



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

ONERA

THE FRENCH AEROSPACE LAB

Kernel Design

Random and Gaussian Processes

sami ben elhaj salah^{1,*} sami.ben_elhaj_salah@onera.fr

26 Janvier 2024

¹ ONERA

Ce document est la propriété de l'ONERA. Il ne peut être communiqué à des tiers et/ou reproduit sans l'autorisation préalable écrite de l'ONERA, et son contenu ne peut être divulgué.

This document and the information contained herein is proprietary information of ONERA and shall not be disclosed or reproduced without the prior authorization of ONERA.

Outline

1 ONERA

Outline

- 1 ONERA
- 2 Introduction

Outline

- ① ONERA
- ② Introduction
- ③ L'IA au service de la mécanique

Outline

- ① ONERA
- ② Introduction
- ③ L'IA au service de la mécanique
- ④ What is a kernel.

Outline

- ① ONERA
- ② Introduction
- ③ L'IA au service de la mécanique
- ④ What is a kernel.
- ⑤ Kernels and positive measures.

Outline

- ① ONERA
- ② Introduction
- ③ L'IA au service de la mécanique
- ④ What is a kernel.
- ⑤ Kernels and positive measures.
- ⑥ Making new from old.

Missions et objectifs

EPIC sous tutelle du ministère des Armées

- EPIC = (Établissement Public à caractère Industriel et Commercial) comme IFREMER, le CEA, le CNES, ...
- Premier parc de souffleries en Europe
- 2 135 salariés en 2022 dont 1 679 ingénieurs et cadres
- 341 doctorants et 19 post-doctorants

L'ONERA a pour mission

- de développer et d'orienter les recherches dans le domaine aérospatial
- d'assurer la diffusion sur le plan national et international des résultats de ces recherches
- de favoriser la valorisation de ces recherches par l'industrie aérospace
- d'en faciliter éventuellement leur application en dehors du domaine aérospatial
- d'assister, en tant qu'expert, les organismes et services officiels
- d'apporter son concours, à la politique de formation à la recherche et par la recherche

Missions et objectifs

EPIC sous tutelle du ministère des Armées

- EPIC = (Établissement Public à caractère Industriel et Commercial) comme IFREMER, le CEA, le CNES, ...
- Premier parc de souffleries en Europe
- 2 135 salariés en 2022 dont 1 679 ingénieurs et cadres
- 341 doctorants et 19 post-doctorants

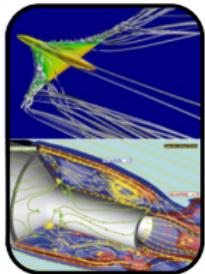
L'ONERA a pour mission

- de développer et d'orienter les recherches dans le domaine aérospatial
- d'assurer la diffusion sur le plan national et international des résultats de ces recherches
- de favoriser la valorisation de ces recherches par l'industrie aérospatiale
- d'en faciliter éventuellement leur application en dehors du domaine aérospatial
- d'assister, en tant qu'expert, les organismes et services officiels
- d'apporter son concours, à la **politique de formation à la recherche et par la recherche**

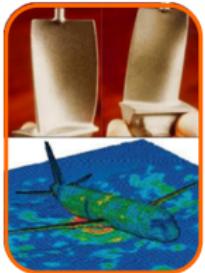
Implantation



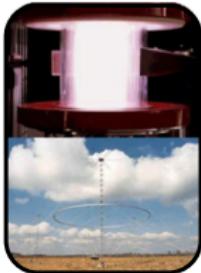
Les domaines scientifiques



Mécanique
des fluides
et énergétique



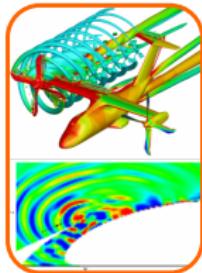
Matériaux
et Structures



Physique



Traitemen
t de
l'informa
tion et Systèmes



Simulation
numérique
avancée



Souffleries



Qu'est ce qu'un numéricien ?



Le numéricien c'est ...

Le numéricien est celui qui conçoit et met en œuvre des algorithmes pour résoudre de manière approchée les équations issus des modèles physiques à l'aide d'un ordinateur.

Le numéricien c'est ...

Le numéricien est celui qui conçoit et met en œuvre des algorithmes pour résoudre de manière approchée les équations issus des modèles physiques à l'aide d'un ordinateur.

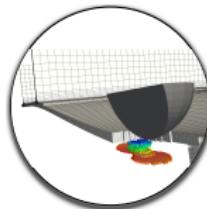
Il se pose des questions telles que :

- comment améliorer la précision des calculs (vis à vis des modèles, ...) ?
- comment calculer plus vite ?
 - utiliser des modèles plus malins ?
 - paralléliser le code de calcul ?
- comment générer des maillages adaptés ?
- comment coupler les codes entre eux (quels algorithmes, quels paradigmes de programmation) ?
- comment intégrer les données massives ?
- comment bien gérer mon code de calcul, le rendre accessible à des utilisateurs, à des stagiaires ?

Mécanique numérique

mathématiques appliquées

$$\int_{\Omega} \underline{\sigma} : \underline{\underline{E}}(\delta \underline{u}) d\Omega - \int_{\partial\Omega_l} \underline{F}_d \delta \underline{u} dS - \int_{\Omega} \underline{f}_v \delta \underline{u} d\Omega = 0 \quad \forall \delta \underline{u} \in \mathcal{U}_0$$



génie mécanique



informatique



Positionnement de la simulation

Dans les projets à financement interne Onera

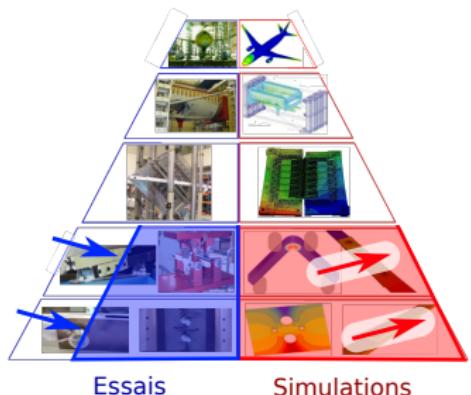
- rôle de compréhension physique
- développement de méthodes numériques
- appuis large sur des outils internes voire développement d'outils dédiés

Dans le cadre de projets institutionnels (DGAC, DGA, ANR, ...) avec les industriels

- volonté de se rapprocher des contraintes industrielles (contraintes des codes commerciaux, ...)

Dans le cadre d'expertises ou de contrats directs avec les industriels

- mise en œuvre des savoir-faire dans le cadre de missions précises



Ingénieur de recherche à l'Onera ?

Les ingénieurs de recherche de l'Onera sont presque tous **docteurs**

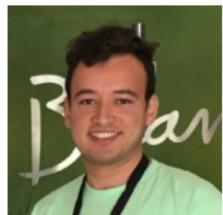
Ingénieur de recherche à l'Onera, c'est :

- participer au montage des projets de recherche
- participer pleinement au travail des projets de recherche
- communiquer les résultats (articles, congrès nationaux et internationaux)
- encadrer des étudiants : stagiaires et doctorants
- enseigner

Informations Personnelles & Parcours

■ Informations personnelles

- ▶ Statut : Docteur-Ingénieur en Mécanique Numérique de Matériaux
- ▶ Qualification : CNU 60 & 26
- ▶ Email : sami.ben_elhaj_salah@onera.fr



■ Depuis janvier 2023

- ▶ Ingénieur de Recherche - ONERA-Châtillon - Département Matériaux et structures (DMAS), équipe MS2.

■ Formations

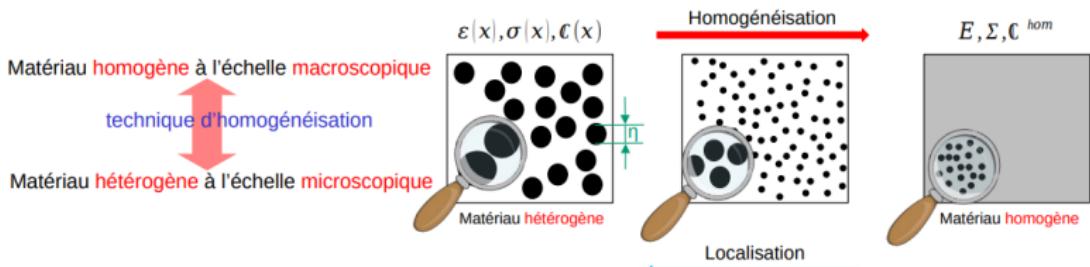
- ▶ 2022, Ingénieur de Recherche Logiciel et Calcul Scientifique, Institut Pprime.
- ▶ 2021, Post-doc, MINES PARIS-TECH & MINES SAINT-ETIENNE.
- ▶ 2020, Ingénieur développeur logiciel, Paris.
- ▶ 2016-2019, Doctorat En Mécanique des Solides, des Matériaux, des Structures et des Surfaces, Institut Pprime, ISAE-ENSMA, Poitiers.

Activités de Recherche

Homogénéisation Asymptotique

η : Petit paramètre d'hétérogénéité pour la structure (porosité).

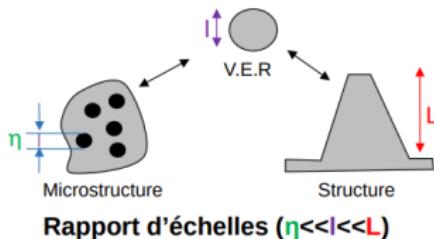
Il se passe quoi quand $\eta \rightarrow 0$?



Conditions de séparation d'échelles:

On considère:

- ✓ L : Dimension caractéristique de la structure macroscopique
- ✓ l : Taille de V.E.R
- ✓ η : Petit paramètre d'hétérogénéité pour la structure (inclusions).



Dans le cas de ma thèse nous avons deux longueurs caractéristiques

Faire une thèse

Les bons côtés

- un très bon moyen d'entrer dans le monde de la R&D
- un très bon moyen de se faire un réseau
- un très bon moyen de continuer “à faire des sciences”
- avoir son sujet pendant 3 ans et en devenir un spécialiste
- les congrès

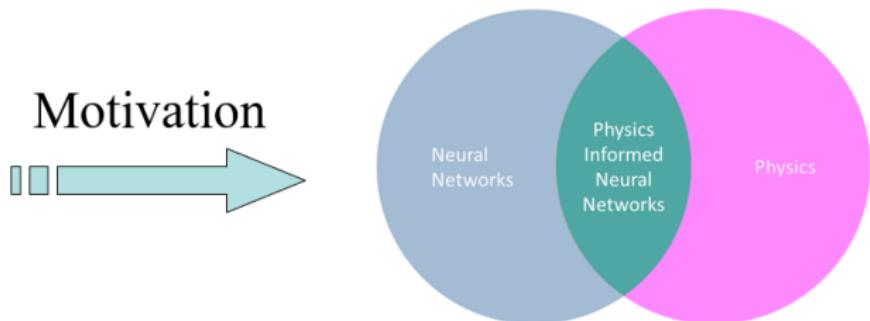
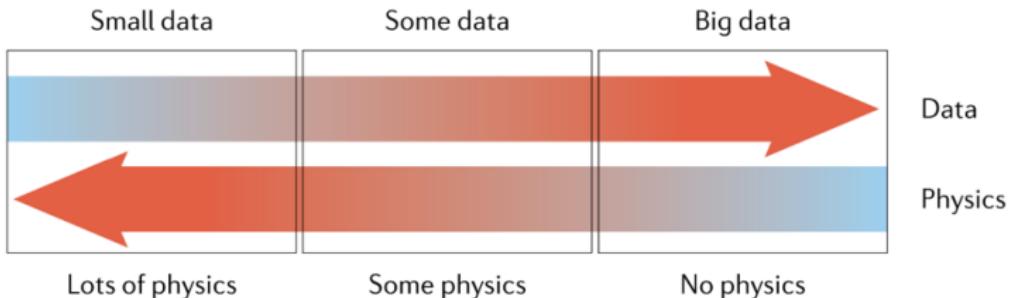
Le moins bon côté

- un salaire parfois moins attractif que celui d'un ingénieur débutant, mais ce n'est même pas toujours vrai (bourse CIFRE, ...)

Après une thèse

- continuer dans le monde académique (enseignant chercheur, chercheur, ...)
- travailler dans un centre de recherche (Onera, CEA, Cetim, ...)
- travailler en R&D chez un grand industriel
 - puis évoluer en BE, en production, en qualité, ...
- travailler en R&D dans une PME
 - puis évoluer
- faire autre chose ...

Motivation



Physics-Informed Neural Networks

Principe

Exemple très simple

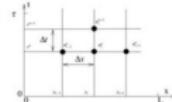
Equation de la chaleur

Equation différentielle

$$\frac{\partial u}{\partial t} - \alpha \frac{\partial^2 u}{\partial x^2} = 0$$

Differences finies
Approximation de la dérivée première

$$\left(\frac{\partial u}{\partial t} \right)_{i,j} = \frac{u_{i,j+1} - u_{i,j}}{\Delta t}$$



Approximation de la dérivée seconde

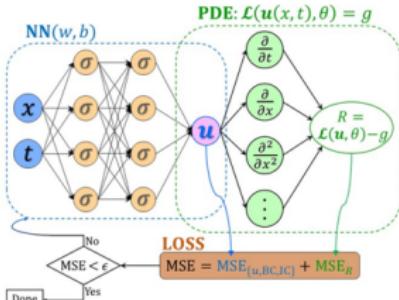
$$\left(\frac{\partial^2 u}{\partial x^2} \right)_{i,j} = \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{(\Delta x)^2}$$

Schéma de discréétisation explicite

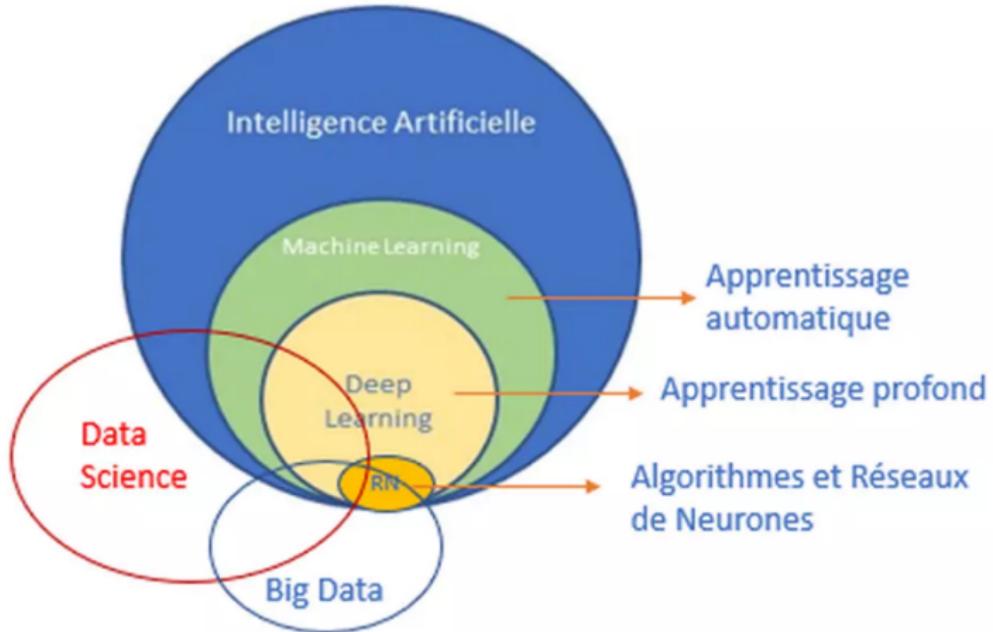
$$\frac{u_{i,j+1} - u_{i,j}}{\Delta t} - \alpha \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{\Delta x^2} = 0$$

Condition de convergence et de stabilité $\alpha \Delta t / \Delta x^2 < 1/2$

Principe des PINNs



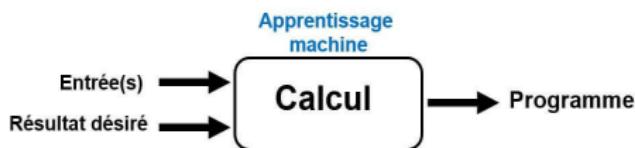
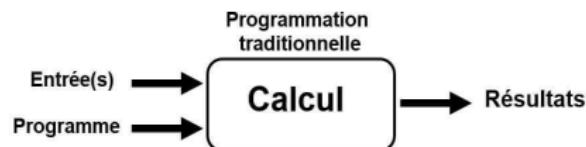
Les concepts importants autour de l'IA



Apprentissage automatique

Définition

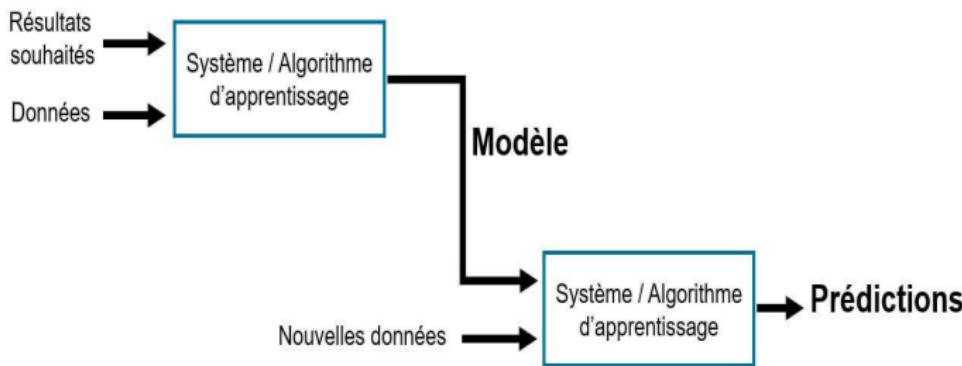
apprentissage d'une fonction cible (f) qui "mappe" le mieux les variables d'entrée (X) à une variable de sortie (Y) : $Y = f(X)$



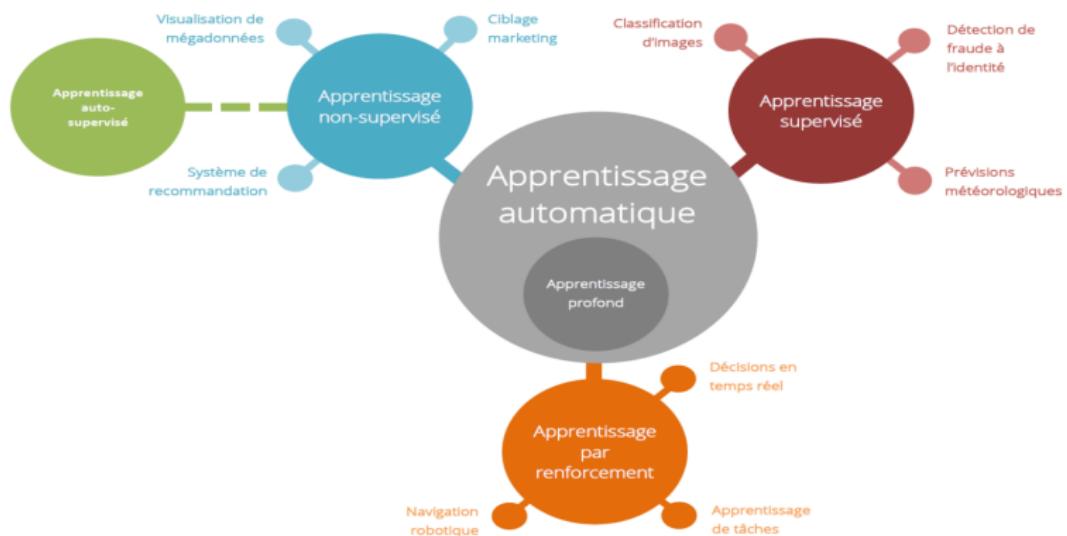
Apprentissage automatique

Définition

apprentissage d'une fonction cible (f) qui "mappe" le mieux les variables d'entrée (X) à une variable de sortie (Y) : $Y = f(X)$



Un aperçu des techniques d'apprentissage automatique



Choix de l'algorithme

Le choix de l'algorithme optimal pour votre problème dépend de ses fonctionnalités telles que

- La vitesse
- La précision des prévisions
- La durée de la formation
- Sa facilité de mise en œuvre
- Sa difficulté à l'expliquer à d'autres

Apprentissage supervisé

Définition

Algorithmes permettant de prévoir ou classifier à partir de données historique et étiquetées



Apprentissage non supervisé

Définition

Algorithmes permettant de classifier, regrouper des profils homogènes à partir de données non étiquetées



Apprentissage semi-supervisé

Définition

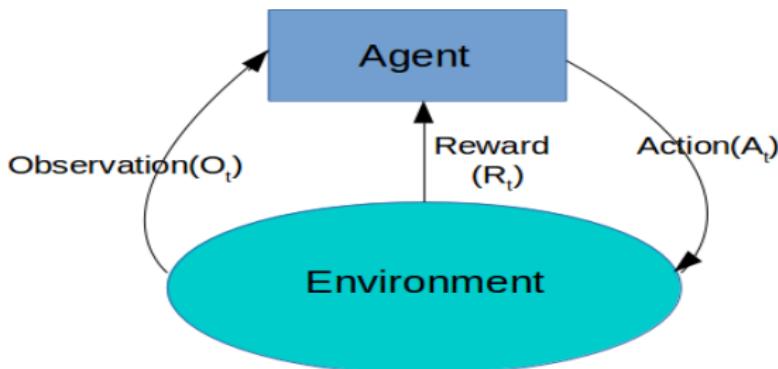
- Apprentissage non supervisé renforcé par un apprentissage supervisé.
- L'apprentissage semi-supervisé est généralement utilisé lorsque nous avons une petite quantité d'entrées étiquetées et beaucoup plus d'entrées non étiquetées.



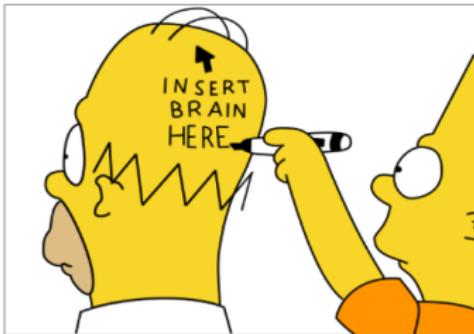
Apprentissage par renforcement

Définition

- Apprendre dans un environnement interactif par essais et erreurs en utilisant les retours sur ses propres actions et expériences
- Algorithmes permettant d'adapter sa stratégie en temps réel en fonction d'un signal positif ou négatif. exemple Robots. jeux ...



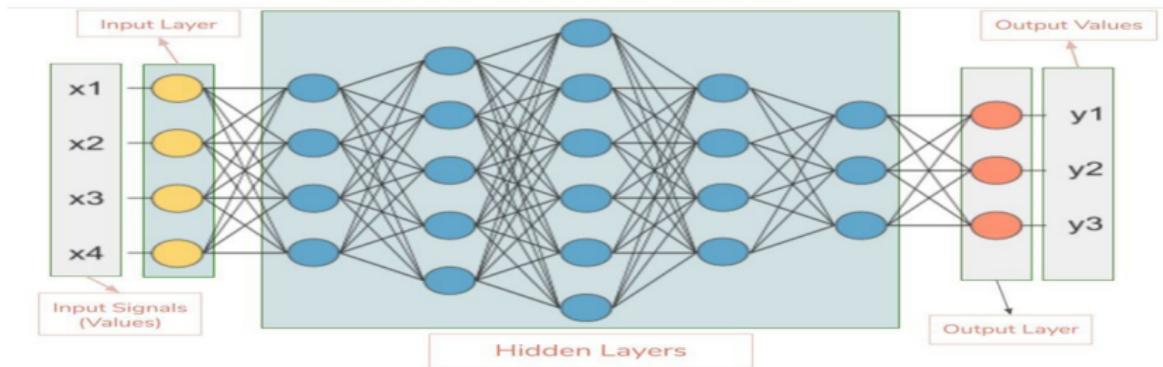
Deep Learning



Fine, but
Deep Learning
What's that?

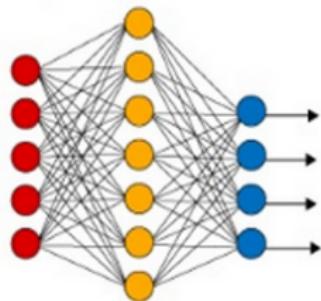
Réseaux de neurones

Réseaux de neurones



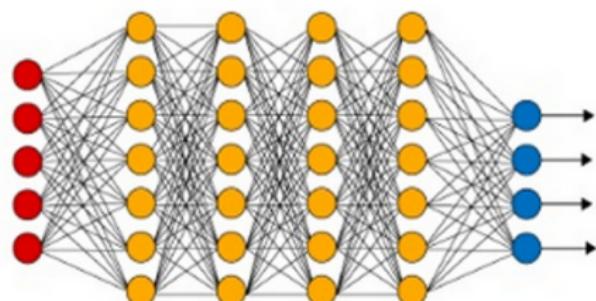
C'est quoi le Deep Learning

Simple Neural Network



● Input Layer

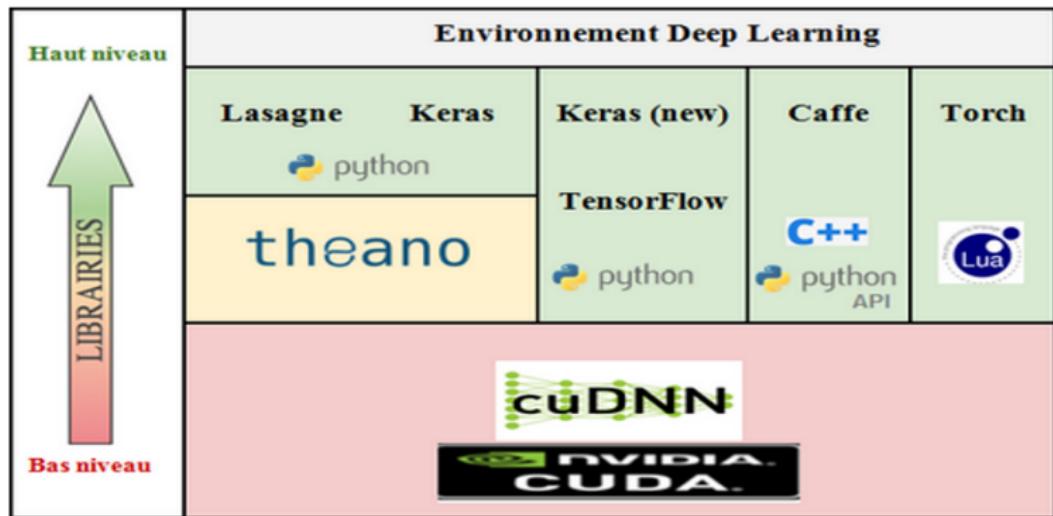
Deep Learning Neural Network



● Hidden Layer

● Output Layer

Environnement de Deep Learning



Environnement de Deep Learning

Python



« Deep learning for humans »

Widely used in the implementation of practical solutions

Keras

By François Chollet (Google)
High level API
Part on TensorFlow since 2017
MIT licence

TensorFlow

Most used DL framework
Supported by Google
Low level API – an hard way
Apache licence

PyTorch Lightning

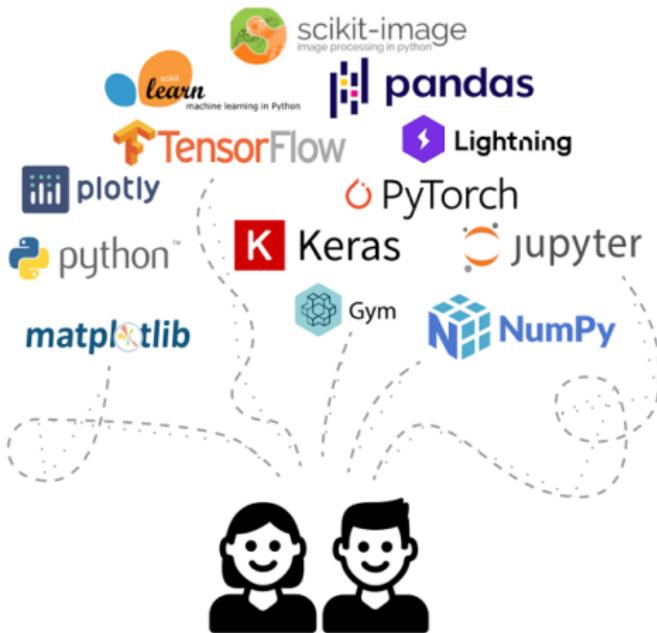
High-level interface for PyTorch, by William Falcon.
Lightning 2.0 is featuring a clean and stable API!!!

PyTorch

From Torch library
Supported by Facebook
BSD licence

Widely used in the field of AI research

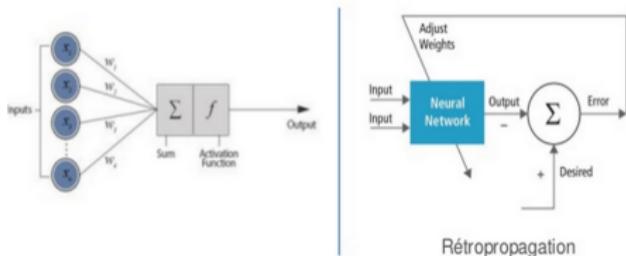
Complexité !!!



Comment apprend un réseau de neurones ?

Rétropropagation des erreurs

Apprentissage = déterminer les poids

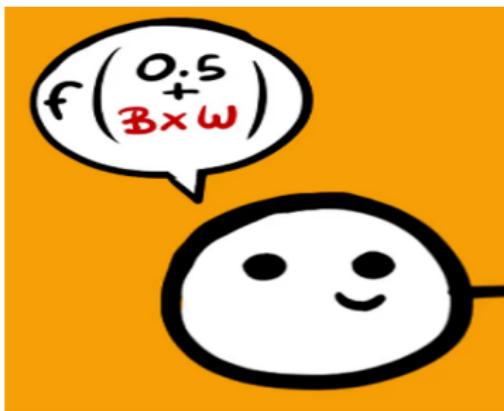


- 1 Présentation d'un motif d'entraînement au réseau
- 2 Comparaison de la sortie du réseau avec a sortie ciblée
- 3 Calcul de l'erreur en sortie de chacun des neurones du réseau.
- 4 Calcul, pour chacun des neurones, de la valeur de sortie qui aurait été correcte
- 5 Définition de l'augmentation ou de la diminution nécessaire pour obtenir cette valeur
- 6 Ajustement du poids de chaque connexion vers l'erreur locale.
- 7 Attribution d'un blâme à tous les neurones précédemment.
- 8 Recommencer à partir de l'étape 4, sur les neurones précédents, le blâme servant d'erreur

Fonction d'activation

Définition

Elle décide au neurone de passer ou pas l'information au prochain neurone, si la valeur passe, elle se rapproche de 1 et si elle ne passe pas, elle se rapproche de 0.



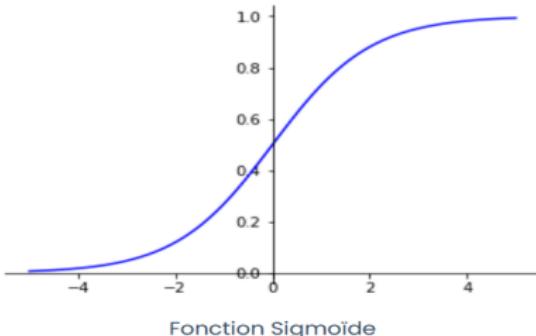
Fonction d'activation

Sigmoid

Définition

prend des valeur des préactivation en $[-\infty, +\infty]$ et donne une activation entre $[0, 1]$, sortie cohérente, exemple dectecter soit un 9 ou pas. On parle de probabilité, son gradient est très petit et diminue brutalement

$$\text{fonction_Sigmoid}(x) = 1 / (1 + \exp(-x))$$



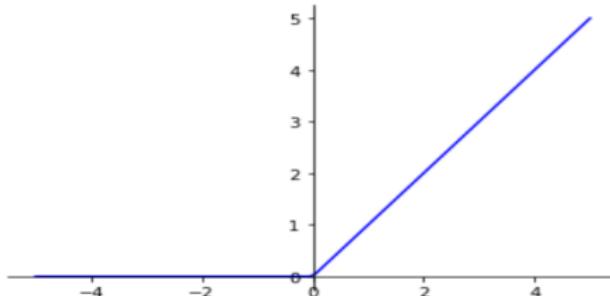
Fonction d'activation

RELU

Définition

Utilisés pour les hidden Layer. Elle marche pour les modèles non-linéaires, très bonnes propriétés de convergence par rapport à Sigmoid et Tanh, mais il faut faire attention lorsque le gradient est nul, donc il y aura risque de non-activation des certains neurones, alors il y aura de neurones qui ne servent à rien.

$$\text{fonction_ReLU}(x) = \max(x, 0)$$



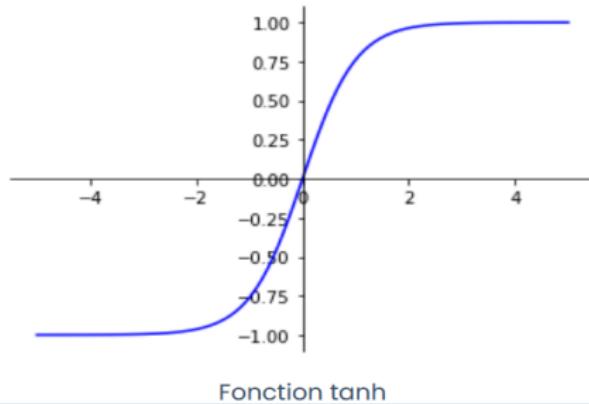
Fonction ReLU – Rectified Linear Unit

Fonction d'activation

TANH

Définition

S'appelle tangente hyperbolique, permet d'éviter les transitions d'échelle (il faut scalar), avoir seulement des valeurs entre $-1,1[$, exemple cas des angles, peut être négative. Son gradient est élevé et diminue doucement. Si on a le choix entre Tanh et Sigmoid, on utilise Tanh. On l'utilise aussi pour les fonctions non linéaire



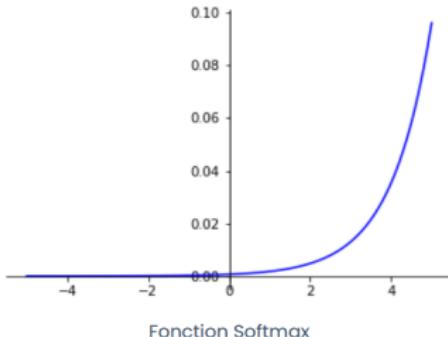
Fonction d'activation

Softmax

Définition

Distribution de probabilité, la somme est égale à 1, si par exemple le cas de reconnaissance de chiffre, il prend tous les valeurs de préactivation de neurones précédents : Déetecter seulement que c'est un neuf, et tous les autres (4,7,5...) sont négligeable. On l'utilise lorsqu'on a différentes classes.

$$\text{fonction_Softmax}(x) = \exp(x) / \sum(\exp(x_i))$$

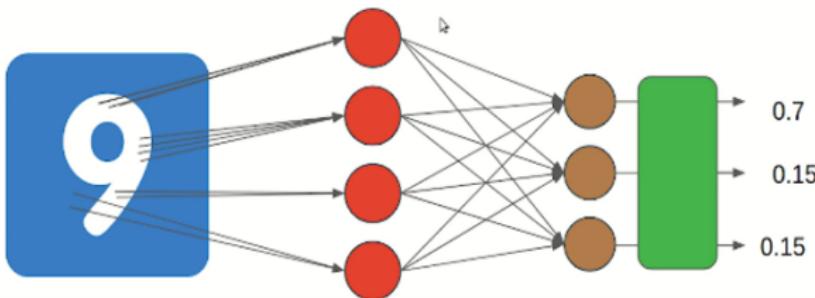


Fonction d'activation

Softmax

Définition

Distribution de probabilité, la somme est égale à 1, si par exemple le cas de reconnaissance de chiffre, il prend tous les valeurs de préactivation de neurones précédents : Déetecter seulement que c'est un neuf, et tous les autres (4,7,5...) sont négligeable. On l'utilise lorsqu'on a différentes classes.



Qualité de l'apprentissage

Mesure de la performance d'un réseau de neurones

- Durée d'apprentissage
- Précision des prédictions
- Capacité de généralisation
- Durée de prédiction

Optimisation stochastique

■ Contexte :

- ▶ Mise à jour des poids des paramètres du modèle (générateur et du discriminateur).
- ▶ Minimiser la fonction coût.
- ▶ Utilisation de mini-lots (**batches**) de données.
⇒ améliorer la performance du modèle

■ Méthodes d'optimisation stochastiques dans un cadre de GAN :

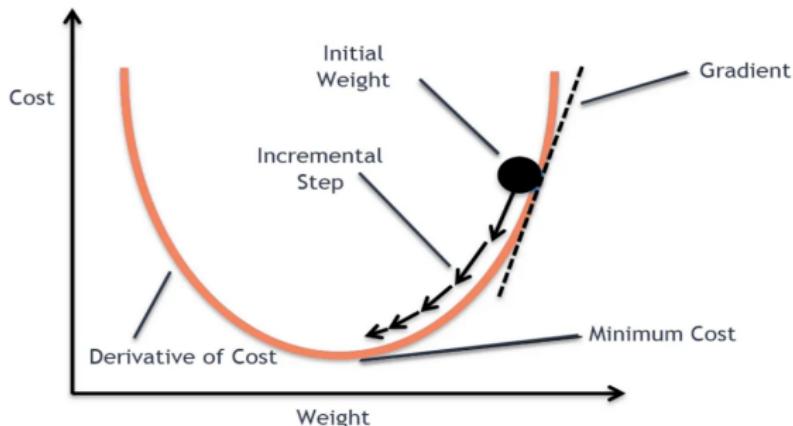
- ▶ GD (Gradient Descent)
- ▶ SGD (Stochastic Gradient Descent)
- ▶ Adam (Adaptive Moment Estimation)

Optimisation stochastique

Différence entre GD et SGD

Définition

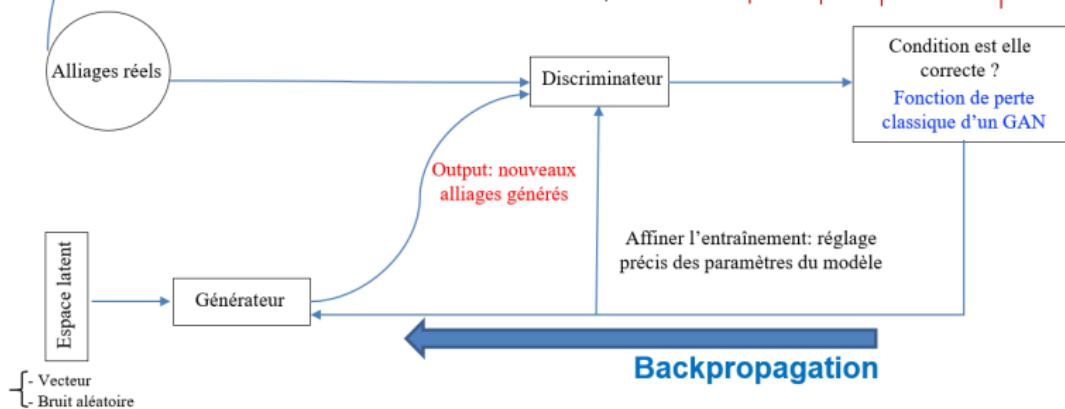
Dans **SGD** pour chaque itération, vous choisissez le lot qui contient généralement **l'échantillon aléatoire de données** mais dans le cas de **GD**, chaque itération contient la **totalité des observations de formation**.



Conception de nouveaux alliages

Exemple d'une base de données

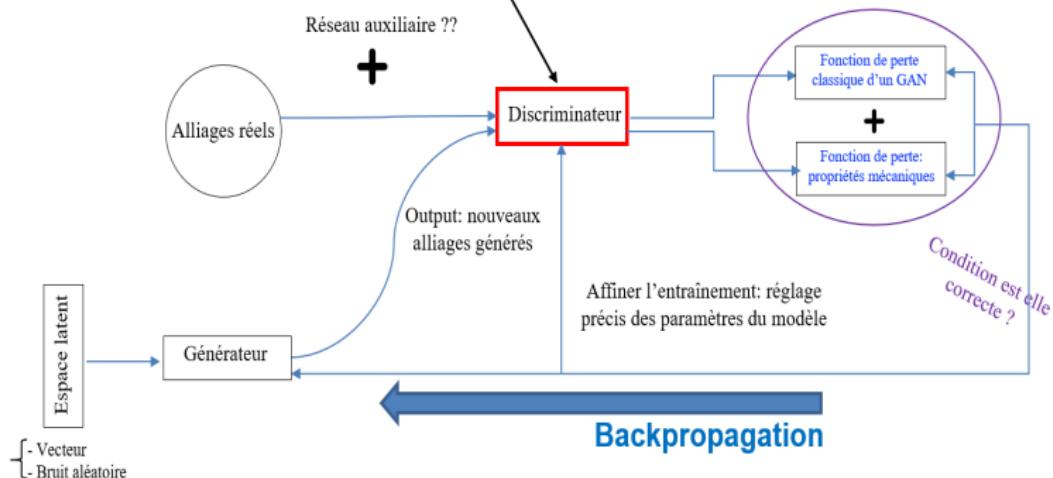
	Compositions Chimiques					Propriétés mécaniques			
	Compo 1	Compo 2	Compo 3	Compo 4	...	Compo n	Fatigue	Fluage	... Elasticité
Alliage 1	%	%	%	%		%	✓	✓	✓
Alliage 2	%	%	%	%		%	✓		✓
Alliage 3	%	%	%	%		%	✓	✓	✓
⋮									
Alliage n	%	%	%	%		%	✓	✓	✓



Conception de nouveaux alliages

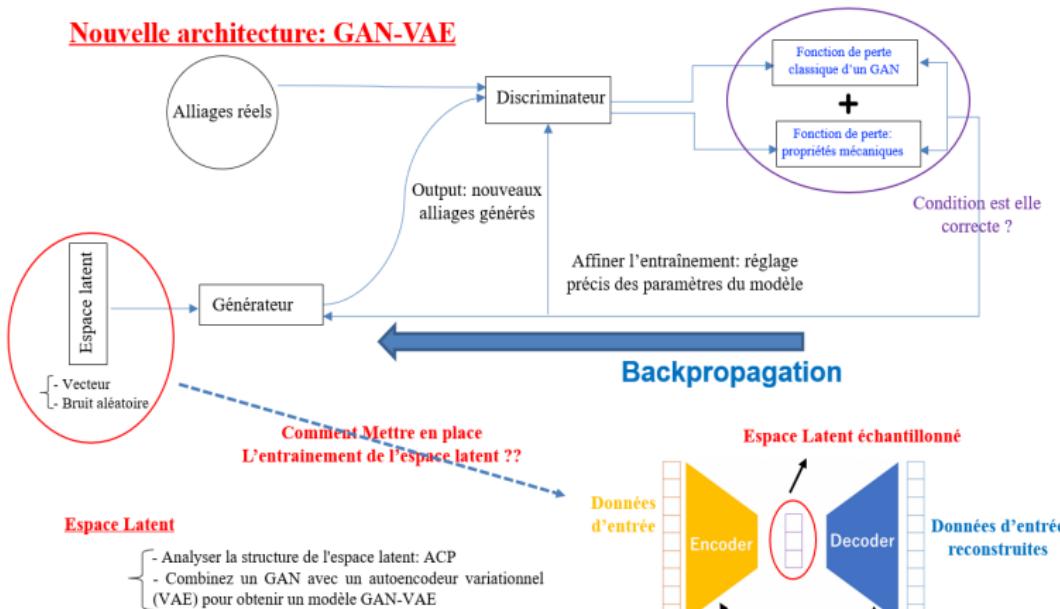
Modification de l'architecture du Discriminateur

- Ajouter des sorties supplémentaires pour les propriétés mécaniques à prédire
- Fonction de perte spécifique
- Pondération des pertes



Conception de nouveaux alliages

Nouvelle architecture: GAN-VAE



Ref: Liu, Y., Yang, Z., Yu, Z., Liu, Z., Liu, D., Lin, H., ... & Shi, S. (2023). Generative artificial intelligence and its applications in materials science: Current situation and future perspectives. *Journal of Materomics*.



Conception de nouveaux alliages

Inconvénients des Réseaux Antagonistes Génératifs (GAN)

- Entraînement instable et lent
 - Le générateur et le discriminateur sont constamment en compétition.
- Données d'entraînement limitées ? ?
 - Les GANs nécessitent un grand volume de données d'entraînement pour produire des bons résultats.
- Exigeant en ressources mémoires.
- Phénomène "**Mode collapse**".
 - Production d'un nombre limité de sorties au lieu de la variété souhaitée.

Conception de nouveaux alliages

■ Réseaux de neurones :

- ▶ Réseaux de neurones artificiels basées sur la physique : **PINNs (Physics-Informed Neural Networks)**
 - Un très bon candidat. Une nouvelle classe de réseaux de neurones qui hybride apprentissage automatique et lois physiques
- ▶ Réseaux de neurones convolutionnels (CNN) : Pour l'analyse de données structurées, comme des images de microstructures d'alliages.

■ Algorithmes génétiques :

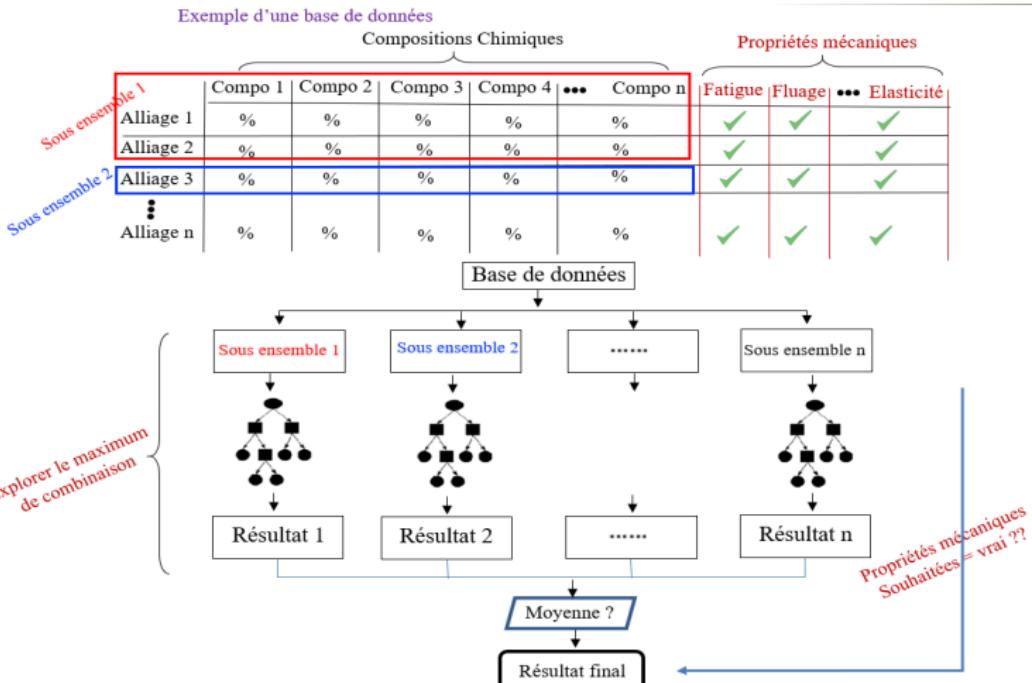
- Utilisez des algorithmes génétiques pour simuler l'évolution naturelle des alliages en ajustant leur composition de manière itérative pour atteindre des propriétés souhaitées.

Conception de nouveaux alliages

- Méthodes basées sur le renforcement :
 - Appliquez des approches de renforcement pour guider la recherche des compositions d'alliages en maximisant ou minimisant une fonction de perte basée sur les propriétés souhaitées.
- Optimisation bayésienne :
 - Appliquez des méthodes d'optimisation bayésienne pour explorer efficacement l'espace des paramètres des alliages et trouver des compositions prometteuses.
- XGBoost :
 - XGBoost utilise des arbres de décision comme modèles de base. Ces arbres sont construits de manière à maximiser la réduction de la fonction de perte.

Conception de nouveaux alliages

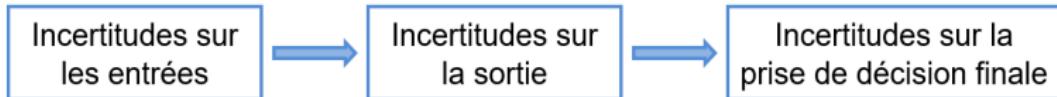
XGBoost



Incertitude

Contexte

- Incertitudes sur les paramètres d'entrée des modèles numériques



- Gestion des incertitudes :

- ▶ Comment propager les incertitudes sur la sortie du code/modèle ?
- ▶ Quelles sont les incertitudes les plus préjudiciables ?
- ▶ Quelles sont les variables les plus influentes

Incertitude

Introduire de l'incertitude dans un modèle ML

■ GAN avec une distribution latente Gaussienne (GAN-Gaussien) :

- ▶ Distribution latente définie comme une distribution gaussienne
- ▶ Cette distribution gaussienne est caractérisée par une moyenne et une covariance (ou écart-type)
⇒ Permettre de modéliser l'incertitude associée à chaque point.
- ▶ L'utilisation d'une distribution gaussienne peut aider à caractériser l'incertitude intrinsèque.
- ▶ Les échantillons générés peuvent varier en fonction de l'incertitude associée à l'espace latent

Référence :

⇒ Di Martino, A., Bodin, E., Ek, C. H., & Campbell, N. D. (2019). Gaussian process deep belief networks : A smooth generative model of shape with uncertainty propagation. In Computer Vision–ACCV 2018 : 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part IV 14 (pp. 3-20). Springer International Publishing



Dessine moi un mouton

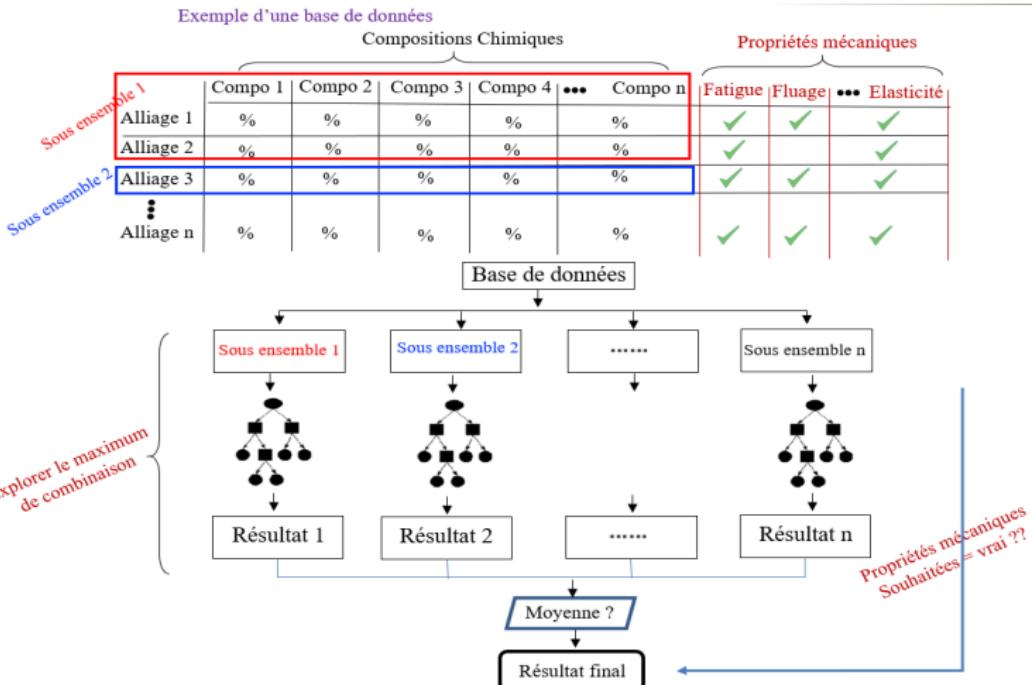


Conclusions

- L'ONERA mène des recherches appliquées ayant vocation à appuyer les industriels de l'aéronautique et de l'espace
- le rôle de la simulation numérique n'est pas de se substituer totalement aux essais, mais d'en être complémentaires
- Il y a encore de gros challenge à relever
 - élaborer de nouveaux matériaux
 - intégrer la science des données
 - exploiter encore mieux les ordinateurs parallèles
 - développer les simulations multiphysiques
 - ...

Conception de nouveaux alliages

XGBoost : utilisé dans le projet ADAMANT



1D normal distribution

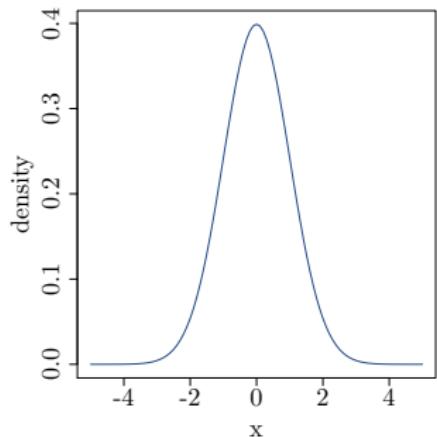
We say that $X \sim \mathcal{N}(\mu, \sigma^2)$ if it has the following pdf :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

The distribution is characterised by

mean : $\mu = E[X]$

variance : $\sigma^2 = E[X^2] - E[X]^2$



One fundamental property : a linear combination of independant normal distributed random variables is still normal distributed.

Multivariate normal distribution

Definition

We say that a vector $Y = (Y_1, \dots, Y_n)^t$ follows a multivariate normal distribution if any linear combination of Y follows a normal distribution :

$$\forall \alpha \in \mathbb{R}^n, \alpha^t Y \sim \mathcal{N}$$

The distribution of a Gaussian vector is characterised by

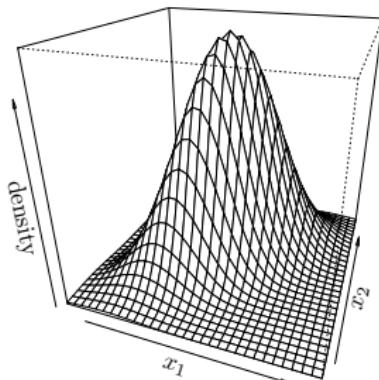
- a **mean vector** $\mu = E[Y]$
- a **covariance matrix** $\Sigma = E[YY^t] - E[Y]E[Y]^t$

Multivariate normal distribution

The density of a multivariate Gaussian is :

$$f_Y(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (x - \mu)^t K^{-1} (x - \mu) \right)$$

2D multivariate Gaussian distribution :



Gaussian processes

The multivariate Gaussian distribution can be generalised to random processes :

Definition

A random process Z over $D \subset \mathbb{R}^d$ is said to be Gaussian if

$\forall n \in \mathbb{N}, \forall x_i \in D, (Z(x_1), \dots, Z(x_n))$ is a Gaussian vector.

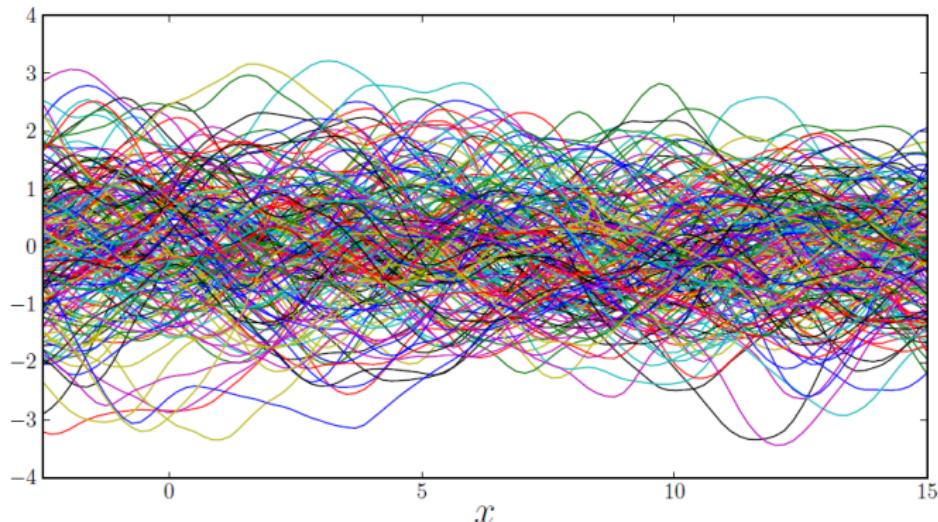
The distribution of a GP is fully characterised by :

- its mean function m defined over D
- its covariance function (or kernel) k defined over $D \times D$:
 $k(x, y) = \text{cov}(Z(x), Z(y))$

We will use the notation $Z \sim \mathcal{N}(m(\cdot), k(\cdot, \cdot))$.

Gaussian processes

Examples of a Gaussian process sample paths :



What would be the next step ?

Let Z be a random process with kernel k . Some properties of kernels can be obtained from their definition.

Example

$$\begin{aligned} k(x, x) &= \text{cov}(Z(x), Z(x)) = \text{var}(Z(x)) \geq 0 \\ \Rightarrow k(x, x) &\text{ is positive} \end{aligned}$$

$$\begin{aligned} k(x, x) &= \text{cov}(Z(x), Z(y)) = \text{cov}(Z(y), Z(x)) = k(y, x) \\ \Rightarrow k(x, x) &\text{ is symmetric} \end{aligned}$$

We can obtain a thinner result...

We introduce the random variable $T = \sum_{i=1}^n a_i Z(x_i)$ where n , a_i and x_i are arbitrary. Computing the variance of T gives :

$$\begin{aligned}\text{var}(T) &= \text{cov}\left(\sum_i a_i Z(x_i), \sum_{j=1}^n a_j Z(x_j)\right) = \sum_i \sum_j a_i a_j \text{cov}(Z(x_i), Z(x_j)) \\ &= \sum_i \sum_j a_i a_j k(x_i, x_j)\end{aligned}$$

since a variance is positive, we have

$$\sum_i \sum_j a_i a_j k(x_i, x_j) \geq 0$$

for any arbitrary n , a_i and x_i

Definition

The function satisfying the above inequality **for all** $n \in \mathbb{N}$, **for all** $x_i \in D$, **for all** $a_i \in \mathbb{R}$ are called **positive semi-definite functions**.

We have just seen :

k is a covariance $\Rightarrow k$ is a positive semi-definite function

The reverse is also true :

Theorem (Loeve)

k corresponds to the covariance of a GP



k is a symmetric positive semi-definite function

Making new from old

Kernels can be :

- Summed together

- ▶ On the same space $k(x, y) = k_1(x, y) + k_2(x, y)$
 - ▶ On the tensor space $k(x, y) = k_1(x_1, y_1) + k_2(x_2, y_2)$

- Multiplied together

- ▶ On the same space $k(x, y) = k_1(x, y) \times k_2(x, y)$
 - ▶ On the tensor space $k(x, y) = k_1(x_1, y_1) \times k_2(x_2, y_2)$

- Composed with a function

- ▶ $k(x, y) = k_1(f(x), f(y))$

All these operations will preserve the positive definiteness.

How can this be useful ?

Composition with a function :

Property

Let k_1 be a kernel over $D_1 \times D_1$ and f be an arbitrary function $D \rightarrow D_1$, then

$$k(x, y) = k_1(f(x), f(y))$$

is a kernel over $D \times D$.

proof

$$\sum \sum a_i a_j k(x_i, x_j) = \sum \sum a_i a_j k_1(\underbrace{f(x_i)}_{y_i}, \underbrace{f(x_j)}_{y_j}) \geq 0$$

Remarks :

- k corresponds to the covariance of $Z(x) = Z_1(f(x))$
- This can be seen as a (nonlinear) rescaling of the input space

For a given set of observations, the model is fully determined by the kernel. A kernel satisfies the following properties :

- It is symmetric : $k(x, y) = k(y, x)$
- It is positive semi-definite (psd) :

$$\forall n \in \mathbb{N}, \forall x_i \in D, \forall \alpha \in \mathbb{R}^n, \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(x_i, x_j) \geq 0$$

Furthermore any symmetric psd function can be seen as the covariance of a Gaussian process. This equivalence is known as the Loeve theorem.

Kernels

Here is a list of the most common kernels in higher dimension :

constant $k(x, y) = \sigma^2$

white noise $k(x, y) = \sigma^2 \delta_{x,y}$

exponential $k(x, y) = \sigma^2 \exp(-\|x - y\|_\theta)$

Matern 3/2 $k(x, y) = \sigma^2 (1 + \sqrt{3}\|x - y\|_\theta) \exp(-\sqrt{3}\|x - y\|_\theta)$

Matern 5/2 $k(x, y) = \sigma^2 (1 + \sqrt{5}\|x - y\|_\theta + \frac{5}{3}\|x - y\|_\theta^2) \exp(-\sqrt{5}\|x - y\|_\theta)$

Gaussian $k(x, y) = \sigma^2 \exp(-\frac{1}{2}\|x - y\|_\theta^2)$

Where

$$\|x - y\|_\theta = \left(\sum_{i=1}^d \frac{(x_i - y_i)^2}{\theta_i^2} \right)^{1/2}$$

The parameter σ^2 is called the **variance** and θ the **length-scale**.

A common general recipee : a product of univariate kernels,

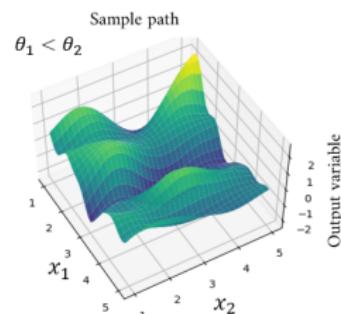
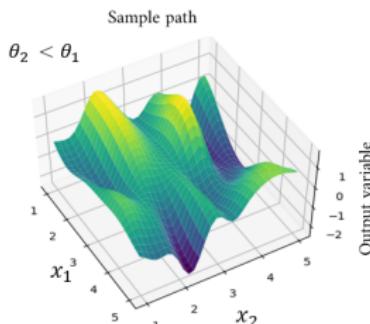
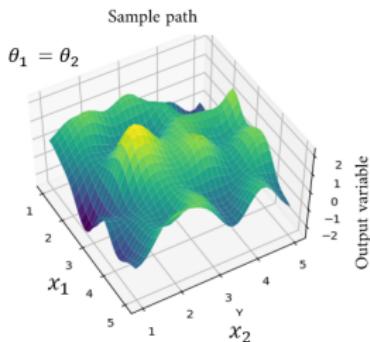
$$k(x, y) = \sigma^2 \prod_{i=1}^d r_i(x_i, y_i)$$

which has $d + 1$ parameters and $r(,)$ the correlation function.

effect of θ , Gaussian Kernel :

$$k(x, y) = \sigma^2 \exp\left(-\frac{1}{2}||x - y||_\theta^2\right) \text{ where } ||x - y||_\theta = \left(\sum_{i=1}^d \frac{(x_i - y_i)^2}{\theta_i^2}\right)^{1/2}$$

$$\begin{cases} \theta_i & \in \{\theta_1, \theta_2\} : \text{length-scale} \\ \sigma^2 & : \text{variance} \end{cases}$$



Exercice :

- Using a Gaussian kernel and a nominal value (mean value), find the analytical form of a path in a 2D case.
- Use a programming language of your preference to implement this path.

Thank you for your attention !

Any questions ?

www.onera.fr