

<http://licence-info.univ-lyon1.fr/LIFAMI>

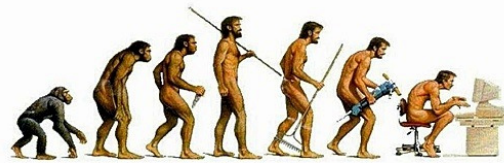


LIFAMI

APPLICATIONS EN MATH ET INFO

BIOLOGIE

Alexandre Meyer
 Equipe SAARA, Laboratoire LIRIS
 Université Lyon 1



1

Bio-informatique vs Informatique bioinspirée

- La bioinformatique est une science interdisciplinaire
 - Analyser l'information biologique disponible et produire de nouvelles connaissances
 - Proposer et développer des modèles, des méthodes et des outils
 - ➔ Il y a un Master BioInfo à Lyon 1 accessible depuis Linfo ou Lbio
- L'informatique bio-inspirée
 - Les scientifiques s'inspirent de la nature pour inventer ...



Nature

≠

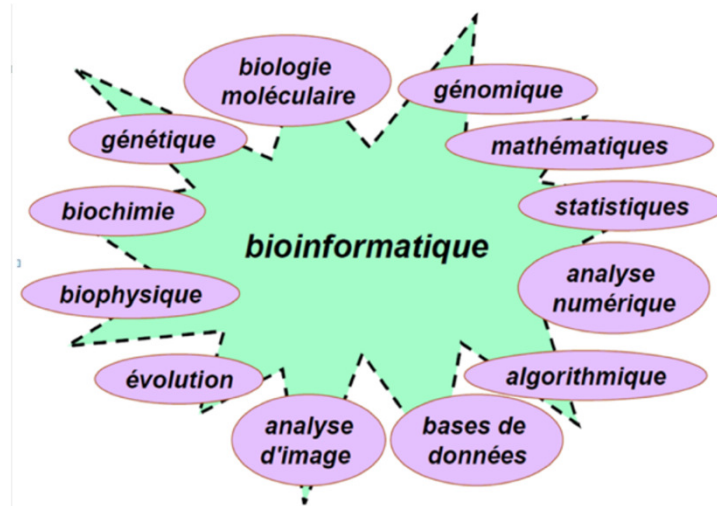
inspire l'



Artificiel

2

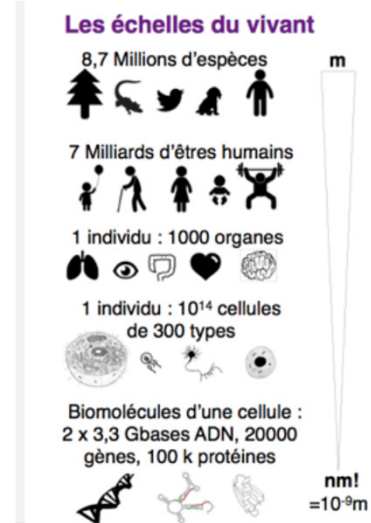
Bio-informatique



3

Bio-informatique

- Pourquoi de l'informatique pour la biologie?
 - Automatisation
 - Stockage
 - Recherche
 - Comparaison
 - Etc.
- La bio-informatique est devenue une partie importante de la biologie



4

Bio-informatique

- Formaliser des problèmes de biologie moléculaire
- Concevoir des solutions computationnelles à la portée des machines
- Développer et valider des outils
- Analyser, structurer, comparer, traiter les informations biologiques
- Stocker, accéder, filtrer ces informations
- Modéliser des processus biologiques
- Prédire des résultats biologiques

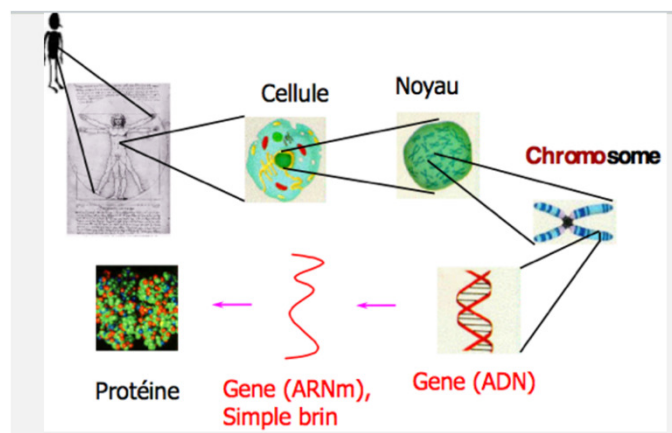
Et donc l'informatique

- Bases de données et de connaissances
- Algorithmique
- Théorie de graphes
- Intelligence artificielle: Data mining, machine learning, etc
- Statistique, analyse de données
- Visualisation de données
- Modélisation et simulation dynamiques
- Calcul parallèle

5

Bio-informatique

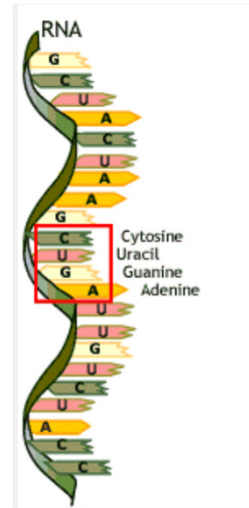
- Quels types d'informations ?



6

Bio-informatique

- Quels types d'informations ?
ADN et ARN



7

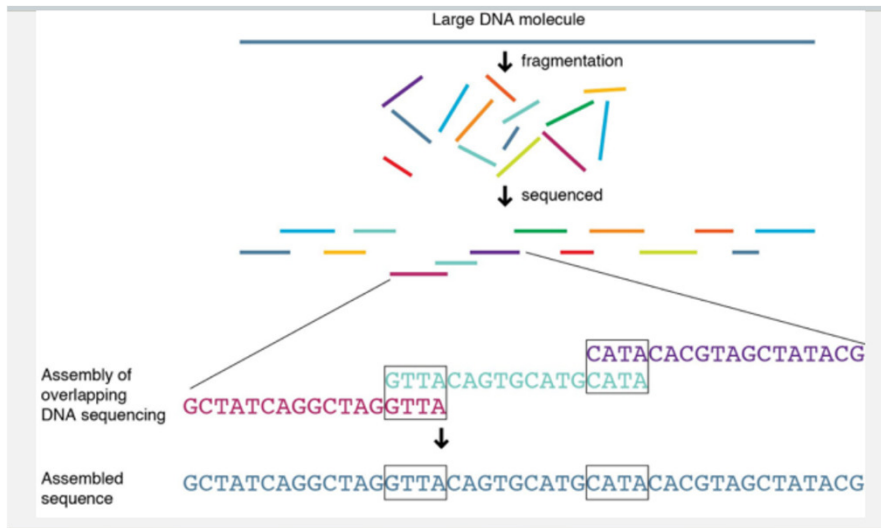
Séquençage ADN

- Homo sapiens – humain
 - 23 paires de chromosomes
 - génome : 3 milliards de bases
 - gènes: de 30000 à 35000
 - l'ADN contenu dans les chromosomes d'une cellule mis bout a bout (formerait un ruban d'1,50 mètre)
- Mus musculus – souris
 - 21 paires de chromosomes
 - génome : 3 milliards de bases
 - gènes: de 30000 à 35000
- Les virus, etc.



