Deep Learning

Hatem & Driss by NEEDEMAND

Promotion DATA IA Microsoft 2018-2019 SIMPLON

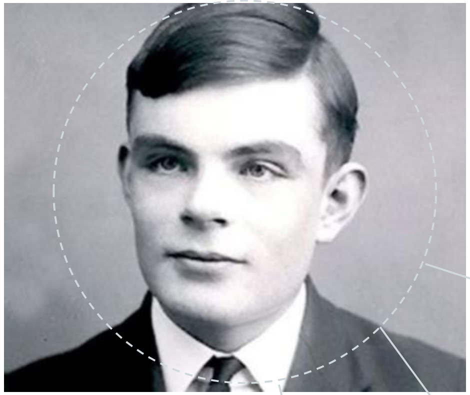
Petit récap Machine Learning

Supervisé ou Non supervisé

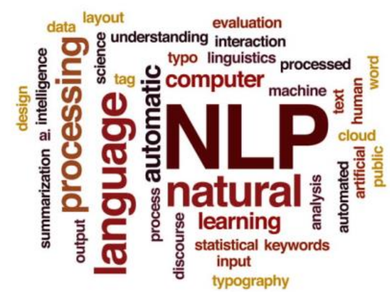
Régression	**Classification**
• Linéaire • Polynomiale • Simple • Multiple	• Régression logistique • Arbre à décision • SVM • Naïve Bayes • kNN • K-Means • Random Forest • Réduction de dimensions • Gradient Boosting
• Mean Absolute Error. • Mean Squared Error. • R^2.	• Précision (accuracy) • Courbe ROC • Matrice de confusion • precision, recall, f1-score et support

La petite histoire !

- 50s : machine Learning
- 43 : Mcculloch / Pitts
- 57 : Perceptron de Rosenblatt
- 86 : Multicouche MLP Rumelhart
- 2010s : Réseaux profonds
  Big data, Cloud, GPU…
- 2015 : Deepmind alphago



https://www.quantmetry.com/une-petite-histoire-du-machine-learning/


NLP
natural
language
processing
computer
learning
automatic
understanding
interaction
linguistics
evaluation
processed
machine
statistical keywords
input
typography
layout
data
typo
tag
science
intelligence
design
summarization
output
process
discourse
analysis
automated
artificial
public
cloud
text
human
word

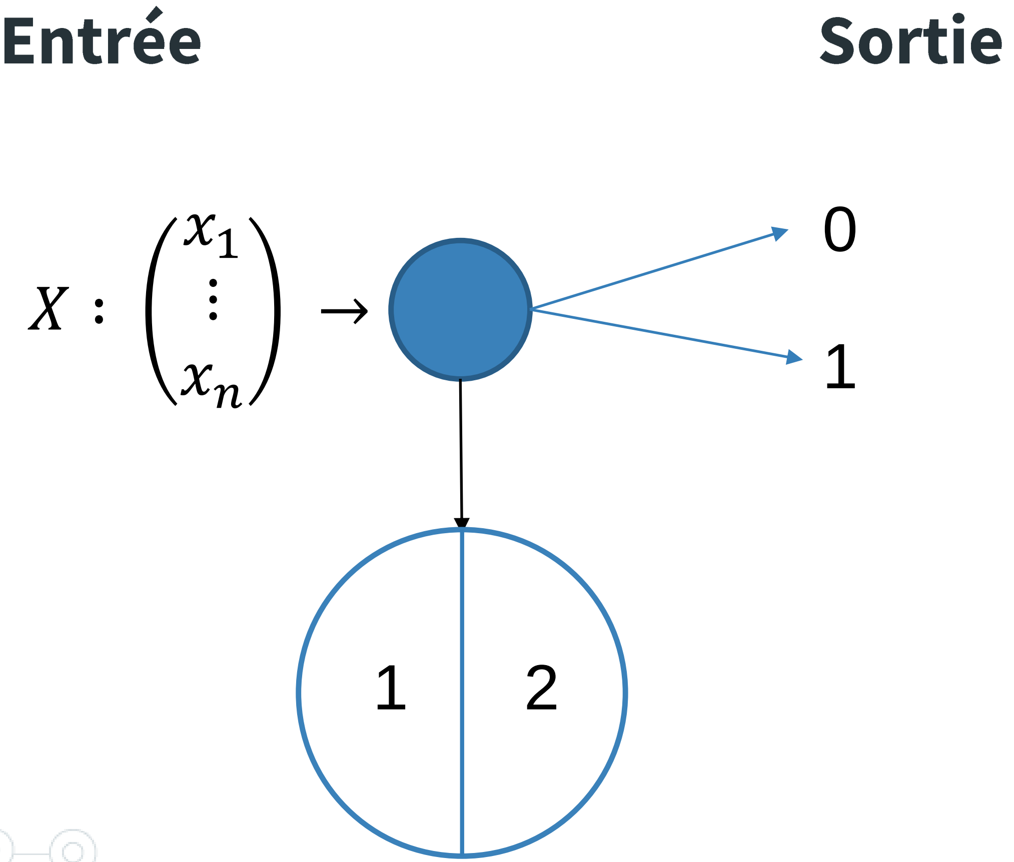

Self-Driving

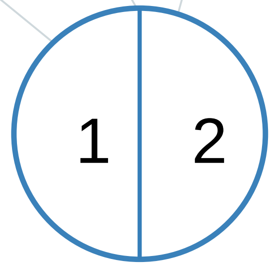
1.

Régression logistique

Un petit rappel

Classifieur binaire





1. Pré-activation

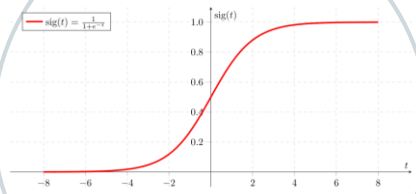
2. Activation

Combinaison linéaire

Fonction sigmoïde

$$X \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \quad z = Z(X) = w_1 x_1 + w_2 x_2 + b \longrightarrow a = A(z)$$

Sigmoïde



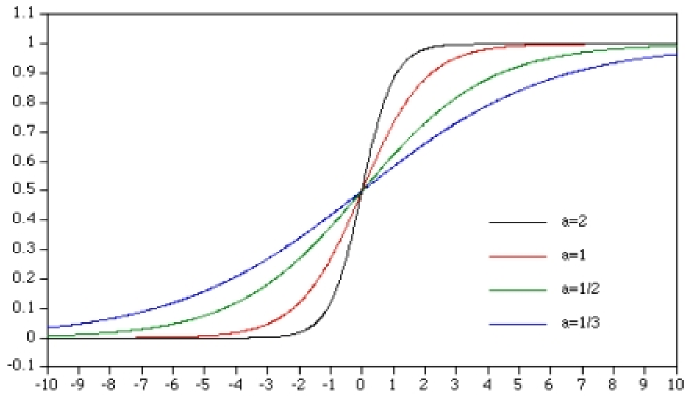
La plus célèbre :

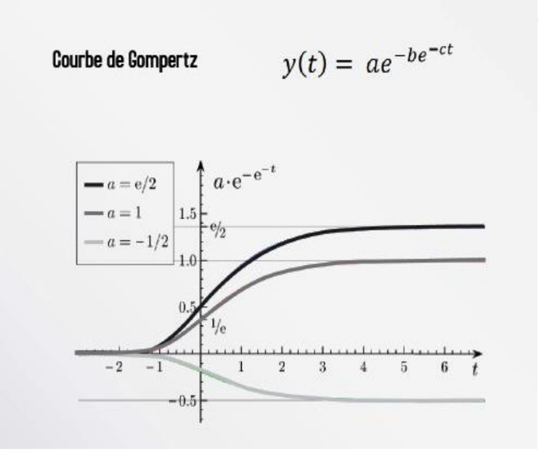
Fonction logistique

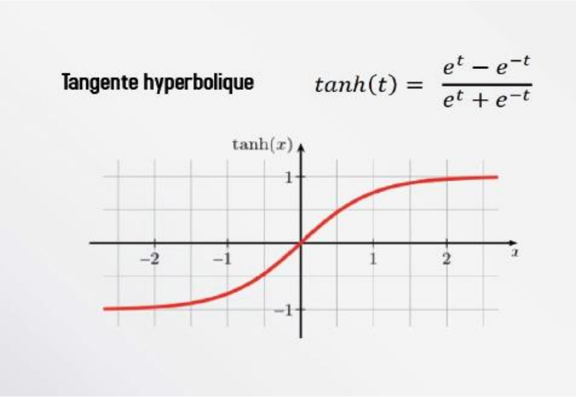
$$\frac{1}{1+e^{-x}}$$

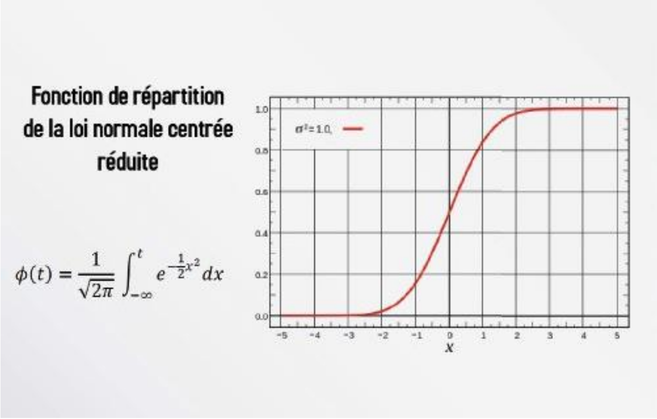
x	$-\infty$	$+\infty$
$\frac{1}{1+e^{-x}}$	0 ↗	1

Et d'autres


a=2
a=1
a=1/2
a=1/3


Courbe de Gompertz
$y(t) = ae^{-be^{-ct}}$
$a\cdot e^{-e^{-t}}$
$a = e/2$
$a = 1$
$a = -1/2$


Tangente hyperbolique
$tanh(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}}$
tanh(x)


Fonction de répartition
de la loi normale centrée
réduite
$\phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}x^2} dx$

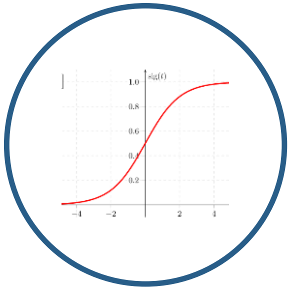
Training supervisé

Dataset

$$\{(X, Y)\}$$

Les données sont labellisées

$$\left(X_k \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, 1\right) \xrightarrow{\text{Forward}} \sum w_i x_i$$



Exemple :
w_i aléatoires

Forward

$$\sum w_i x_i = 0.3 \rightarrow a = 0 \qquad E = 0.7$$

Backward

Modification des Paramètres w_i

$Y_k = 0$
w_i modifiés

$$\sum w_i x_i = 0.2 \rightarrow a = 0 \qquad E = 0.2$$

Backward

Au clavier

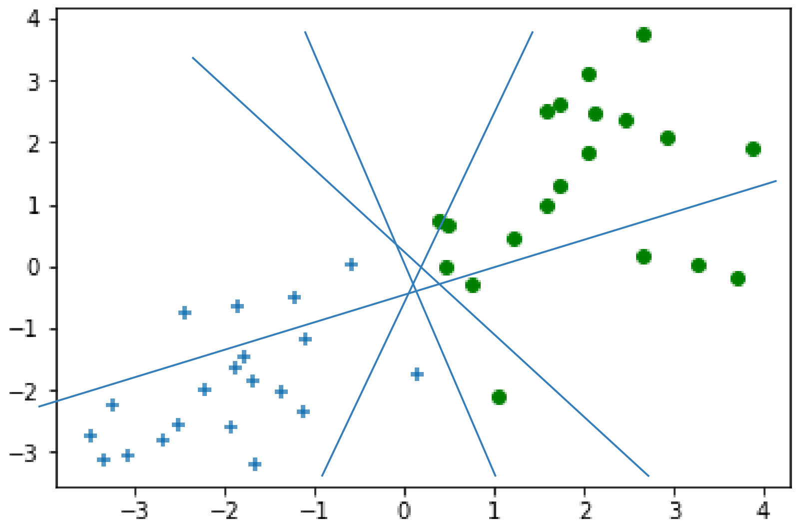
Code Forward

```
def init_weights():
    weights = np.random.normal(size=2)
    b = 0
    return weights, b
```

```
def pre_activation(features, weights, b):
    return np.dot(features, weights) + b
```

```
def activation(z):
    return 1 / (1 + np.exp(-z))
```

