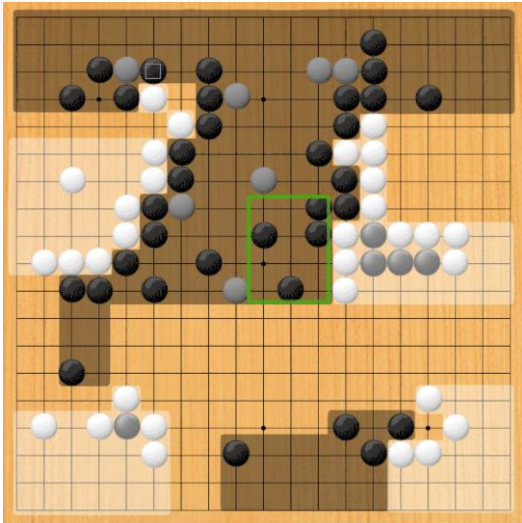


Machine Learning

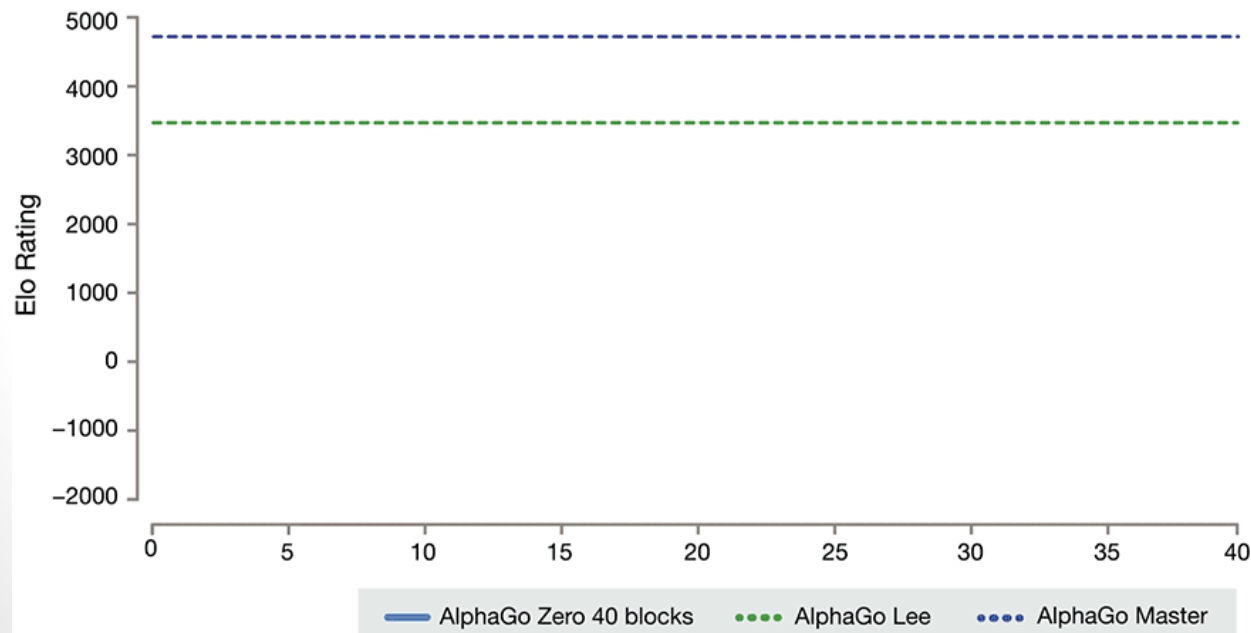
ANN : Artificial Neural Network

I. PERFORMANCE DU MACHINE LEARNING

Jeu de Go

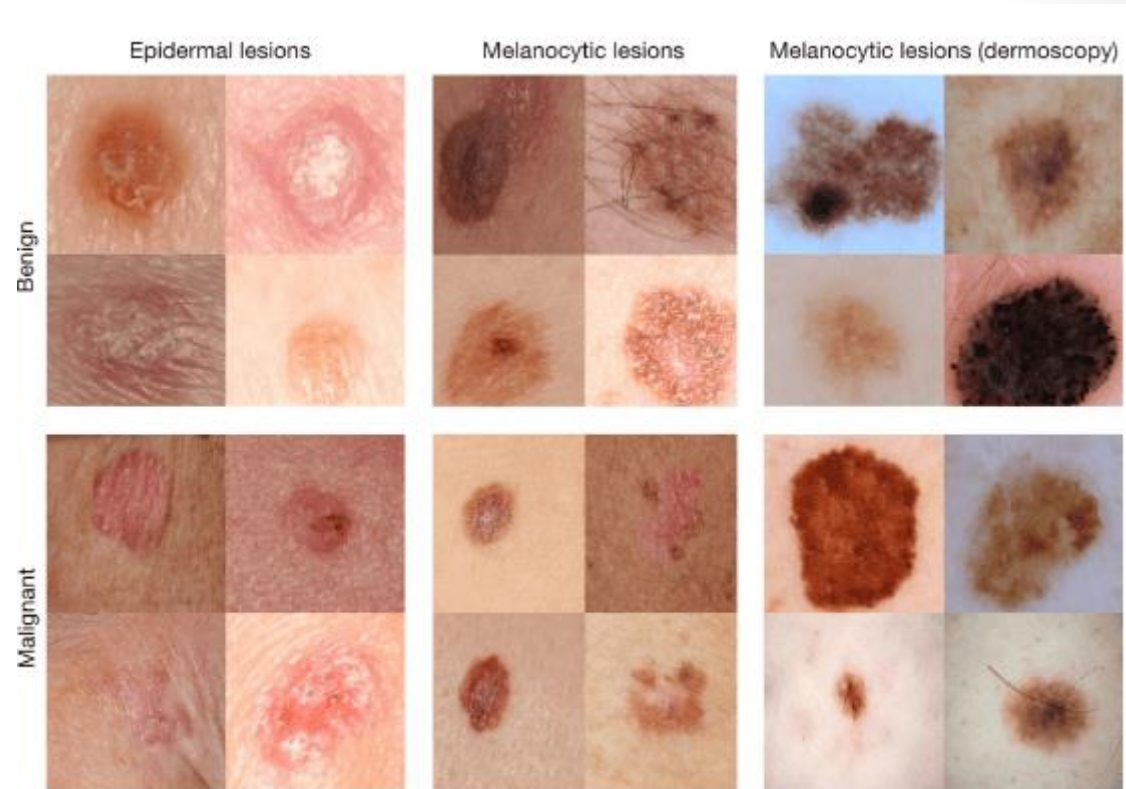


- Sondage auprès des experts présents à ICML 2015:
“L’intelligence Artificielle battra l’humain au Go en 2027.”
- En mars 2016, Alphago de Google deepMind bat le N°3 Mondial (Lee Sedol) 4-1.
- En 2017, AlphaGo Zero réapprend le Go en partant des règles et bat le premier AlphaGo par 100 parties à 0.

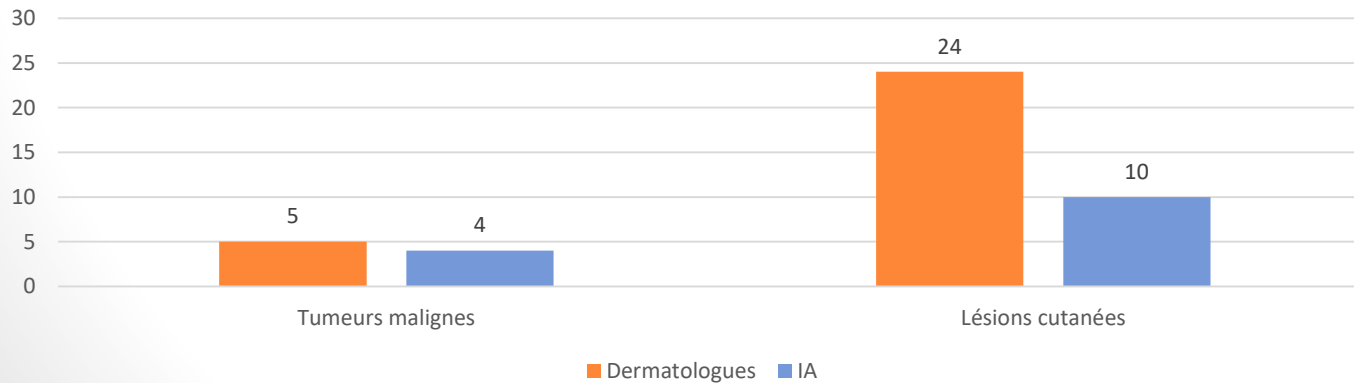


Médecine

- L'IA distingue mieux que les dermatologues les tumeurs des grains de beauté.

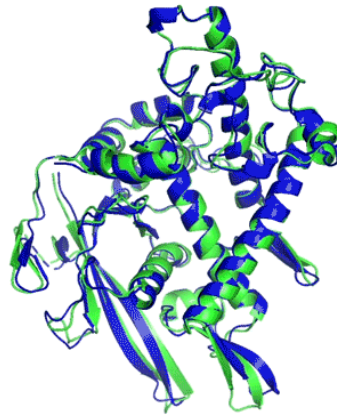


Taux d'erreur Diagnostic

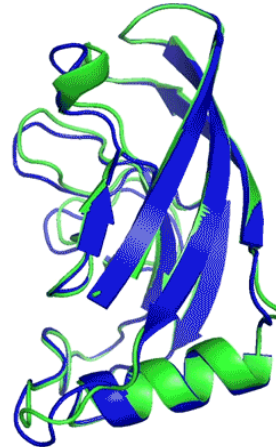


Biochimie

Déterminer la forme 3D d'une protéine à partir de sa séquence d'acides aminés



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

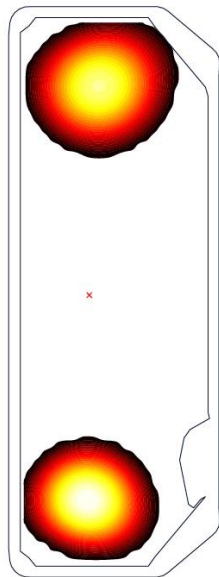
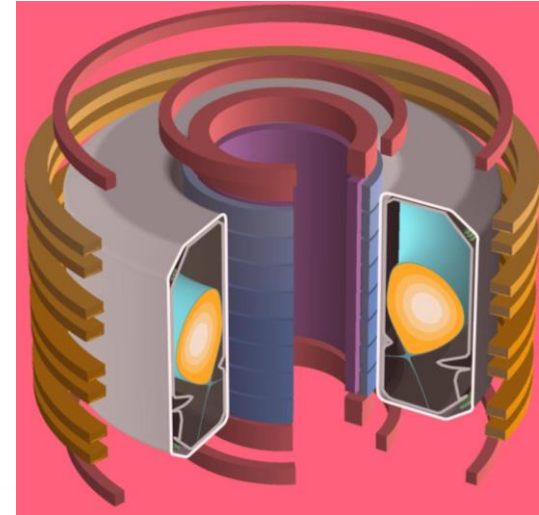


T1049 / 6y4f
93.3 GDT
(adhesin tip)

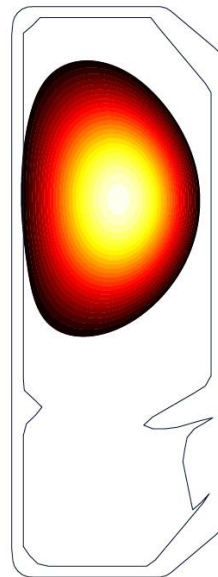
- Experimental result
- Computational prediction

Fusion nucléaire

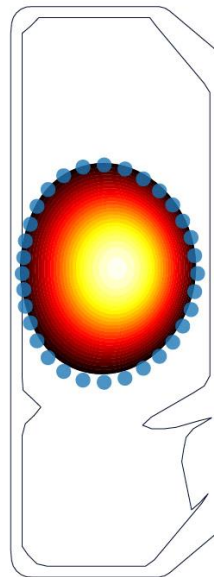
- Les générateurs Tokamak ont pour ambition de répliquer l'énergie du soleil : La fusion nucléaire.
- L'IA de DeepMind Contrôle par champs magnétique la forme du plasma à l'intérieur de la cuve.
- De nouvelles formes ont été inventées pour le plasma dans la cuve.
- <https://actu.epfl.ch/news/epfl-and-deepmind-use-ai-to-control-plasmas-for-nu/>



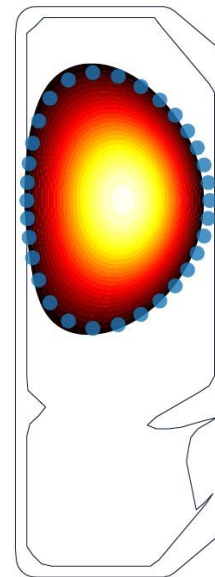
Droplets



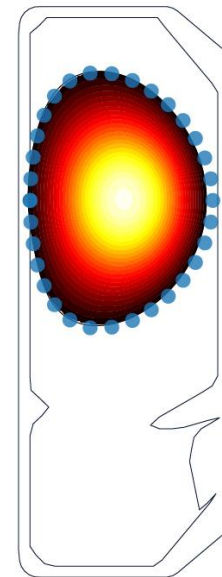
Negative
Triangularity



ITER-like
shape

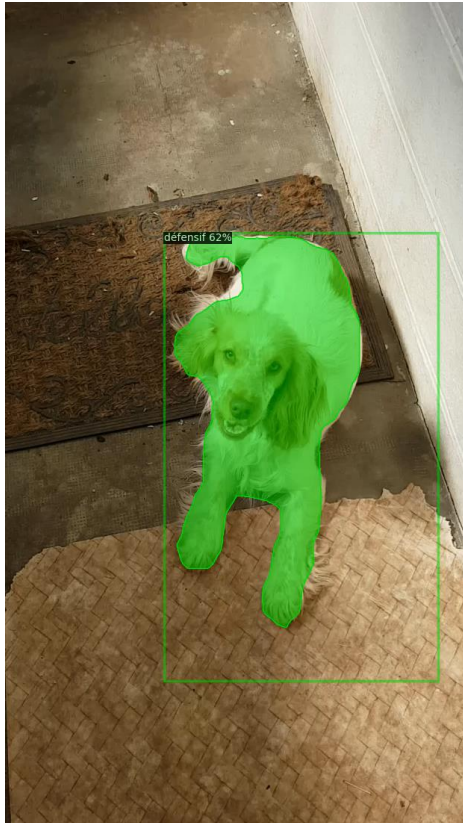


Snowflake

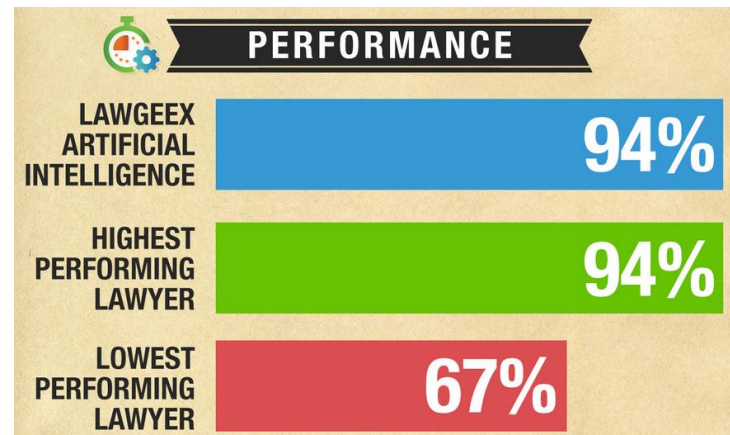
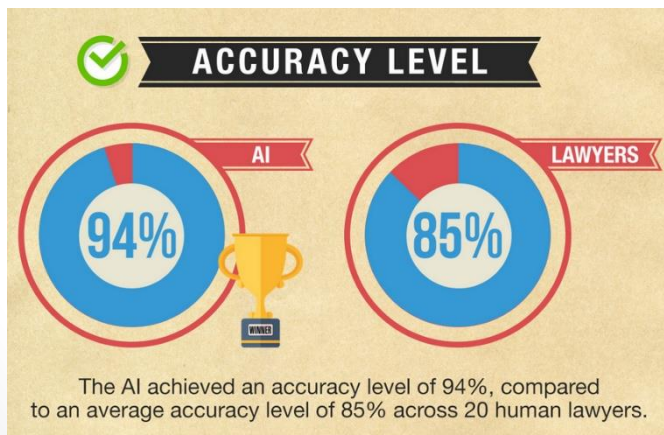
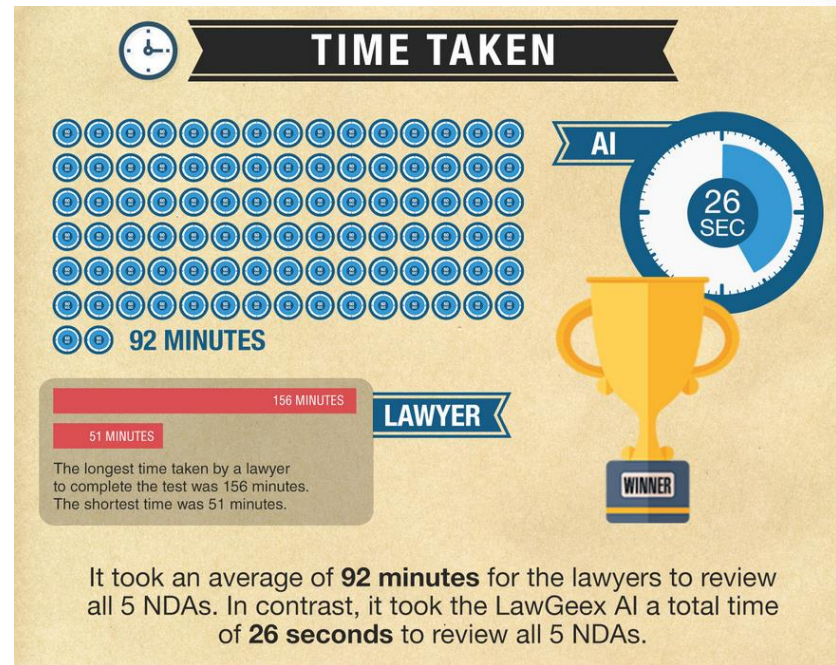


Elongated
Plasma

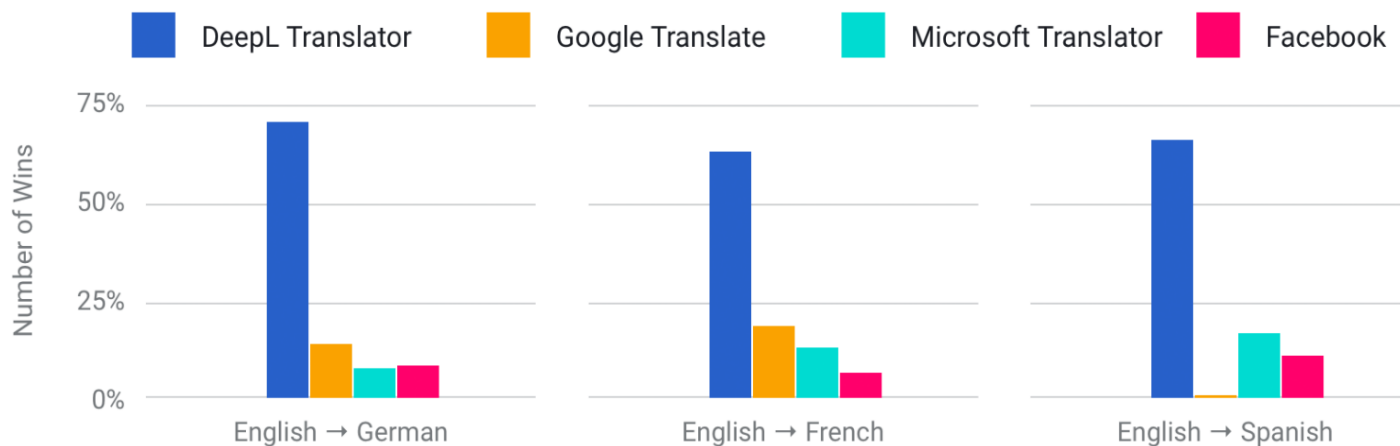
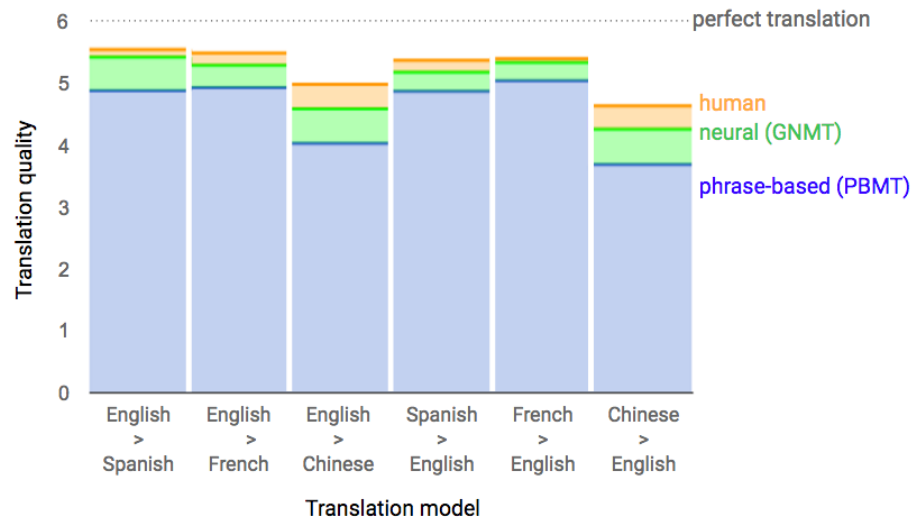
Comportement Animal



Droit : L'expérience LawGeex (2018)



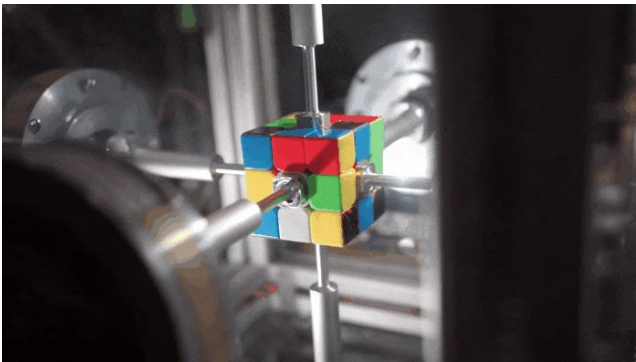
Traduction et Langage



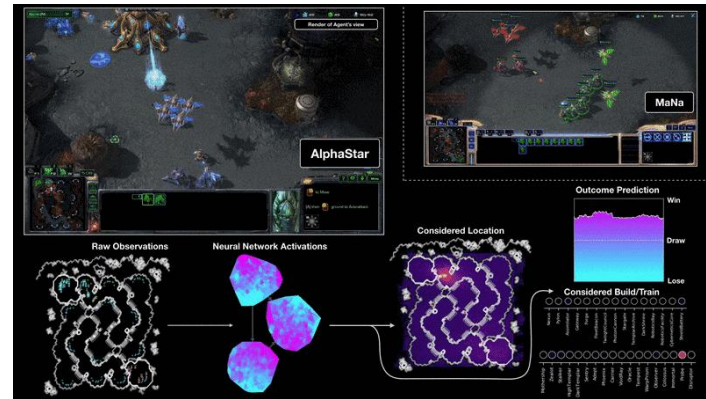
<https://www.deepl.com/>

Jeux

Parfait	Sur-humain	Expert	Humain
Dames Rubik's Cube Poker Limit Holdem	Jeu de Go Poker Jeopardy	Bridge Mots Croisés Dota 2/StarCraft	Conduite Automobile Reconnaissance d'images



Humain : 4''59 / IA : 0''38

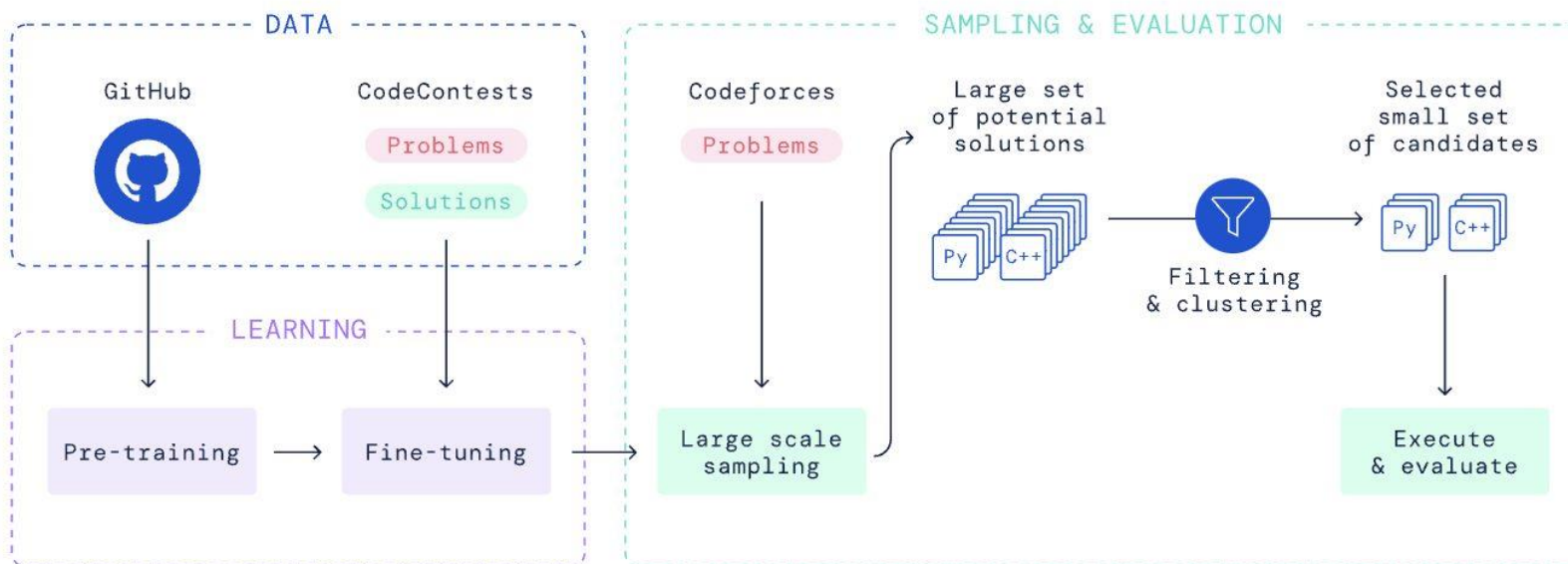


starCraft

Jeu	Années	Positions (log10)	Complexité (log10)	Information complète
Dames	1994	21	31	Oui
Othello	1997	28	58	Oui
Echecs	1997	46	123	Oui
Scrabble	2006			Oui
Shogi	2017	71	226	Oui
Go	2016	172	360	Oui
Poker	2017			Non
StarCraft	?	270+		Non

Génération de code

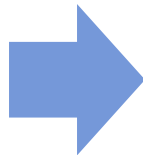
- DeepMind vient de dévoiler **AlphaCode**, une intelligence artificielle, qui « *écrit des programmes informatiques à un niveau compétitif* ».
- Entraîné sur plus de 715 Go de code GitHub
- En fonction du système de notation Elo de Codeforces, AlphaCode se positionne parmi les 28 % des meilleurs compétiteurs des six derniers mois.
- **Copilot** de Github et **CodeT5** de Salesforce sont des offres concurrentes.