8

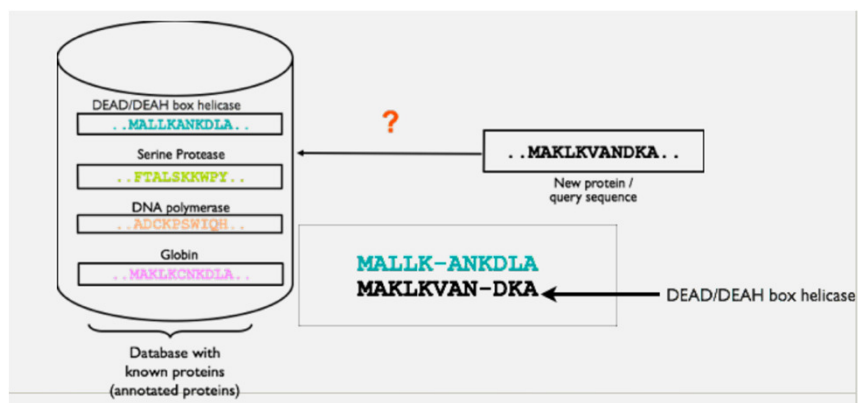
Séquençage ADN



9

Analyse des séquences protéines

- Etude de la fonction des protéines
 - Par comparaison à d'autres protéines



10

BIO-INFORMATIQUE

Modélisation de la dynamique des populations

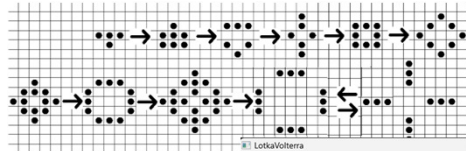
Evolution et propagation de virus, feu de forêts, espèces, eco-système, etc.

11

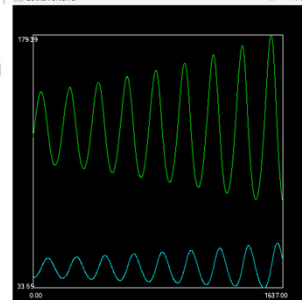
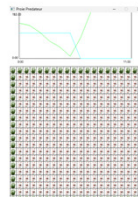
Modélisation en dynamique des populations

La modélisation des dynamiques des populations vise à expliquer, et éventuellement à prévoir, les évolutions d'une population dans un cadre écologique ou géographique donné

- Jeu de la vie (Conway)



- Équations de compétition de Lotka-Volterra
 - Equation différentielle
- Simulation plus complète



12

Etude du comportement de groupe

- Boids est le nom d'un programme informatique de vie artificielle, développé par Craig W. Reynolds en 1986, simulant le comportement d'une nuée d'oiseaux en vol.
- Le mot boid est par ailleurs une contraction de bird-oid (qui a la forme d'un oiseau)

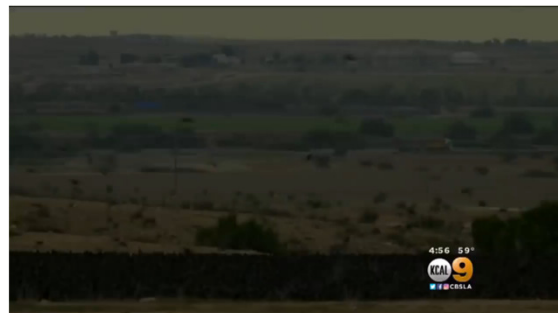


13

Etude du comportement de groupe

- Boids
 - **la cohésion** : pour former un groupe, les boids se rapprochent les uns des autres
 - **la séparation** : 2 boids ne peuvent pas se trouver au même endroit au même moment
 - **l'alignement** : pour rester groupés, les boids essayent de suivre un même chemin

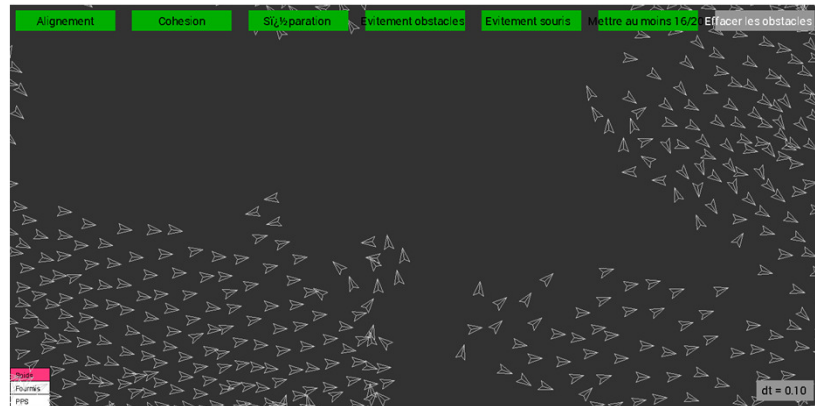
Faire varier :
zone de répulsion,
une zone d'orientation
et une zone d'attraction



14

Etude du comportement de groupe

- Démo d'un projet passé



15

Simuler un éco-système

- Simuler la concurrence des espèces d'arbres dans un environnement donnée



16

Simulation de feu de forêts

- Simulation et prédiction de l'évolution de feu de forêts
 - Fonctions des espèces
 - Du terrain
 - Du vent



17

Simulation de feu de forêts

- Simulation et prédiction de l'évolution de feu de forêts
 - À partir de règles simples du type jeu de ma vie peut faire un bon mini-projet

Fire in Paradise: Mesoscale Simulation of Wildfires
 T. Hadrlich¹, D. T. Banuti², W. Palubicki³, S. Pirk⁴, D. L. Michels⁵
¹KAUST, ²UNM, ³UAM, ⁴Google AI

We present a novel method for simulating wildfires with the goal to realistically capture the combustion process of individual trees and the resulting propagation of fires at the scale of forests.

