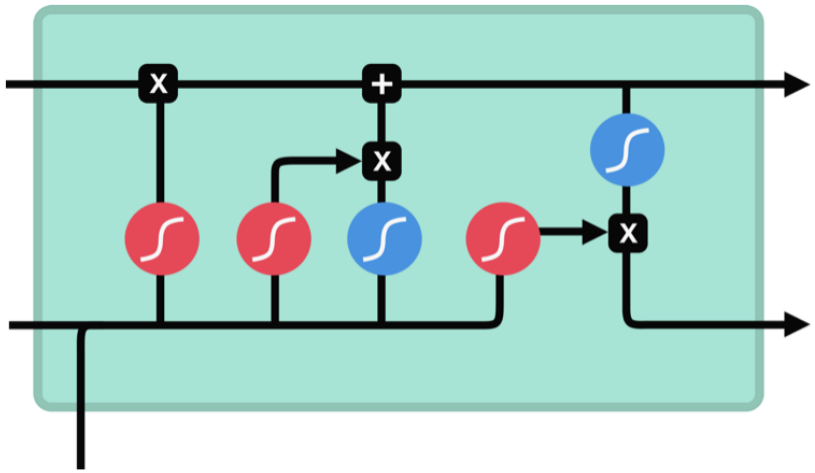
Réseau récurrent



Réseau récurrent



Réseau récurrent



Problématique du deeplearning

- Boite noire
- Nécessite beaucoup de données
- Répète le biais des données


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS