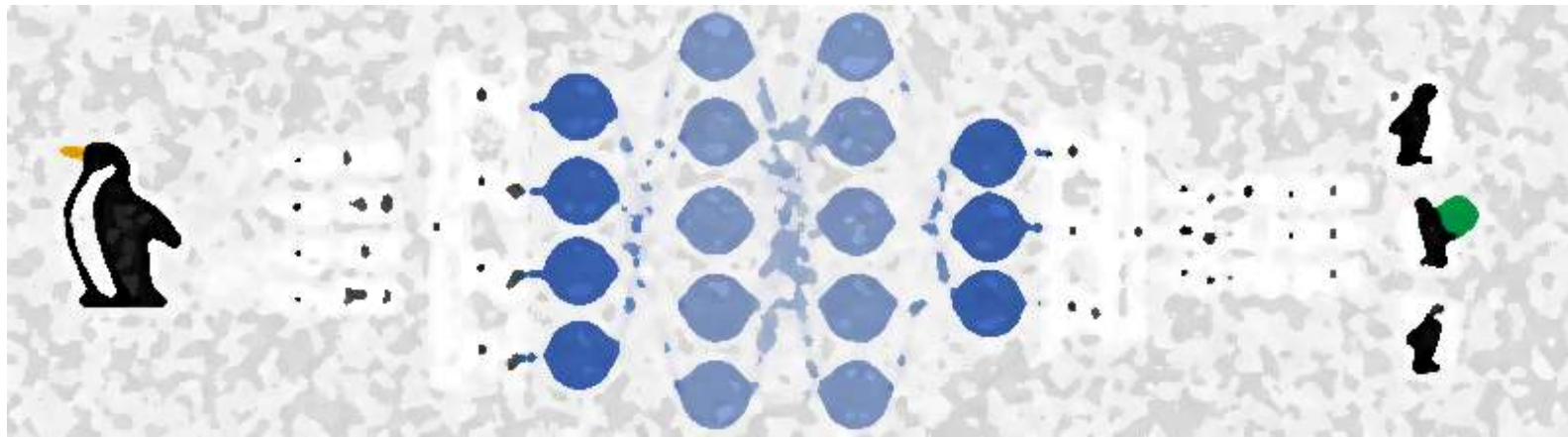


# Machine Learning démystifié: principes basiques

Avril 2024



# Sommaire

-  Principes basiques du Machine Learning
-  Plateforme Data Science et IA à l'OCSIN
-  Exemple à l'Etat de Genève
-  Références

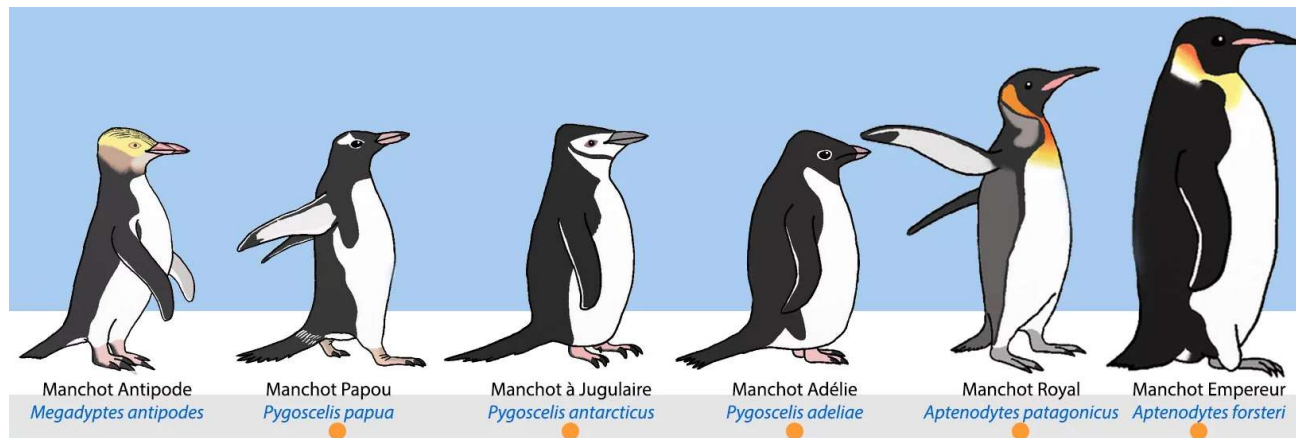
# Principes basiques du Machine Learning

## Définition

Le Machine Learning (ML) utilise la modélisation mathématique pour prédire des résultats à partir de données d'observations passées.

### Par exemple:

Utiliser des observations passées pour identifier des espèces de manchots ("Adélie", "Empereur" ou "à Jugulaire") en fonction des dimensions des nageoires, becs et autres.

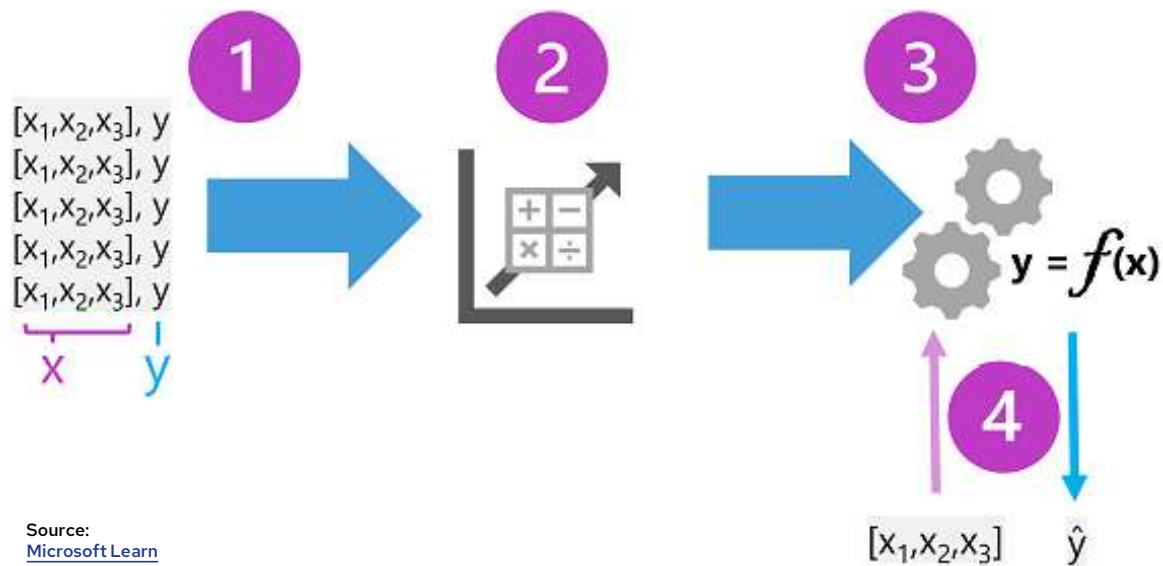


Source:  
[Latitud Argentina](#)

# Principes basiques du Machine Learning

## Définition (suite)

Un modèle ML est une fonction qui retourne une valeur de sortie à partir de valeurs d'entrée. Le processus pour définir la fonction est appelé "entraînement" ou "apprentissage" (learning). Une fois définie, la fonction est utilisée pour prédire des valeurs lors du processus appelé "inférence".



# Principes basiques du Machine Learning

① Les valeurs d'entrée ont  $n$  caractéristiques et la valeur de sortie que le modèle doit apprendre à prédire est appelée "étiquette" (tag). On regroupe les caractéristiques dans un vecteur  $x=[x_1, \dots, x_n]$  et on note l'étiquette  $y$ .

Par exemple:

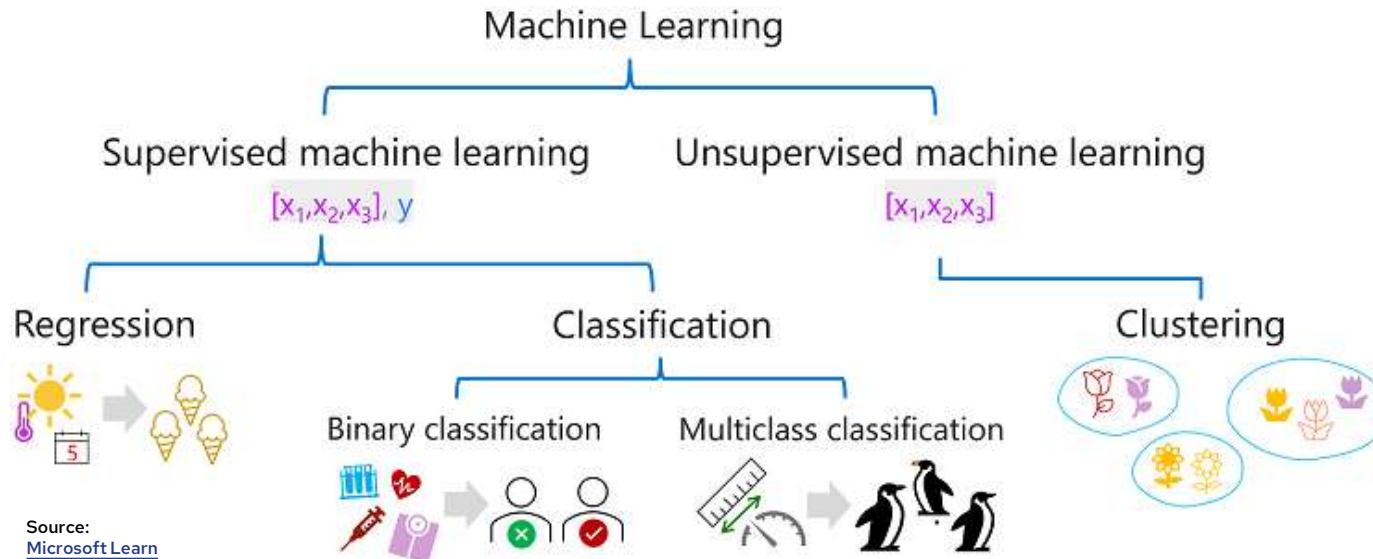
Les mesures du manchot sont les caractéristiques de  $x$  et l'espèce du manchot est l'étiquette  $y$  (0 pour "Adélie", 1 pour "Empereur", 2 pour "à Jugulaire").

② On applique un algorithme aux données pour déterminer une règle de calcul donnant  $y$  à partir de  $x$ .

③ Le résultat est un modèle entraîné qui encode la règle de calcul dans une fonction  $y=f(x)$ .

④ La fonction  $f$  est utilisée pour l'inférence. Sa valeur de sortie est notée  $\hat{y}$  car c'est une prédiction, et non une valeur "réelle".

# Principes basiques du Machine Learning



- ❑ Le ML supervisé est une forme d'apprentissage automatique consistant à apprendre une fonction de prédiction à partir d'exemples étiquetés.
- ❑ Le ML non supervisé est une forme d'apprentissage automatique consistant à apprendre une fonction de prédiction à partir d'exemples non étiquetés.

# Principes basiques du Machine Learning

Le clustering est une forme de ML non supervisé dans laquelle les données sont regroupées en clusters sur la base de similitudes de caractéristiques.

Par exemple:

On observe un échantillon de fleurs et on note le nombre de pétales et feuilles de chaque fleur.