18

INFORMATIQUE BIO-INSPIRÉ

Algorithmes évolutionnistes / Optimisation

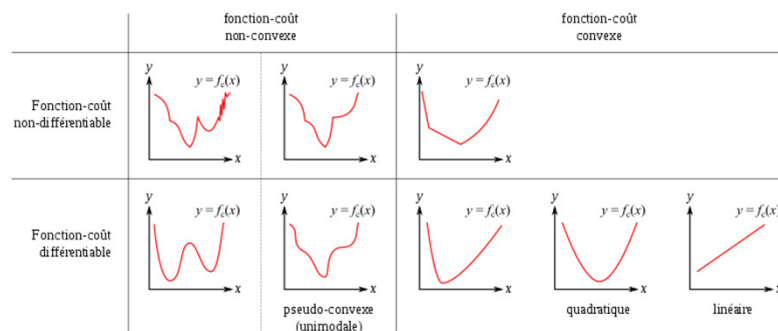
Colonie de fourmis

Réseaux de neurones

19

Qu'est-ce que l'optimisation ?

- Le but d'un problème d'optimisation est de **trouver une solution maximisant (resp. mini-misant) une fonction objectif donnée**
- En math, trouver le minimum d'une fonction



20

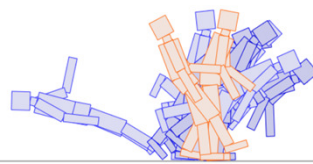
Qu'est-ce que l'optimisation ?

- Le but d'un problème d'optimisation est de **trouver une solution maximisant (resp. mini- misant) une fonction objectif donnée**
- Besoin d'algorithme car souvent itératif
 - il en existe de nombreux
- Exemples de problèmes
 - Maximiser le nombre de cours avec un minimum de salles
 - Apprendre à marcher sans tomber et avec le moins d'énergie
 - Plus court chemin
 - Voyageur de commerce
 - Ordonnancer des processus (aux urgences, système d'exploitation, ...
 - SUDOKU, Echec, etc.
 - ...

21

Algorithmes génétiques

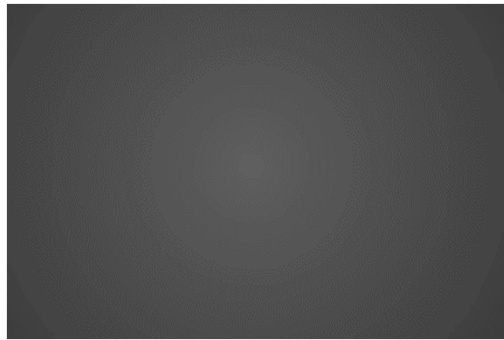
- Les algorithmes génétiques appartient a la famille des algorithmes évolutionnistes
- Inspiré de la biologie/Darwin: utilisent la notion de sélection naturelle sur une population de solutions
- Initialement introduit de manière formelle par John Holland [1975] puis vulgariser par Goldberg [1989].
- Voir l'exemple ici : https://rednuht.org/genetic_walkers/



22

Algorithmes génétiques

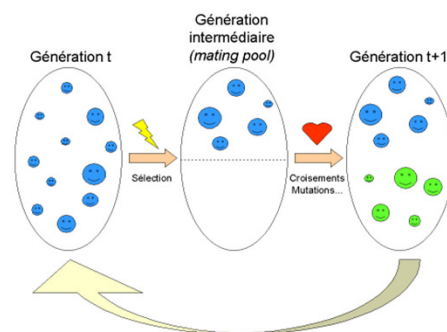
- Gènes \Leftrightarrow caractéristiques à optimiser
 - Dans l'exemple d'apprendre à marcher, les gènes codent les angles de chaque articulation pour une animation de N poses
 - Tableau de nombres



23

Algorithmes génétiques

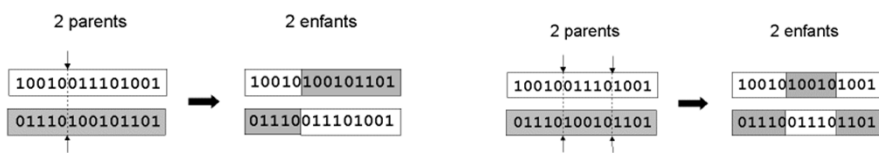
- Gènes \Leftrightarrow caractéristiques à optimiser
- Opérations répétées itérativement
 - Sélection des meilleurs
 - Génération T+1 avec
 - Croisement entre les gènes
 - Et mutation (perturbation des gènes)



24

Algorithmes génétiques

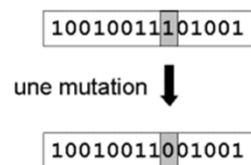
- Gènes \Leftrightarrow caractéristiques à optimiser
- Opérations répétées itérativement
 - Sélection
 - **Croisement (entre les gènes)**
 - Mutation : perturbation des gènes



25

Algorithmes génétiques

- Gènes \Leftrightarrow caractéristiques à optimiser
- Opérations répétées itérativement
 - Sélection
 - Croisement (entre les gènes)
 - **Mutation : perturbation des gènes**



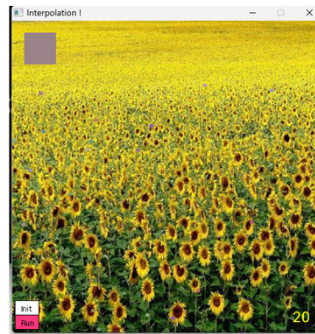
Ou choisir un nombre au hasard dans la liste des gènes et le perturber entre + ou - 10%

26

Algorithmes génétiques

- TD : Optimiser la couleur des insectes dans leur environnements
 - Une population a des couleurs tirées au hasard
 - On sélectionne ceux qui sont mangés par la souris en dernier
 - La nouvelle population utilise la couleur de la génération précédente, avec des variations

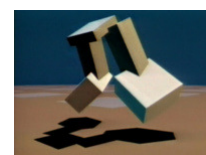
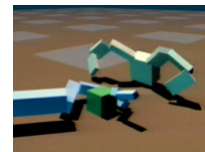
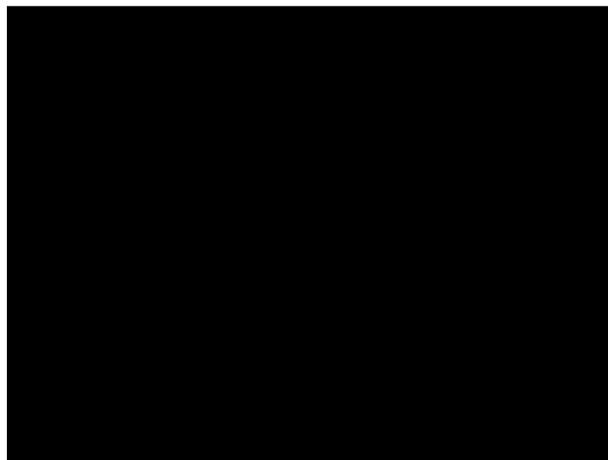
Démo du TD couleur des insectes



27

Algorithmes génétiques

- Karl Sims : **Evolved Virtual Creatures (1994)**



28

Algorithmes génétiques

Inconvénients

- Nécessite beaucoup de temps de calcul
- Souvent difficiles à mettre en œuvre
- Impossible d'assurer que la solution trouvée est la meilleure
- Problème de convergence vers un optimum local, si celui-ci est le plus majoritaire

→ mais ils sont FUN !!! et facile à comprendre ...

29

INFORMATIQUE BIO-INSPIRÉ

Algorithmes évolutionnistes / Optimisation

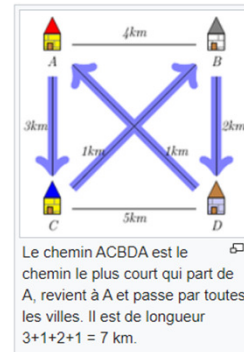
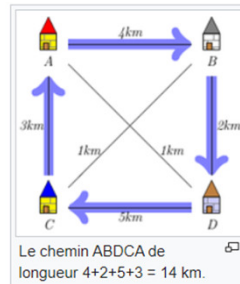
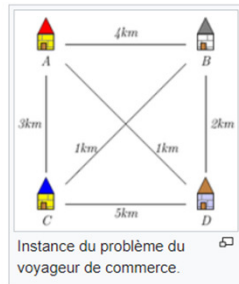
Colonie de fourmis

Réseaux de neurones

30

Problème du voyageur de commerce

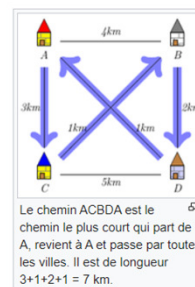
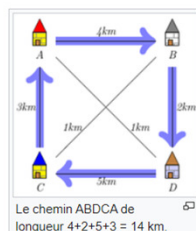
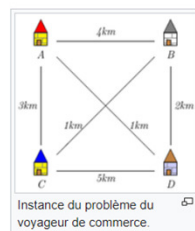
- Plus court chemin / voyageur de commerce
 - N villes avec des distances entre chaque ville
 - Visiter toutes les villes avec un minimum de distance



31

Problème du voyageur de commerce

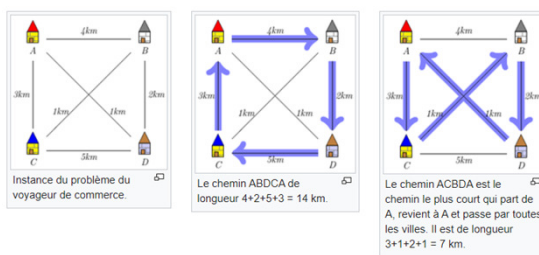
- Plus court chemin / voyageur de commerce
 - Ce problème est plus compliqué qu'il n'y paraît ;
 - on ne connaît pas de méthode de résolution permettant d'obtenir des solutions exactes en un temps raisonnable pour de grandes instances (grand nombre de villes) du problème.



32

Problème du voyageur de commerce

- Plus court chemin / voyageur de commerce
 - Ce problème est plus compliqué qu'il n'y paraît ;
 - Pour ces grandes instances, on devra donc souvent se contenter de solutions *approchées*, car on se retrouve face à une [explosion combinatoire](#).



Nombre de villes n	Nombre de chemins candidats $\frac{1}{2}(n-1)!$
3	1
4	3
5	12
6	60
7	360
8	2 520
9	20 160
10	181 440
15	43 589 145 600
20	$6,082 \times 10^{16}$
71	$5,989 \times 10^{99}$

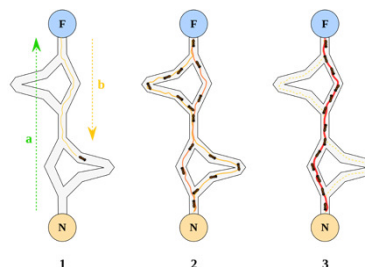
33

Colonie de fourmis et le voyageur

Colonie de fourmis = une bonne heuristique à ce problème

1. une fourmi parcourt au hasard l'environnement autour de la colonie
2. découvre une source de nourriture et rentre directement au nid, en laissant sur son chemin une piste de [phéromones](#) ;
3. ces phéromones étant attractives, les fourmis passant à proximité vont avoir tendance à suivre, cette piste ;
4. ces mêmes fourmis vont *renforcer* la piste ;

F=Food
N=Nid

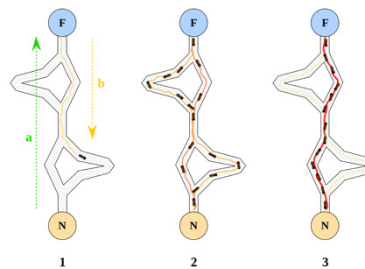


34

Colonie de fourmis

5. si deux pistes sont possibles pour atteindre la même source de nourriture, celle étant la plus courte sera, dans le même temps, parcourue par plus de fourmis que la longue piste ;
6. la piste courte sera donc de plus en plus renforcée, et donc de plus en plus attractive ;
7. la longue piste finira par disparaître, les phéromones étant volatiles ;
8. à terme, l'ensemble des fourmis a donc déterminé et « choisi » la piste la plus courte

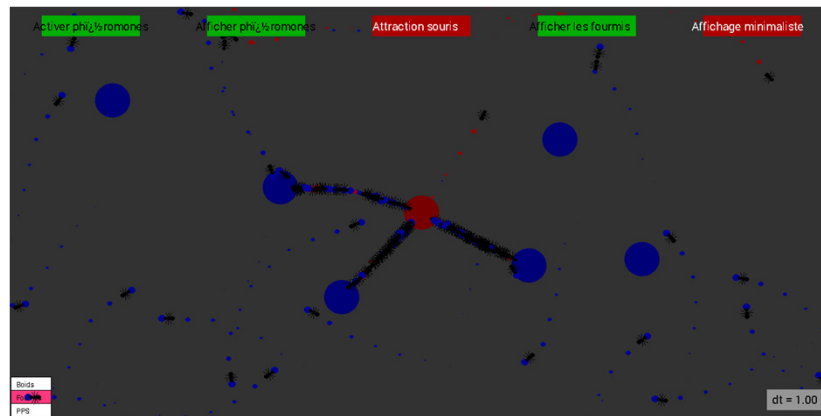
F=Food
N=Nid



35

Colonie de fourmis

- Plus court chemin
 - Il semblerait que Waze utilise un algorithme issue de colonie de fourmis



36

INFORMATIQUE BIO-INSPIRÉ

Algorithmes évolutionnistes / Optimisation

Colonie de fourmis

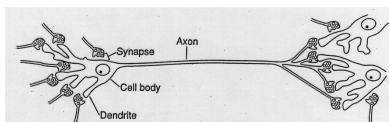
Réseaux de neurones

37

38

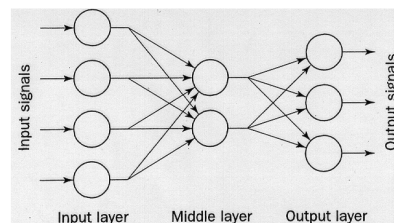
Réseau de neurones artificiel (RNA) un modèle de calcul inspiré du cerveau humain

- Cerveau humain
 - 80 milliards de neurones
 - 1000 à 10000 connexions (synapses) par neurones
- RNA
 - Un nombre fini de processeurs élémentaires (neurones).
 - Liens pondérés passant un signal d'un neurone vers d'autres.
 - Plusieurs signaux d'entrée par neurone



Cerveau
cellule (soma)
dendrites
synapses
axon

RNA
neurone
entrées
poids
sortie

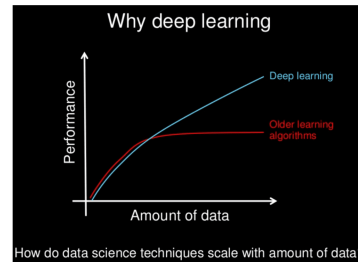


[McCulloch-Pitts, 1943]

38

Réseau de neurones artificiel (RNA) un modèle de calcul inspiré du cerveau humain

- AI
 - ...
 - Machine Learning (apprentissage machine)
 - Random Forest
 - SVM
 - Baysien
 - ...
 - Neural Network
 - **Deep Learning**
 - Apprentissage par renforcement



- Pourquoi les réseaux ont ce succès ?
 - Quand des quantités de données sont devenues plus grandes, les réseaux ont mieux mieux que les autres approches

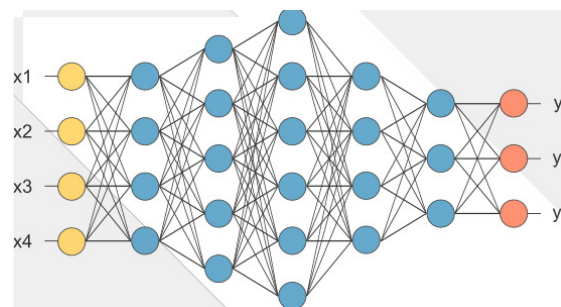
39

Réseaux de neurones



*Pour lui plus de plaisir, plus d'ignorer, plus d'innover.
Ses parents lui font un cadeau spécial.
Dans un monde plein de soleil et de pluie,
Sur les sentiers de terre des sentiers secrets,
Il marche d'un pas, et dans sa tête,
Rappelle à l'instant, un petit souvenir.
De son père, de sa mère, de son frère,
Il dit dans sa tête, et lui fait un sourire.
Le vent, le bruit des feuilles qui tombent,
L'air, le goût de la terre, le goût de la vie,
Il se souvient de tout, et se sent heureux.
Avec le dernier regard, il dit son nom :
D'Europe, en France, maintenant le dit son nom :
Europe ! L'Europe ! Europe de son pays.
Les échos répètent, Europe ! Europe !*

...



Label,
Une image,
Un texte,
etc.

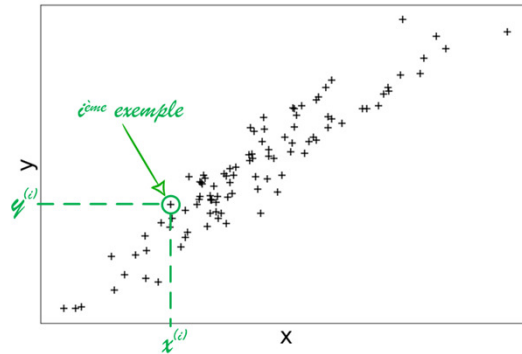
**Une « machine » qui transforme des nombres en d'autres nombres
en « regardant » une base d'apprentissage**

40

Réseau de neurones artificiels

- Soit un problème simple, par exemple prédire le poids d'un enfant en fonction de son âge
 - un jeu de mesures du poids de l'enfant en fonction de son âge
 - $x = \text{âge}$
 - $y = \text{poids}$

Input	Actual output	Desired output
0	0	0
1	3	2
2	6	4
3	9	6
4	12	8



41

Réseau de neurones artificiels

- Soit un problème simple, par exemple prédire le poids d'un enfant en fonction de son âge
 - un jeu de mesures du poids de l'enfant en fonction de son âge
 - $x = \text{âge}$
 - $y = \text{poids}$
- Avec un unique neurone (artificiel) « simpliste »



Régression linéaire : il faut trouver a

42

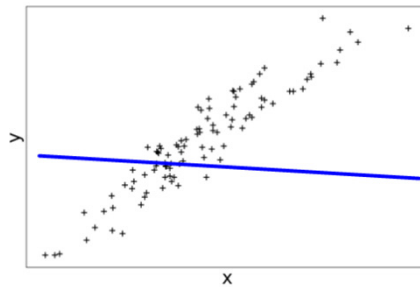
Réseau de neurones artificiels

- Soit un problème simple, par exemple prédire le poids d'un enfants en fonction de son âge



- Avec un unique neurone : il faut trouver **a**

Choisir **a** au hasard



43

Réseau de neurones artificiels

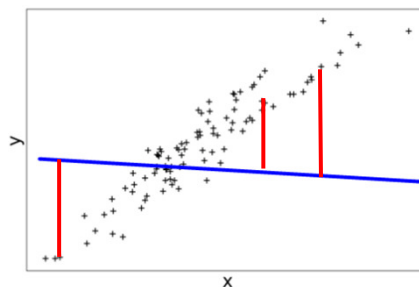
- Soit un problème simple, par exemple prédire le poids d'un enfants en fonction de son âge



- Calculer l'erreur

Pour chaque échantillon des données

$$J(a, b) = \frac{1}{2m} \sum_{i=1}^m (f(x^{(i)}) - y^{(i)})^2$$



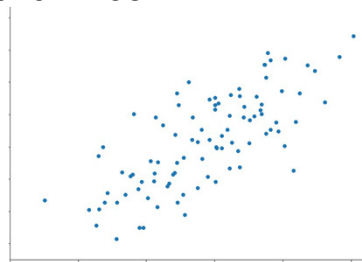
44

Réseau de neurones artificiels

- Soit un problème simple, par exemple prédire le poids d'un enfant en fonction de son âge



- Optimiser l'erreur → changer a petit à petit de manière à ce que l'erreur diminue



Descente de gradient

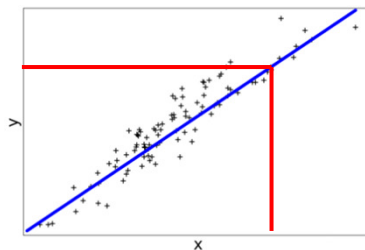
45

Réseau de neurones artificiels

- Soit un problème simple, par exemple prédire le poids d'un enfant en fonction de son âge



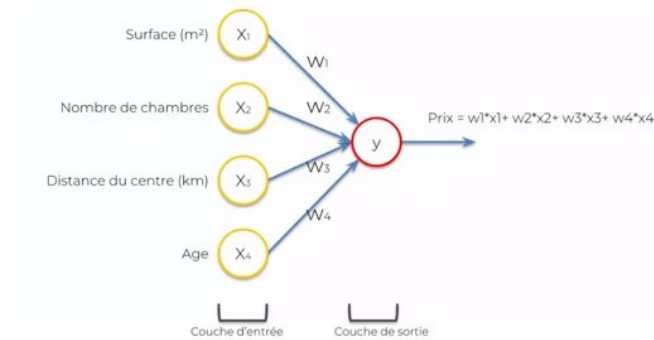
- Exploiter le neurone pour prédire
 - Pour un âge → estimation du poids



46

Réseau de neurones artificiels

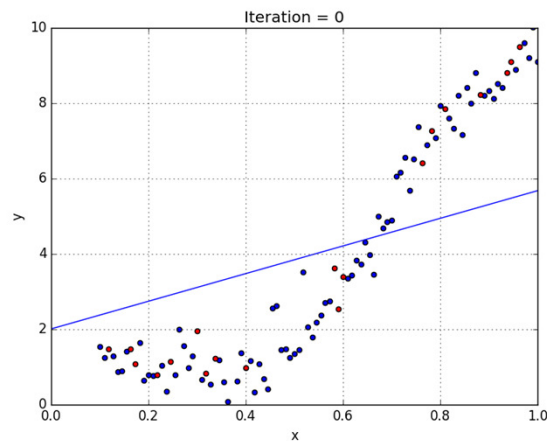
- Souvent les données d'entrées comportent plusieurs dimensions
- ➔ Combinaison linéaire des entrées



47

Réseau de neurones artificiels

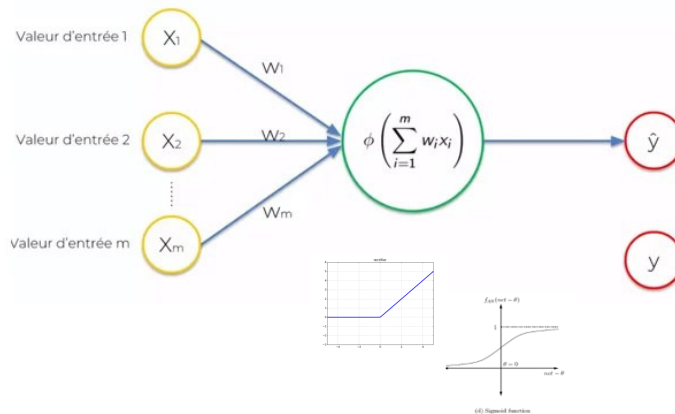
- Souvent le problème est moins simple : **moins linéaire**



48

Réseau de neurones artificiels

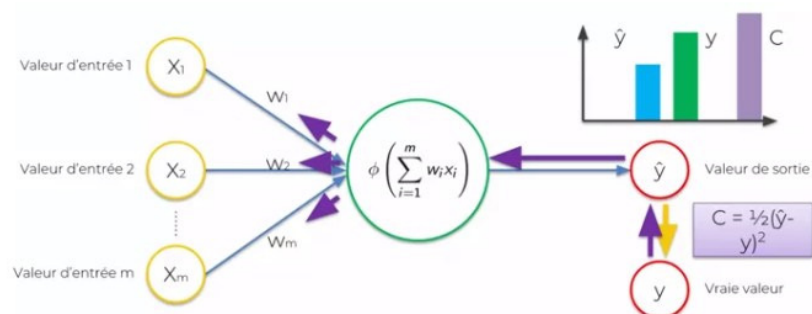
- Seulement souvent le problème est moins simple
 - **Moins linéaire**
- ➔ **Ajout d'une fonction d'activation (non linéaire)**



49

Réseau de neurones artificiels

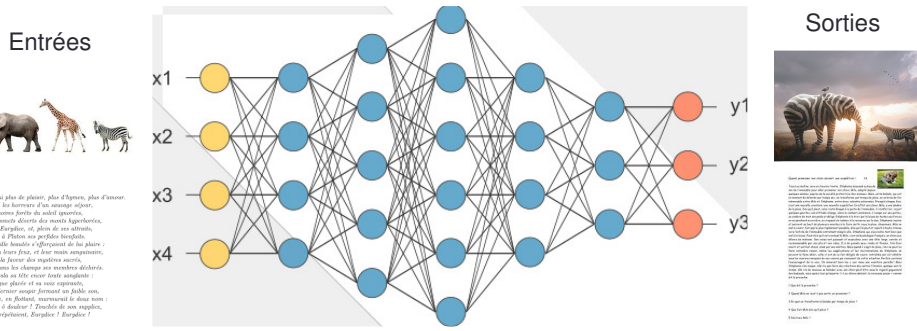
- Backpropagation



50

Réseau de neurones profonds

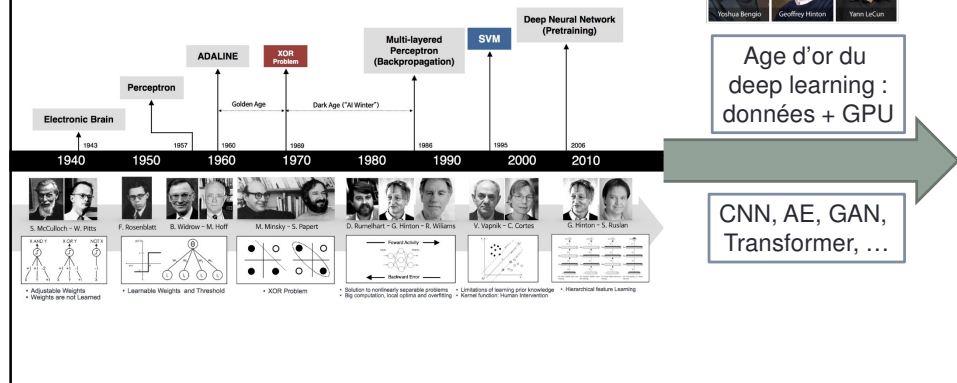
- Chaque cercle est un neurone $f(A.X+B)$
 - Qui prend en entrée des nombres et sort des nombres
 - En bleu les couches cachées