Créer une image à partir d'une description.

Dalle 2 d'OpenAI

Un vélociraptor travaillant
à un stand de hotdogs.
lomographie.



Le paradoxe de Movarec



- “Il est relativement facile de faire en sorte que les ordinateurs affichent des performances de niveau adulte aux tests d'intelligence ou aux jeux de dames, et difficile ou impossible de leur donner les compétences d'un enfant d'un an lorsqu'il s'agit de perception et de mobilité.

Steven Pinker: “the most significant discovery uncovered by AI researchers”



II. NOTIONS ET VOCABULAIRES

(COMMUN AU MACHINE LEARNING EN GÉNÉRAL)

Apprentissage et Inférence



- L'apprentissage consiste à faire apprendre le traitement d'un problème au programme qui ne sait rien faire au départ.
- Après l'apprentissage on peut questionner le programme pour le problème concerné. C'est l'inférence.

Apprentissage



Inférence



Panda
Roux

Apprentissage supervisé

- Dans l'apprentissage **supervisé**, on donne la « réponse » pour chaque ligne de donnée.

Features (1 par pixel)



Souris



Éléphant



Souris



Éléphant



Éléphant



Souris



Souris



Éléphant



Souris



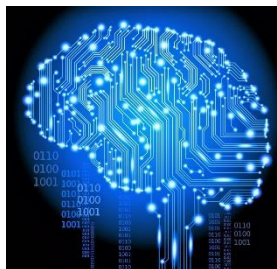
Éléphant



Labels (numérique)



+



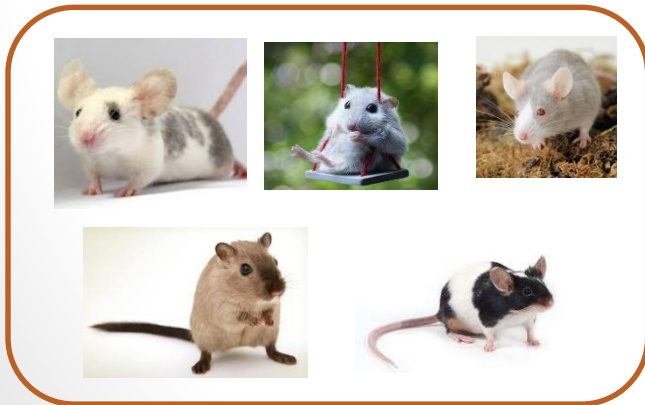
=

Souris



Apprentissage non supervisé

- Dans l'apprentissage **non supervisé**, on laisse le programme chercher des régularités dans les données.



Set 1



Set 2

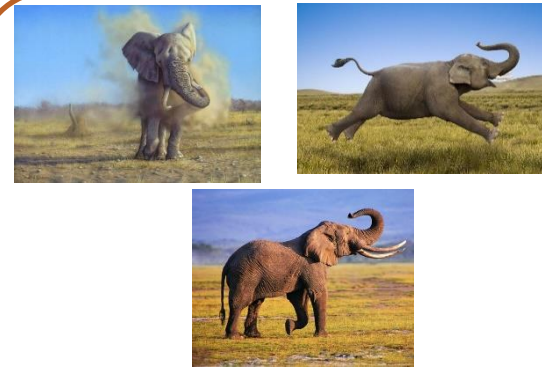


Surprise du non supervisé

- Un apprentissage **non supervisé** peut trouver des régularités imprévues.



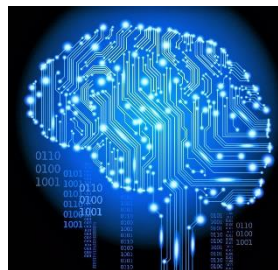
Set 1



Set 2



+



=

Set 2

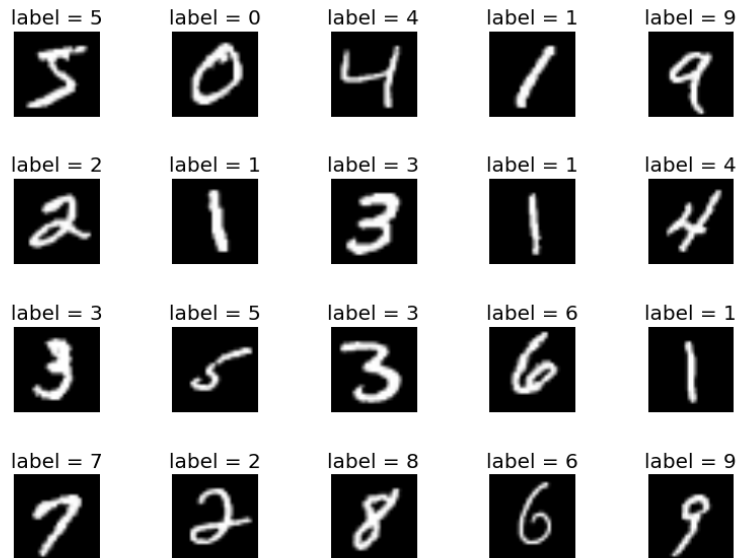


2 Types de problème

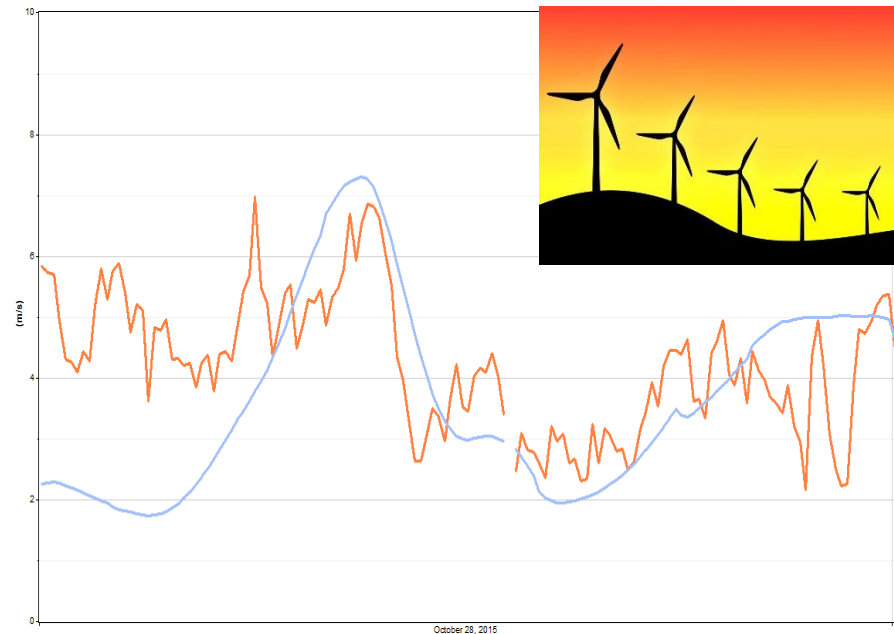
- Les problèmes de classification mettent les exemples dans des « classes ».
 - Exemple : Reconnaissance d'image, détection d'anomalie, applications médicales et juridiques.... => majorité des cas...
- Les problèmes de régression associe chaque exemple à une valeur continue.
 - Exemple : Prévission de production, prix d'une maison, météo...
 - Attention : Ne pas confondre avec la « Droite de Régression ».



2 Types de problème



Classification (MNIST)



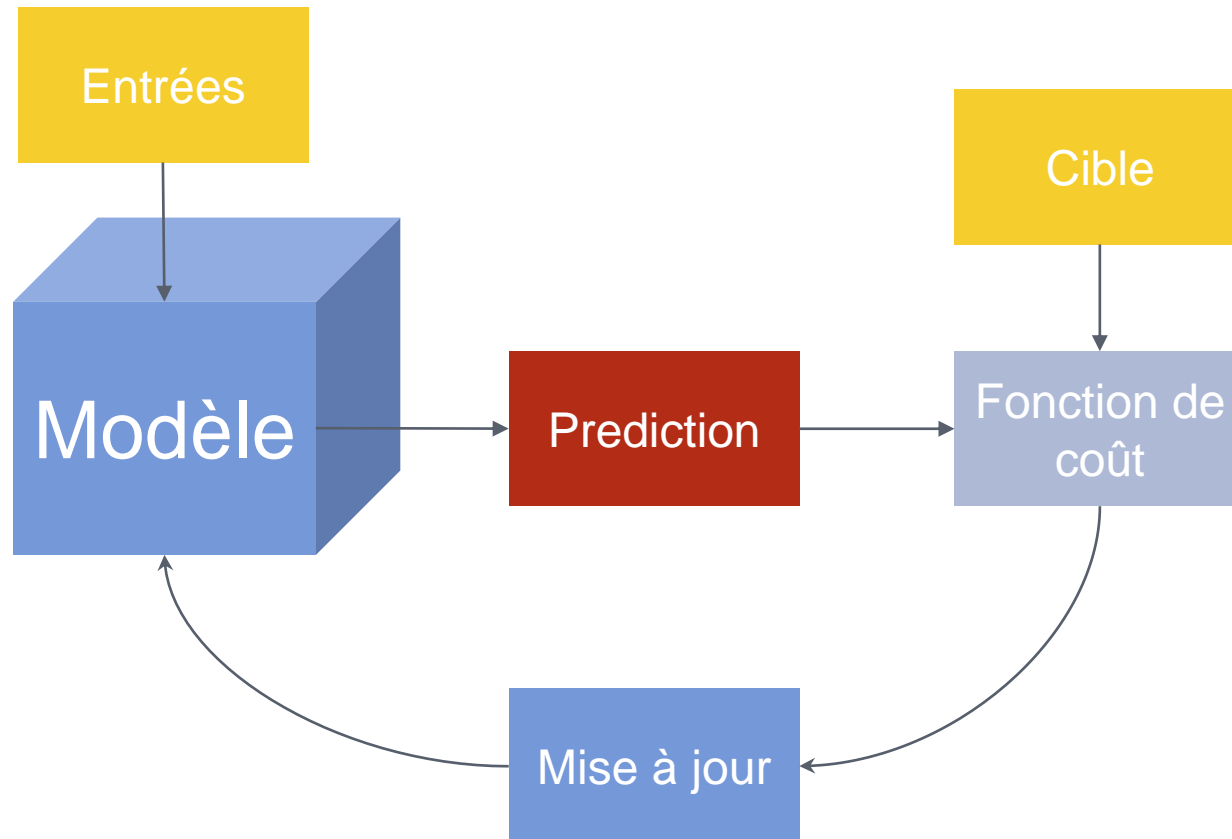
Régression

date_time	winddir100m	windspeed100m	t100m	density100m	gfs_dir100m	gfs_speed100m	gfs_t100m	gfs_density100m	gfs_dir150m	gfs_speed150m	gfs_t150m	gfs_density150m	measured_production
19/11/2016 03:30	36.00	2.05	6.65	1.20	49.00	2.10	13.04	1.20	51.00	2.20	13.04	1.19	414.95
19/11/2016 03:45	35.00	2.07	6.62	1.20	51.00	2.10	13.00	1.20	52.00	2.20	13.01	1.19	393.50
19/11/2016 04:00	35.00	2.09	6.60	1.20	52.00	2.10	12.97	1.20	54.00	2.20	12.98	1.19	419.55
19/11/2016 04:15	35.00	2.12	6.58	1.20	54.00	2.10	12.94	1.20	56.00	2.20	12.94	1.19	459.90
19/11/2016 04:30	34.00	2.14	6.56	1.20	56.00	2.10	12.91	1.20	57.00	2.20	12.91	1.19	419.25
19/11/2016 04:45	34.00	2.15	6.54	1.20	57.00	2.10	12.88	1.20	59.00	2.20	12.88	1.19	455.30
19/11/2016 05:00	34.00	2.17	6.53	1.20	59.00	2.10	12.85	1.20	60.00	2.20	12.86	1.19	482.50
19/11/2016 05:15	34.00	2.18	6.51	1.20	60.00	2.10	12.83	1.20	62.00	2.20	12.83	1.19	461.30

III. FONCTIONNEMENT MACHINE LEARNING

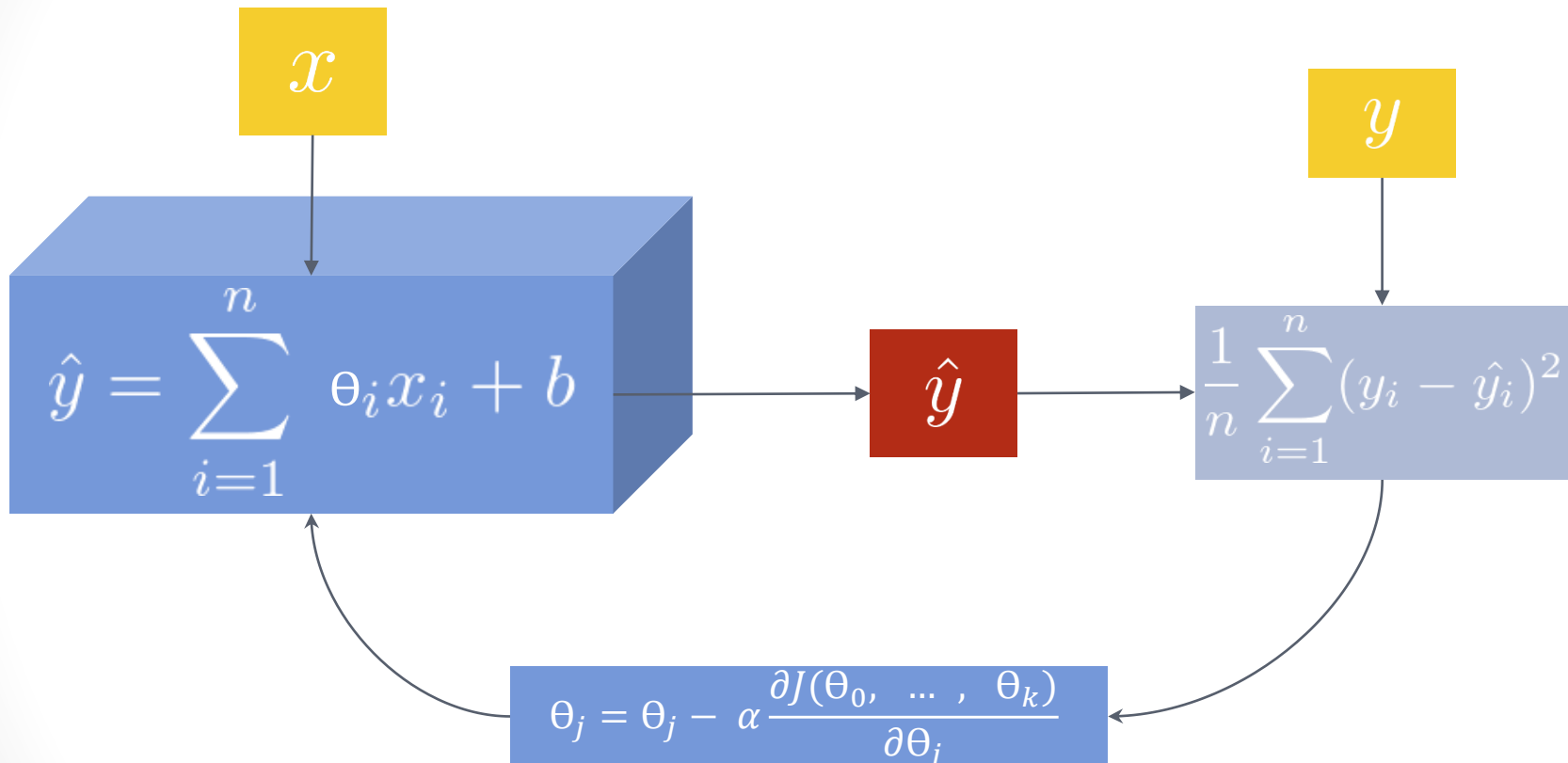
(COMMUN AU MACHINE LEARNING EN GÉNÉRAL)

Apprentissage





Apprentissage (maths)



Fonction hypothèse et de coût



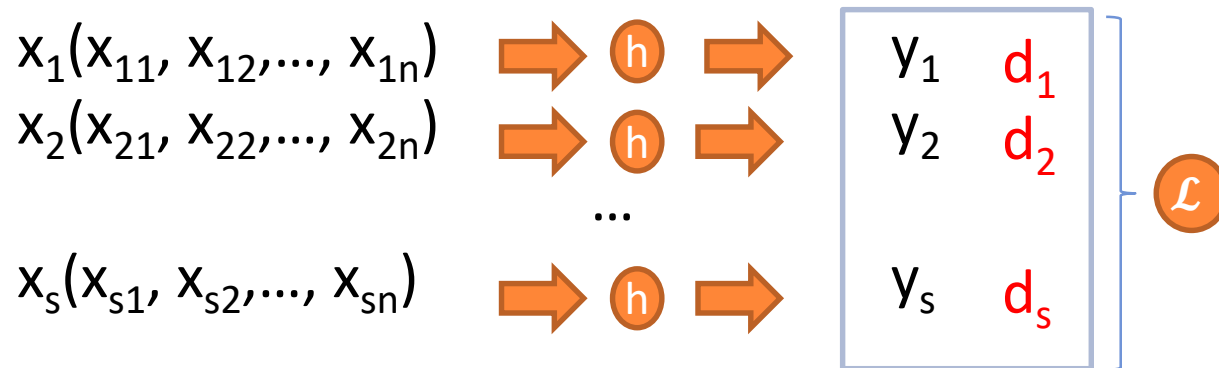
- La fonction **hypothèse** est celle qu'on essaie de paramétrer pour obtenir la bonne sortie à partir des entrées.
- Le type de fonction hypothèse caractérise la méthode : ANN, Arbre, Régression...
- La fonction **de coût** est la fonction que l'on va chercher à minimiser dans l'apprentissage. On l'appelle aussi **Loss**.
- La fonction de coût peut varier suivant le problème considéré.
- Nous verrons bientôt (Descente de gradient) qu'une fonction de coût devrait être continue, dérivable et convexe.





Fonction hypothèse et de coût

- $D=\{(x_1,d_1),,(x_s,d_s)\}$ est le training set de taille s , où x_j est le vecteur d'entrée j de dimension n : $(x_{j1},x_{j2},...,x_{jn})$.
- d_j est la valeur attendue pour le vecteur x_j



- h est la fonction hypothèse, Elle possède des paramètres $\theta_1, \dots, \theta_k$.
- \mathcal{L} est la fonction de coût. Elle dépend des paramètres et des valeurs d .



y et d peut être des simples valeurs ou des vecteurs (par exemple pour une classification).