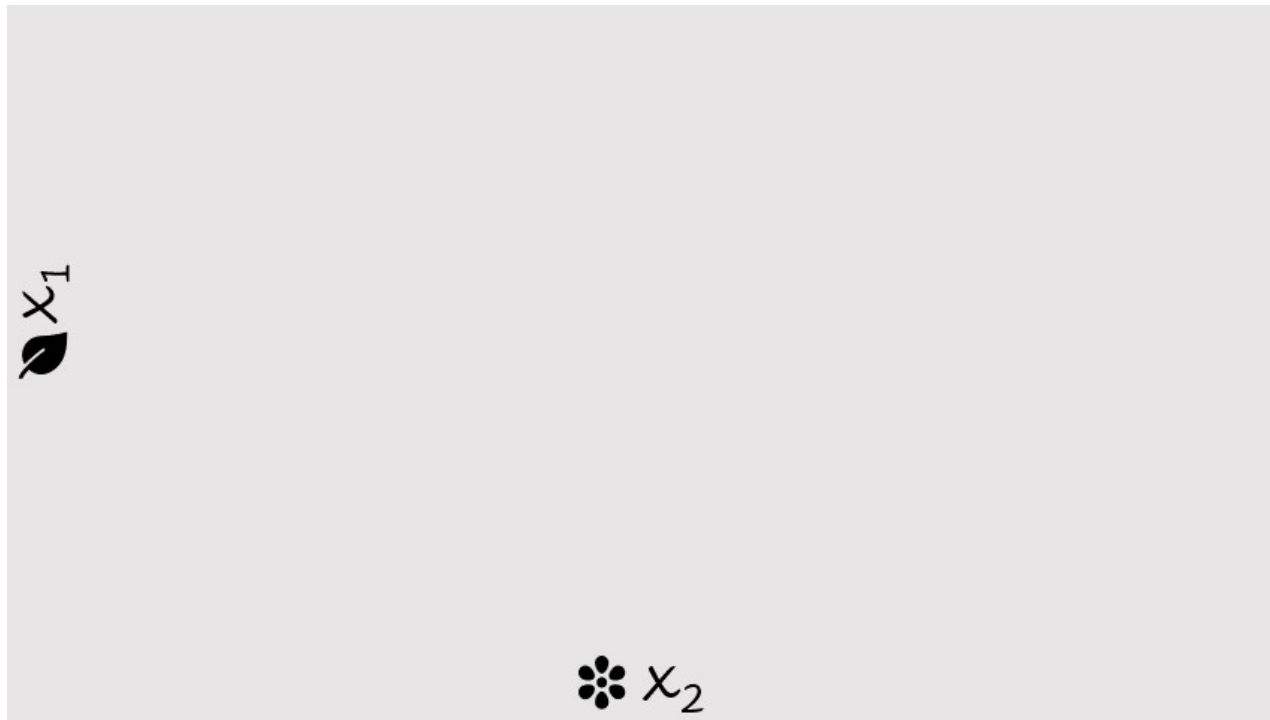


Source:  
[Microsoft Learn](#)

L'objectif n'est pas d'identifier les différentes espèces, mais simplement de regrouper des fleurs semblables en fonction du nombre de feuilles et pétales.

# Principes basiques du Machine Learning

Étant donnés des points et un entier  $k$ , l'algorithme de clustering itératif  $k$ -means permet de répartir les points en  $k$  clusters en minimisant pour chaque cluster la distance entre points et barycentre du cluster.



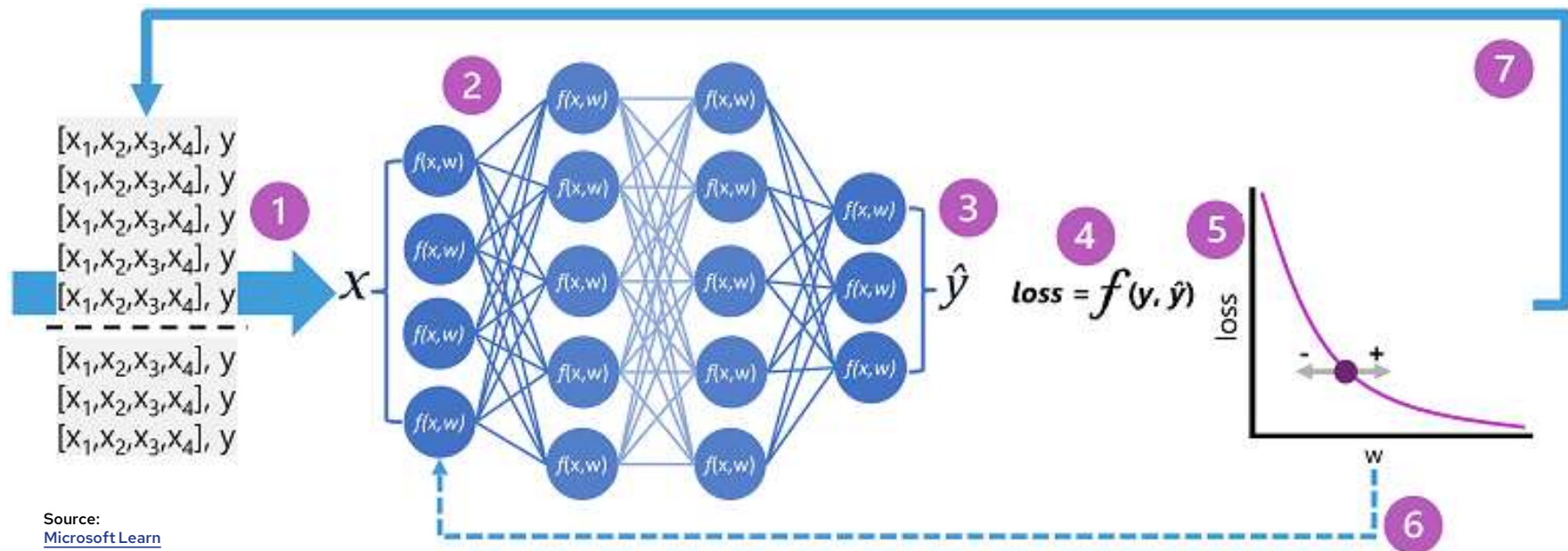
Source:  
[Microsoft Learn](#)



# Principes basiques du Machine Learning

Le Deep Learning est du ML basé sur un réseau neuronal artificiel profond (deep) qui simule l'activité neuronale humaine avec des fonctions mathématiques.

Pendant l'entraînement, le modèle apprend des paramètres numériques (poids  $w$ ) permettant les prédictions. Dans le cas supervisé, le processus schématisé est le suivant:

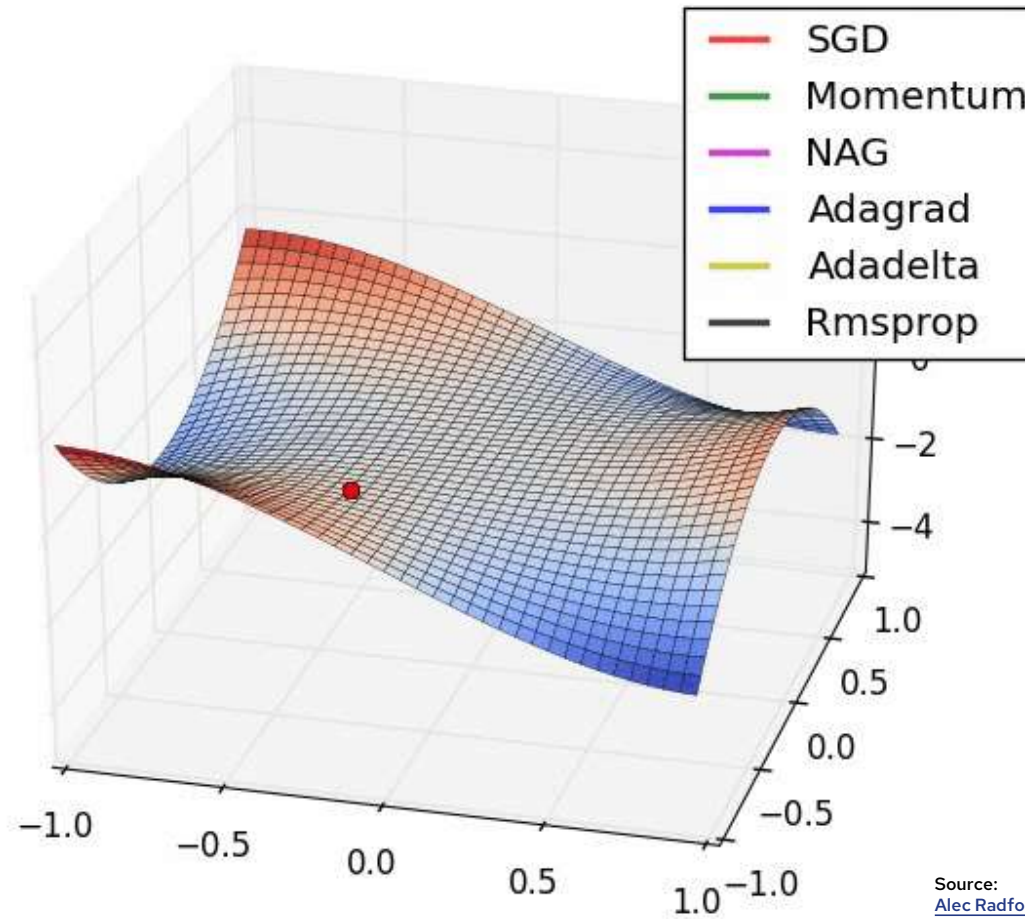


Source:  
[Microsoft Learn](#)

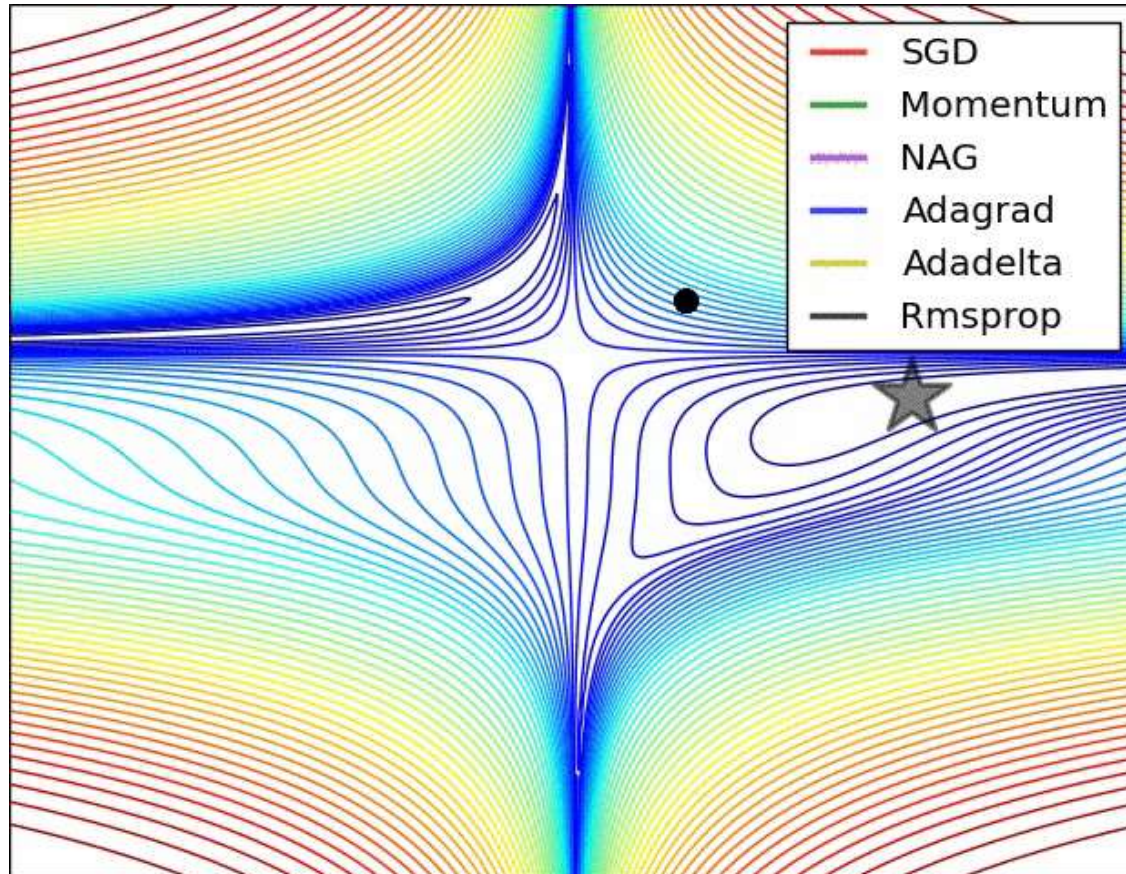
# Principes basiques du Machine Learning

- ① Les données d'entraînement (variables  $x$ ) sont définies. Les poids  $w$  sont initialisés de manière aléatoire. Les variables d'entraînement sont injectées dans la couche d'entrée.
- ② Les neurones de chaque couche agissent sur leurs valeurs d'entrée en appliquant une fonction linéaire dont les paramètres sont les poids  $w$ , puis alimentent les neurones de la couche suivante.
- ③ La couche finale produit un vecteur  $\hat{y}$  contenant les valeurs prédites.
- ④ Une fonction de perte  $f(y, \hat{y})$  (loss function) est utilisée pour calculer la distance entre valeurs réelles  $y$  et valeurs prédites  $\hat{y}$  (erreur quadratique  $f(y, \hat{y}) = \sum_i (y_i - \hat{y}_i)^2$  ou autre).
- ⑤ On optimise les poids pour que  $f(y, \hat{y})$  soit minimal, i.e. on cherche une vallée sur une surface de grande dimension (la dimension est le nombre de poids). Ceci se fait en se déplaçant dans la direction où la pente est négative (gradient descent).
- ⑥ Les modifications apportées aux poids sont rétropropagées dans les couches du réseau et viennent remplacer les valeurs antérieures.
- ⑦ Le processus est répété jusqu'à ce que la perte soit minimale et la prédiction  $\hat{y}$  acceptable.

# Principes basiques du Machine Learning



# Principes basiques du Machine Learning



Source:  
[Alec Radford](#)

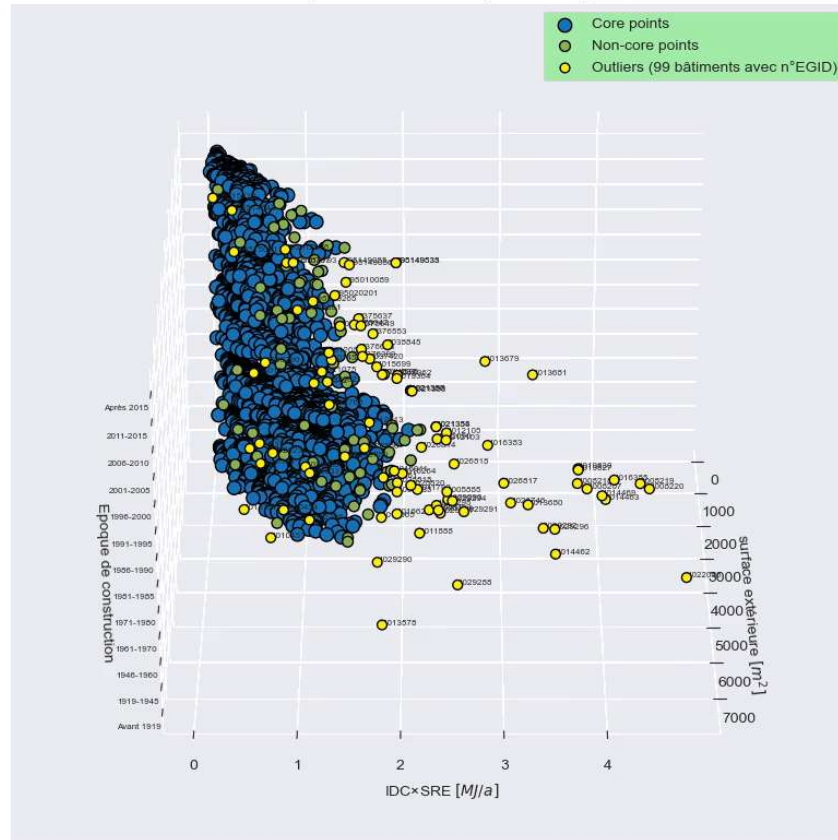
# Plateforme Data Science et IA à l'OCSIN

La plateforme de data science et IA est l'ensemble des processus et ressources IT (hardwares, softwares, librairies python,...) permettant au secteur INGDD la réalisation des projets data science et IA; ce depuis la réception initiale d'une demande métier jusqu'au déploiement final de la solution.



# Exemple à l'Etat de Genève

Clustering DBSCAN pour (surface extérieure, IDC×SRE, époque de construction) en 2022 pour les habitations à plusieurs logements





OCSIN | Département des Institutions et du Numérique

OCSIN

 Rafael Tiedra

 rafaël.tiedra@\*\*\*\*.\*\*.\*\*

# Références

[1] Fondamentaux du Machine Learning (Microsoft)

[2] A Non-technical Introduction to Machine Learning (Olivier Colliot)

[3] A journey into Optimization algorithms for Deep Neural Networks (Sergios Karagiannakos)

[4] Alec Radford's animations for optimization algorithms (Deniz Yuret's Homepage)

[5] Cafés de l'innovation: Machine Learning et indice IDC (Rafael Tiedra)