51

Explosion du deep learning

- Il faut des données : souvent images, texte
 - Computer vision
 - Natural Language Processing

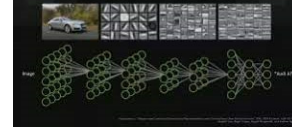


52

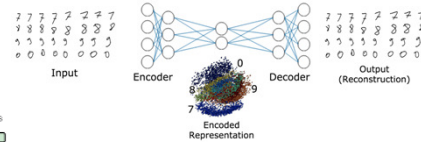
Explosion du deep learning

- Invention de différents type de réseaux

- ConvNN



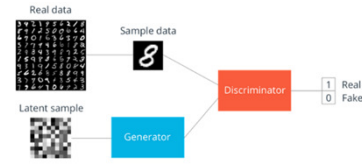
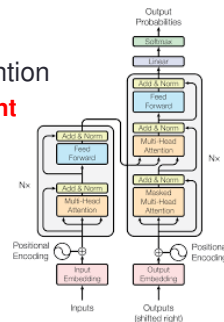
- Auto-encoder



- GAN

- Transformer + attention

- **la star du moment**



53

NNLP et Vision par ordinateur

- Traitement du Langage Naturel

- Machine translation (traduction)
- Natural language generation
- Web Search
- Spam filters
- Sentiment Analysis
- Chatbots
- ...

ChatGPT
DeepL

DALL-E 2

...

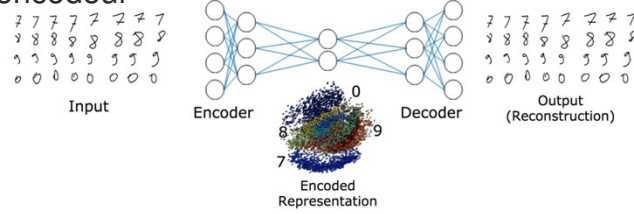
- Vision par ordinateur

- Reconnaissance d'objets, de texte, de comportements, d'expressions, etc.
- Segmentation, conduite automatique
- Diagnostic, imagerie médicale
- Classification, indexation d'images, de video, etc.
- Capture de mouvements, etc.
- ➔ industrie lourde, cinéma, jeux video, médecine, etc.

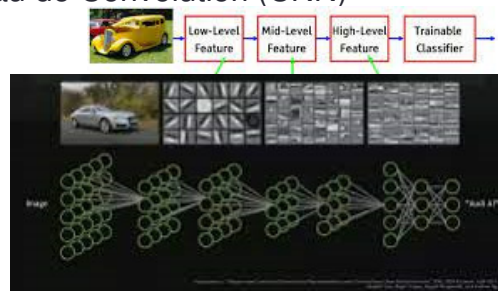
54

Modèle de réseau

- Auto-encodeur



- Réseau de Convolution (CNN)

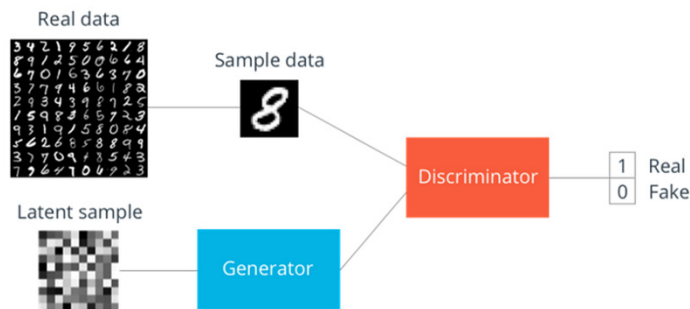


55

GAN

- 2 réseaux en concurrence

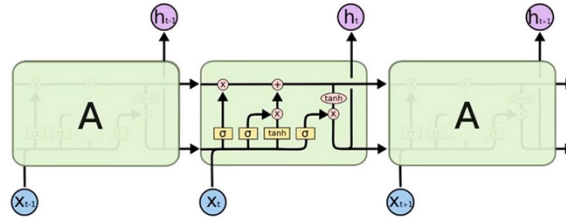
- Générateur : celui qui est gardé à la fin
- Discriminateur : utile que pour entrainer le générateur



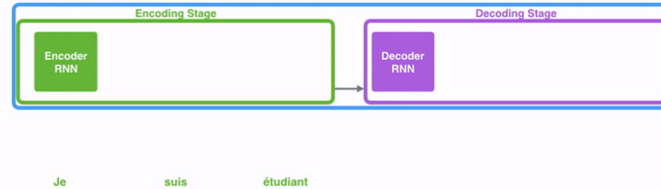
56

Modèle de réseau

- LSTM pour les séquences temporels



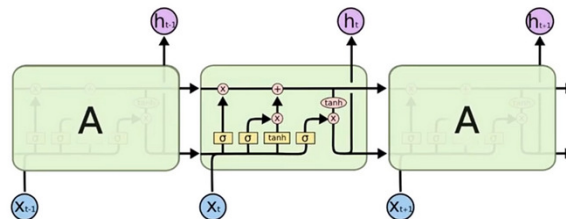
Neural Machine Translation SEQUENCE TO SEQUENCE MODEL



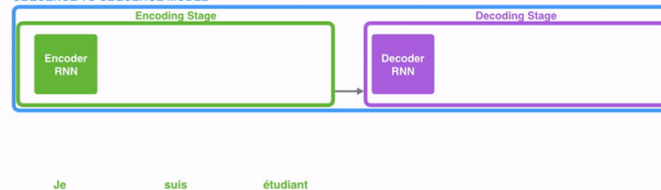
57

Modèle de réseau : séq. temporelle

- LSTM pour les séquences temporelles



Neural Machine Translation SEQUENCE TO SEQUENCE MODEL



58

Modèle de réseau : séq. temporelle

- Transformer avec réseau d'Attention
 - Mots = suite de lettres : mal approprié
 - espace latent de mots, indépendant des lettres
 - L'ordre des mots est importants
 - Réseau d'Attention donne une note à l'importance d'un mot par rapport à un autre

Encoder self-attention: tokens look at each other queries, keys, values are computed from encoder states

Decoder self-attention (masked): tokens look at the previous tokens queries, keys, values are computed from decoder states

Decoder-encoder attention: target token looks at the source queries – from decoder states; keys and values from encoder states

Feed-forward network: after taking information from other tokens, take a moment to think and process this information

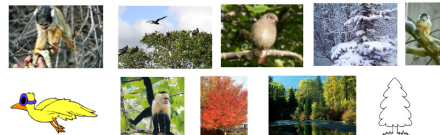
Attention: What part of the input should we focus?