Fonction hypothèse et de coût

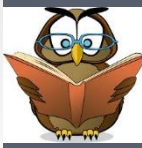


- Exemple de fonction hypothèse (régression linéaire):

$$h(x_i) = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_n x_{in}$$

dans ce cas, h a n+1 paramètres

- Les valeurs y dépendent de h et des paramètres.
- La fonction de coût pour un problème de régression va généralement être la **MAE** ou la **RMSE**.
- Exemple de fonction de coût : $J(\theta_1, \dots, \theta_k) = \sum (d_j - y_j)^2$
- Peut aussi être représentée par un L ou parfois \mathcal{L} , pour « **Loss** »



Fonction de coût Régression

- Dans un problème de régression, la fonction hypothèse produit une valeur y unique.
- La fonction de coût pour un problème de régression va généralement être basée sur la **MAE** ou la **RMSE**. Mathématiquement, il s'agit des normes 1 et 2.
- La RMSE (Root Mean Square Error) est égale à:

$$\text{RMSE} = \sqrt{\frac{\sum (d_j - y_j)^2}{n}}$$

... qu'on peut ramener à la fonction $\sum (d_j - y_j)^2$

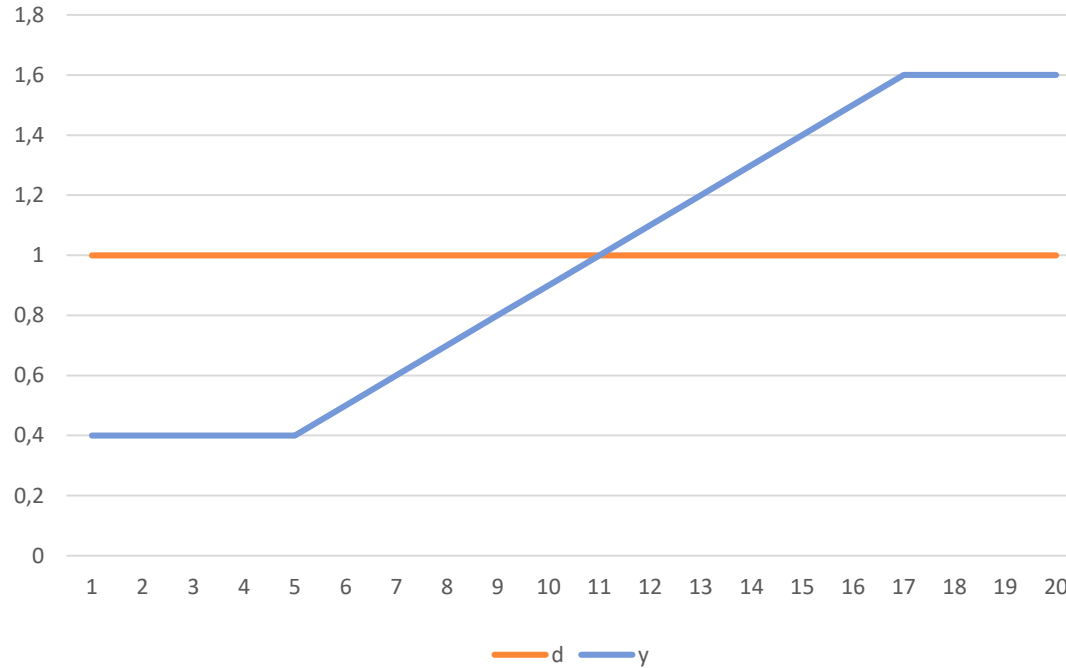
- La MAE (Mean Absolute Error) est égale à:

$$\text{MAE} = \frac{\sum |d_j - y_j|}{n}$$

... qu'on peut ramener à la fonction $\sum |d_j - y_j|$

- Par construction, $\text{RMSE} > \text{MAE}$. On va généralement préférer le RMSE pour éviter les barres de valeur absolue.
- $\frac{\sum (d_j - y_j)}{n}$ est le **Biais**. Il est intéressant, mais ne peut servir de fonction de coût.

Problème du BIAIS



MAE = 0,43

RMSE = 0,47

BIAIS = 0 !!!

Fonction de coût Classification

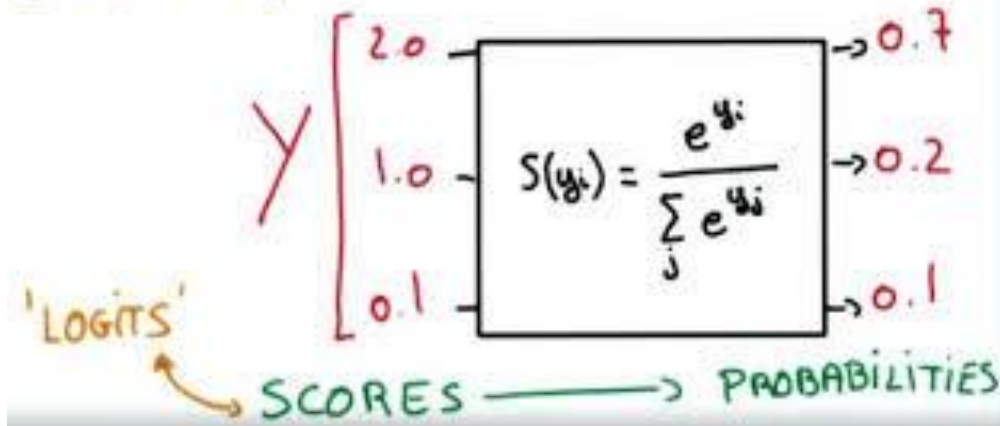


- Dans un problème de classification, la fonction hypothèse, la fonction hypothèse produit un vecteur y dont la taille est égale au nombre de classes.
- On va procéder en 3 étapes :
 - Normaliser les sorties du vecteur y pour que la somme des valeurs soient égales à 1.
=> **Softmax**
 - Représenter le label sous la forme d'un vecteur de longueur égale au nombre de classes, avec 1 pour le label et 0 pour les autres valeurs.
=> **Hot Encoding**
 - Mesurer la « distance » entre le vecteur y et le vecteur label.
=> **Cross Entropy**

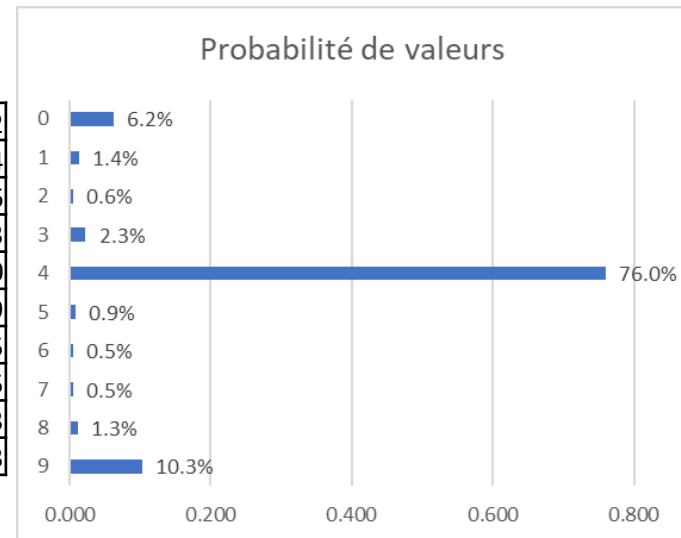
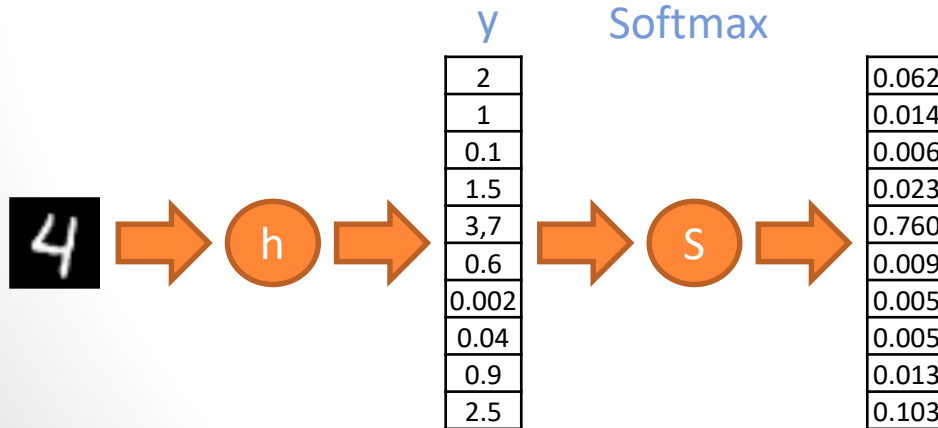


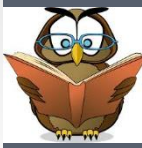
Fonction Softmax

SOFTMAX



Images issues du cours «Deep Learning» d'Udacity





Hot Encoding

- On représente chaque label comme un vecteur de longueur égale au nombre de classe.
- Chaque classe est alors définie par une position dans le vecteur.
- Lorsque le label est de la classe C, son vecteur est composé de 0 sauf pour la position associée à C qui vaut 1.
- Exemple :

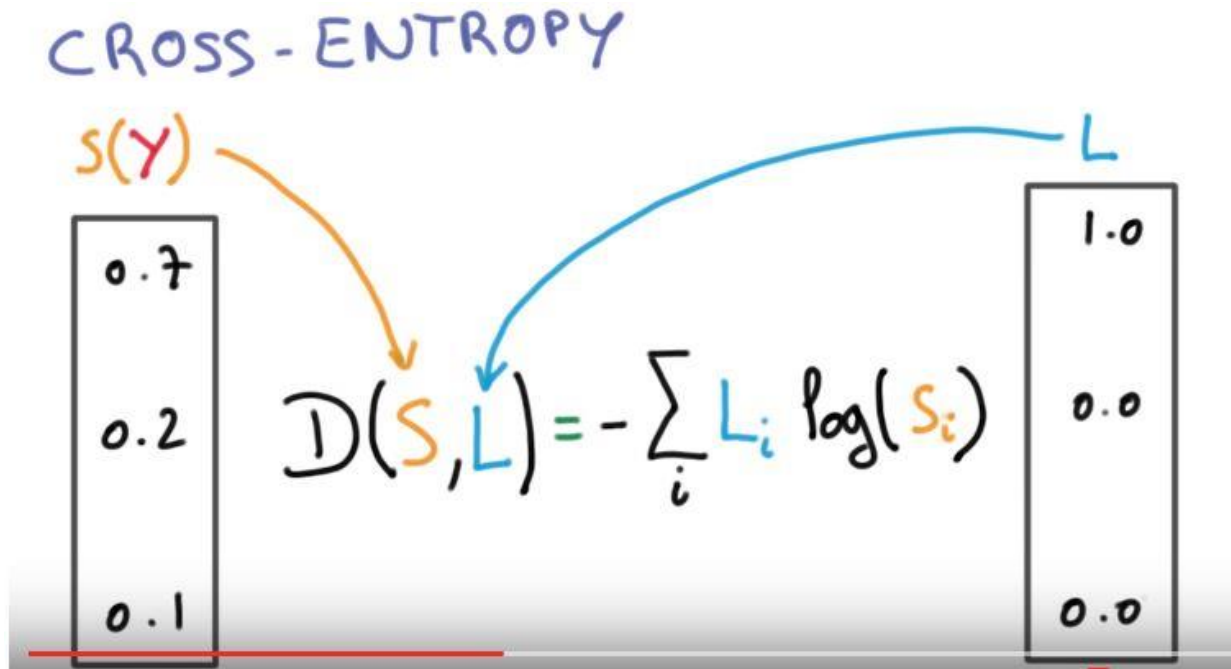
0	1	2	3	4	...	9
1	0	0	0	0		0
0	1	0	0	0		0
0	0	1	0	0		0
0	0	0	1	0		0
0	0	0	0	1		0
0	0	0	0	0		0
0	0	0	0	0		0
0	0	0	0	0		0
0	0	0	0	0		0
0	0	0	0	0		1

- Nous avons pris l'ordre logique, de 0 à 9. Nous aurions pu prendre n'importe quel autre ordre.



Cross Entropy

- La Cross Entropy est une mesure de la distance entre deux vecteurs:

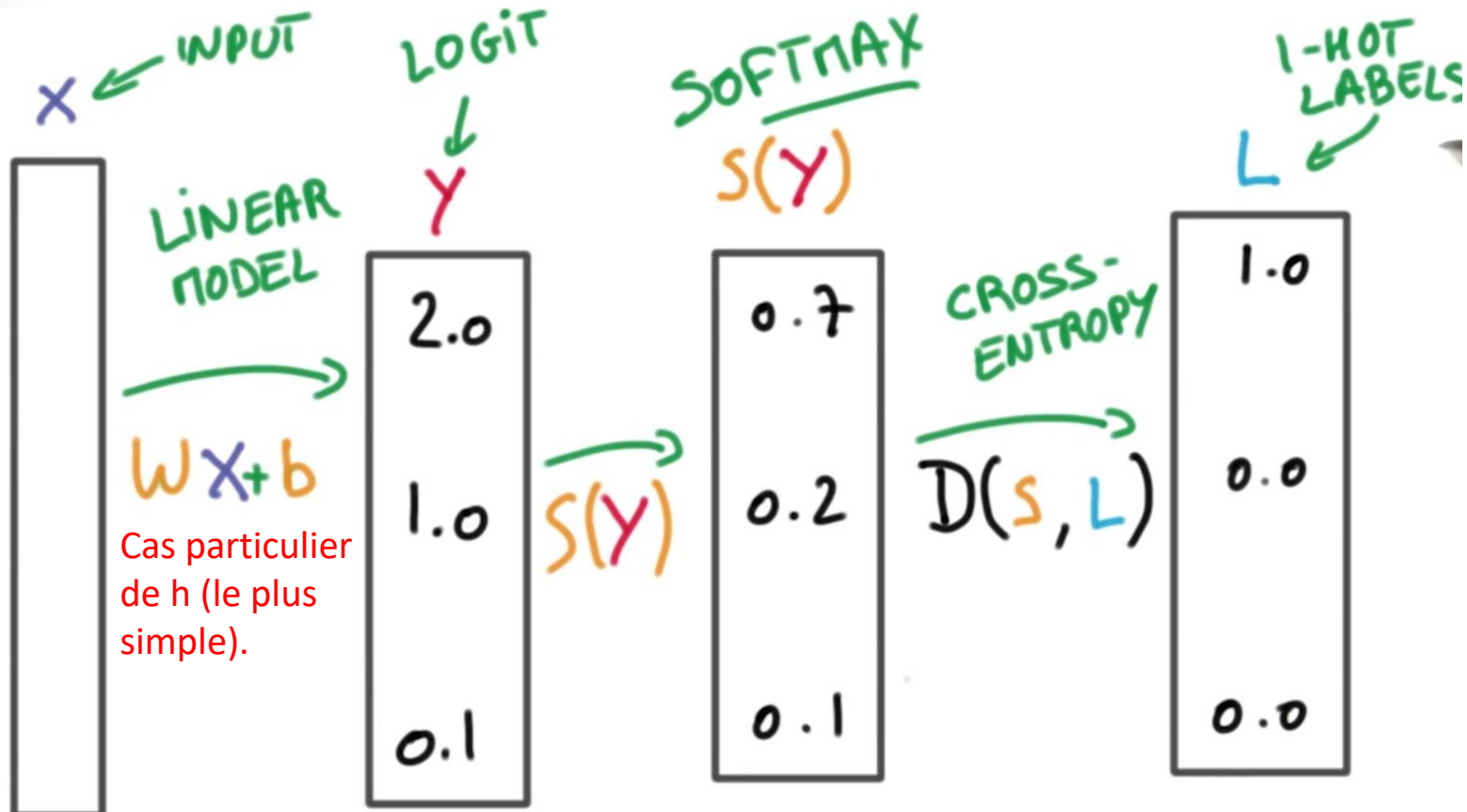


Images issues du cours
«Deep Learning»
d'Udacity

- Attention : $D(S, L) \neq D(L, S)$. $D(L, S)$ est indéfini, car $\log(0)$ n'est pas défini.
- Pour tout comprendre à la cross entropy :

<http://colah.github.io/posts/2015-09-Visual-Information/>

On se résume





Et la fonction de coût est...

- On obtient une valeur de cross entropy pour chaque exemple.
- La fonction de coût qu'on va chercher à minimiser est alors la somme de la Cross Entropy sur le training set (de taille n):

$$J = \frac{\sum_{i=1}^n D(S(h(x_i)), L_i)}{n}$$

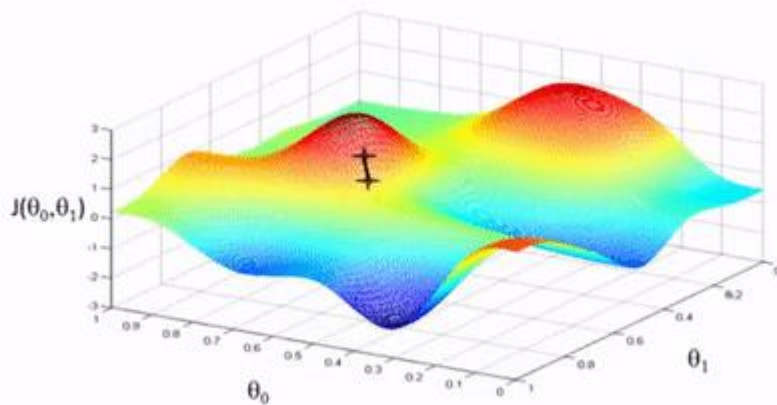
J : Fonction de coût , D : Cross Entropy, S : Softmax,
 h : fonction hypothèse, L : Label « Hot Encoded ».

- La fonction de coût est souvent trop énorme pour être calculée sur tous les exemples. On verra une technique pour réduire la charge de calcul nécessaire (minibatch).



Descente de Gradient


- Ainsi, un problème de machine Learning devient un problème d'optimisation: Comment minimiser la fonction de coût en fonction des paramètres $(\theta_0, \dots, \theta_k)$.
- Pour cela, nous disposons d'un algorithme classique : La Descente de Gradient.

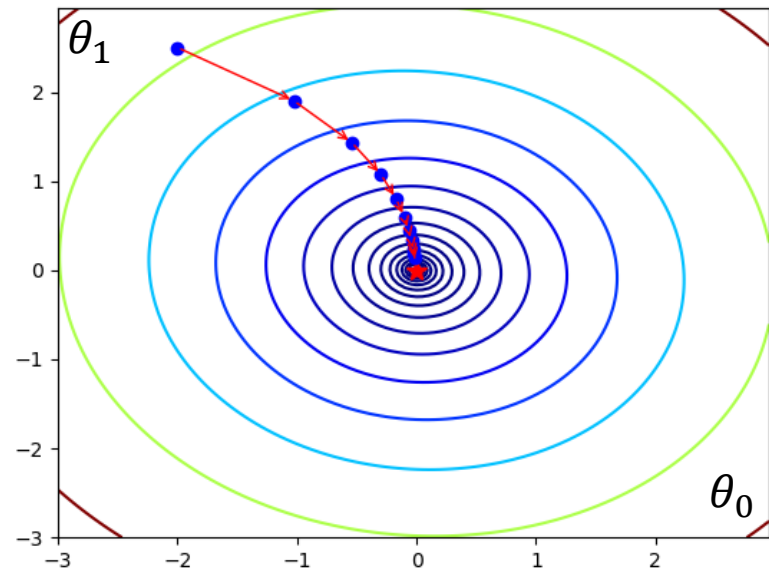
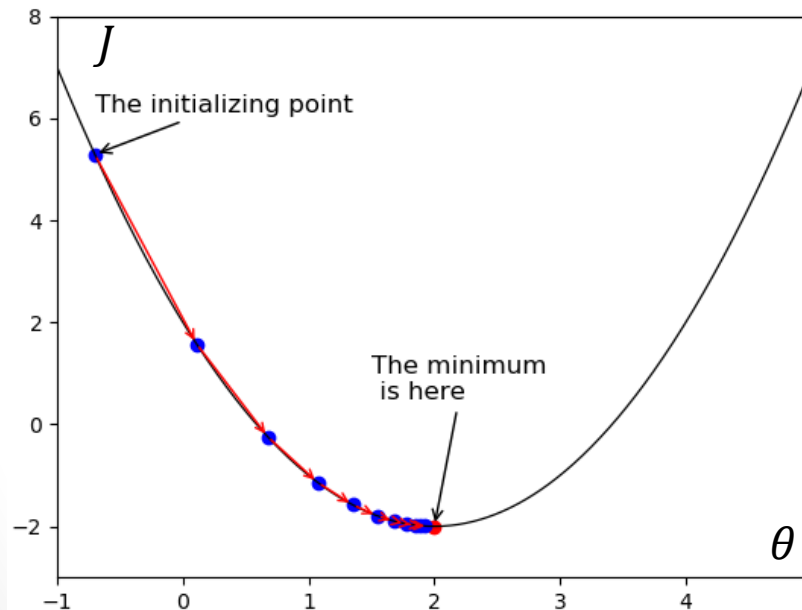


- Intuitivement, L'algorithme est analogue à la descente d'une colline douce par un randonneur.
- A chaque pas, il va chercher la direction qui descend le plus

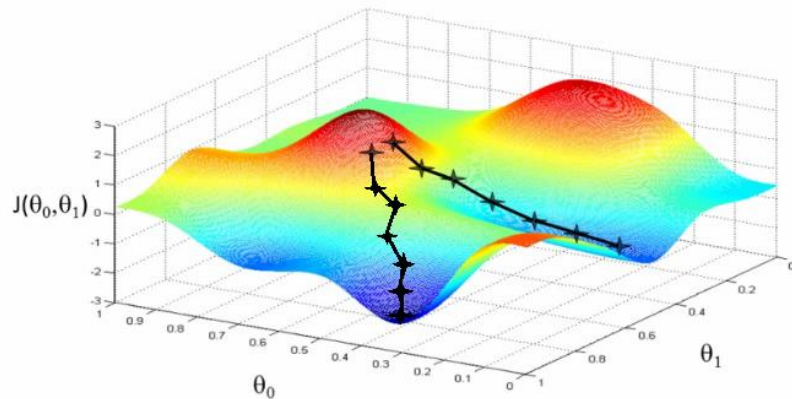
Descente de Gradient



- On applique la descente simultanément pour tous les paramètres θ_j
- J est la fonction de coût. α est le **Learning Rate** (> 0).
- Pour j dans $\{0, 1, \dots, k\}$ $\theta_j = \theta_j - \alpha \frac{\partial J(\theta_0, \dots, \theta_k)}{\partial \theta_j}$ 

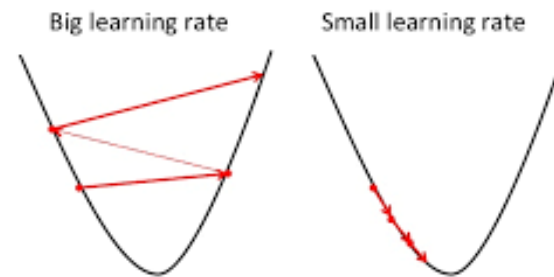


Descente de Gradient - problèmes



- Si la fonction de coût n'est pas convexe, il peut y avoir plusieurs minimum locaux.

Gradient Descent

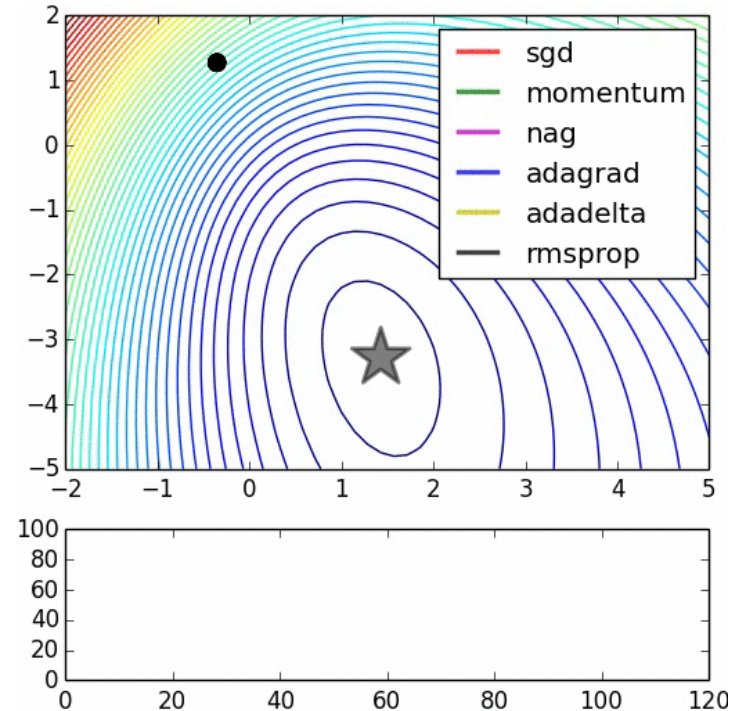
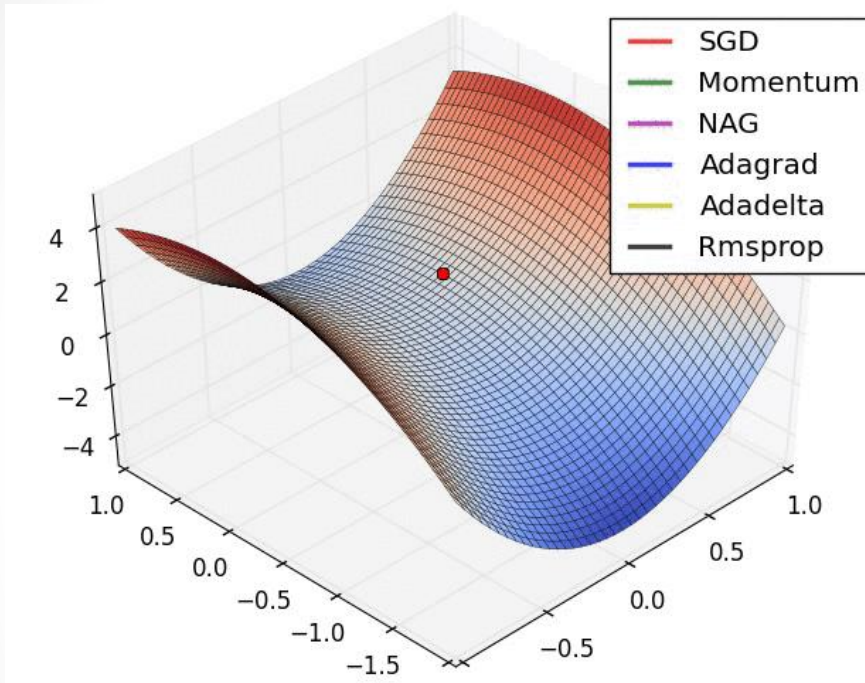


- Si le Learning Rate est trop grand => Divergence
- Si Learning Rate est trop petit => Trop lent.

Descente de Gradient



- Des méthodes de descente de gradient plus évolués existent:



- On joue alors sur:
 - Un sous-ensemble des exemples qui varie à chaque itération.
 - Une évolution du Learning Rate au fil des itérations, ou dépendant de la dimension.
- Adam est l'algorithme le plus utilisé.