Frontière de décision



Forward

Backward

Erreur

Data : $\{(X, Y)\}$

Erreur

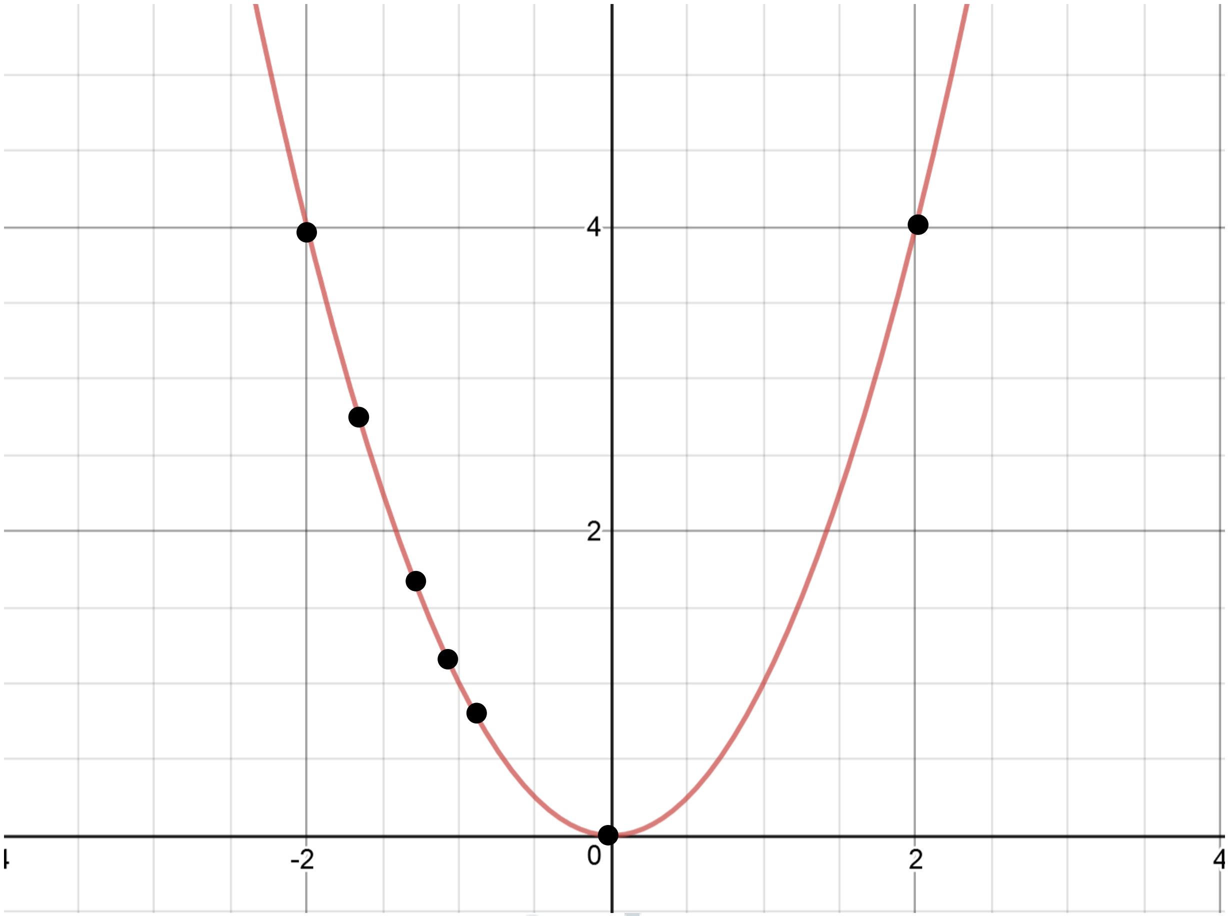
$$\boldsymbol{X} \rightarrow \boldsymbol{Z} \rightarrow \boldsymbol{A} \longrightarrow \boldsymbol{E} = \frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{A} - \boldsymbol{Y})^{\mathbf{2}}$$