Focus	Attention Vectors
The → The big red dog	[0.71 0.04 0.07 0.18] ^T
big → The big red dog	[0.01 0.84 0.02 0.13] ^T
red → The big red dog	[0.09 0.05 0.62 0.24] ^T
dog → The big red dog	[0.03 0.03 0.03 0.91] ^T

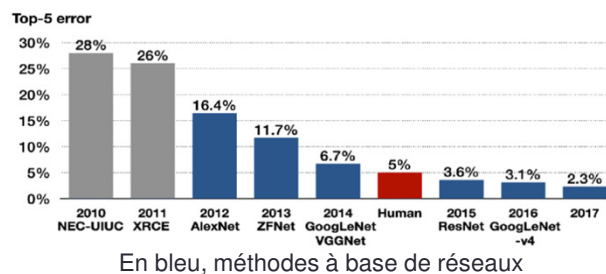
LE RESEAU STAR DU MOMENT

59

Corpus d'images



- Classification
 - Vision humaine : repérer un prédateur ou un membre de sa famille
 - Concours IMAGENET → mettre un label sur une image
 - 10 millions d'images avec 11000 classes, 1.4To



60

Corpus de texte

- Corpus of Historical American English
 - 400 millions de mots
- Google Books Ngram Viewer
 - 500 milliards de mots
- Traduction : Dataset Chris Callison-Burch
 - 22 millions de phrases courtes
- WikiMatrix
 - 135 millions de phrases dans 1620 langues
- CCMatrix
 - 4.5 milliards de phrases courtes issues du Web dans 576 langues

61

ChatGPT Generative Pre-trained Transformer

- Dans son moteur, il y a deux moteurs
 - modèle géant de prédiction de texte combiné avec un modèle encyclopédique
 - modèle conversationnel
 - Apprentissage par renforcement à partir de retour humains
 - Seulement pour la conversation
- **Entraîné avec 500 milliards de mots**
- **175 Milliards de paramètres**
 - 100 milliards de neurone dans un cerveau humain mais ATTENTION à la comparaison

62

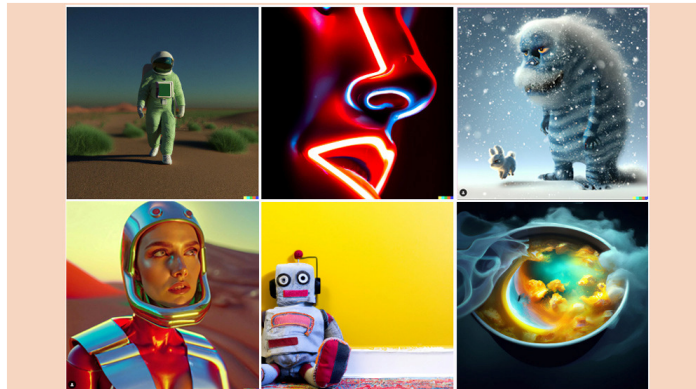
ChatGPT

- Existe dans les labos depuis 2018 mais
 - Grand public demandent de la puissance
 - Risque de mauvais buzz s'il se trompe !
 - Le modèle linguistique n'est pas actualisé (stop en 2021)
 - Le modèle conversationnel
 - Mémoire de 3000 mots pour la conversation
 - continue de s'affiner en fonction du retour des utilisateurs toutes les 3-4 semaines en moyenne
- pas de nouvelles connaissances mais générations phrases de meilleures qualités et, inversement, pénalise davantage les générations malvenues

63

Générateur d'images

- DALL-E : générateur d'images à partir de texte
 - Un auto-encoder basée sur un bruit (mais inversible)
 - Couplé à un modèle de texte



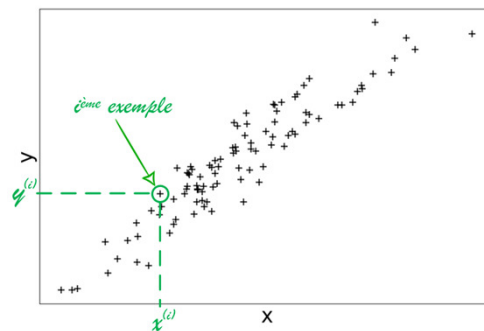
64

Questions ?

65

Réseau de neurones artificiels

- Soit un problème simple, par exemple prédire le poids d'un enfant en fonction de son âge
 - un jeu de mesures du poids de l'enfant en fonction de son âge
 - $x = \text{âge}$
 - $y = \text{poids}$



66

Réseau de neurones artificiels

- Soit un problème simple, par exemple prédire le poids d'un enfant en fonction de son âge
 - un jeu de mesures du poids de l'enfant en fonction de son âge
 - $x = \text{âge}$
 - $y = \text{poids}$
- Avec un unique neurone (artificiel)



Régression linéaire : il faut trouver a

67

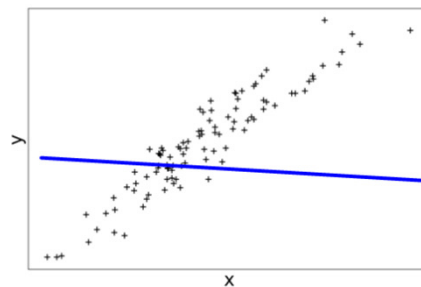
Réseau de neurones artificiels

- Soit un problème simple, par exemple prédire le poids d'un enfant en fonction de son âge



- Avec un unique neurone : il faut trouver a

Choisir a au hasard



68

Réseau de neurones artificiels

- Soit un problème simple, par exemple prédire le poids d'un enfants en fonction de son âge

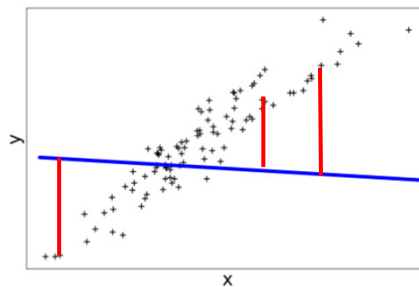


- Calculer l'erreur

Pour chaque échantillon des données

Mean Squared Error

$$J(a, b) = \frac{1}{2m} \sum_{i=1}^m (f(x^{(i)}) - y^{(i)})^2$$



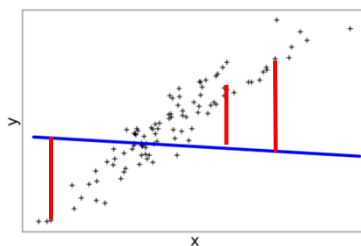
69

Réseau de neurones artificiels

- Soit un problème simple, par exemple prédire le poids d'un enfants en fonction de son âge



- Optimiser l'erreur → changer a petit à petit de manière à ce que l'erreur diminue



70

Réseau de neurones artificiels

- Soit un problème simple, par exemple prédire le poids d'un enfant en fonction de son âge



- Optimiser l'erreur → changer a petit à petit de manière à ce que l'erreur diminue