E > 0
Maximise les grandes erreurs

Hatem & Driss by NEEDEMAND
Promotion DATA IA Microsoft 2018-2019 SIMPLON

Minimiser l'erreur : Descente de gradient

$$y = x^2 \qquad y' = 2x$$



$$x = -2 \ \rightarrow \ \nabla = -4$$

$$correction$$

$$x = x - \nabla = 2$$

$$Learning\ Rate$$

$$x = x \ - \eta\nabla$$

$$\eta = 0.1$$
$$x = -2 + 0.4 = -1.6$$

$$-1.6 \ \rightarrow -1.28$$

$$-1.28 \ \rightarrow -1.02$$

$$-1.02 \ \rightarrow -0.82$$

...

...

$$0 \ \rightarrow 0$$

Fonctions de plusieurs variables

Exemple :

$$f(x, y) = 5x^2 + 2xy + 6y^2$$

Fonctions de plusieurs variables

Exemple :

$$f(x, y) = 5x^2 + 2xy + 6y^2$$

$$\frac{\partial f}{\partial x} \qquad et \qquad \frac{\partial f}{\partial y}$$

Fonctions de plusieurs variables

Exemple :

$$f(x, y) = 5x^2 + 2xy + 6y^2$$

$$\frac{\partial f}{\partial x} = 10x + 2y \qquad et \qquad \frac{\partial f}{\partial y}$$

Fonctions de plusieurs variables

Exemple :

$$f(x, y) = 5x^2 + 2xy + 6y^2$$

$$\frac{\partial f}{\partial x} = 10x + 2y \qquad et \qquad \frac{\partial f}{\partial y} = 2x + 12y$$

Fonctions de plusieurs variables

Exemple :

$$f(x,y) = 5x^2 + 2xy + 6y^2$$

$$\frac{\partial f}{\partial x} = 10x + 2y \quad et \quad \frac{\partial f}{\partial y} = 2x + 12y$$

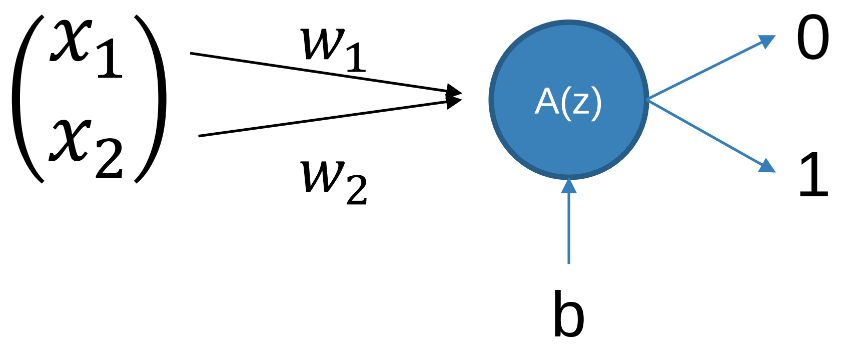
Descente de Gradient pour un couple (x_i, y_i) choisi aléatoirement.

$$x_i \leftarrow x_i - \eta \nabla_x(x_i) \quad et \quad y_i \leftarrow y_i - \eta \nabla_y(y_i)$$

```
# Fonction
def f(x,y) :
    return (5 * x**2) + (2*x * y) + (6 * y**2)
…
```

Au clavier

Application à la minimisation de l'erreur



- $E = \frac{1}{2}(a - y)^2$
- $a = A(z) = \frac{1}{1 - e^{-z}}$
- $z = Z(X) = w_1 x_1 + w_2 x_2 + b$

Calcul des dérivées partielles de l’erreur

- Règle de dérivation des fonctions composées :

$$\frac{dE}{dw} = \frac{dE}{da} \times \frac{da}{dz} \times \frac{dz}{dw}$$

$$\frac{dE}{da} = a$$

$$\frac{da}{dz} = a(1 - a)$$

$$\frac{dz}{dw} : \textit{dérivées partielles}$$

Au clavier

Thanks!

Any questions?

Hatem & Driss

By Needemand