

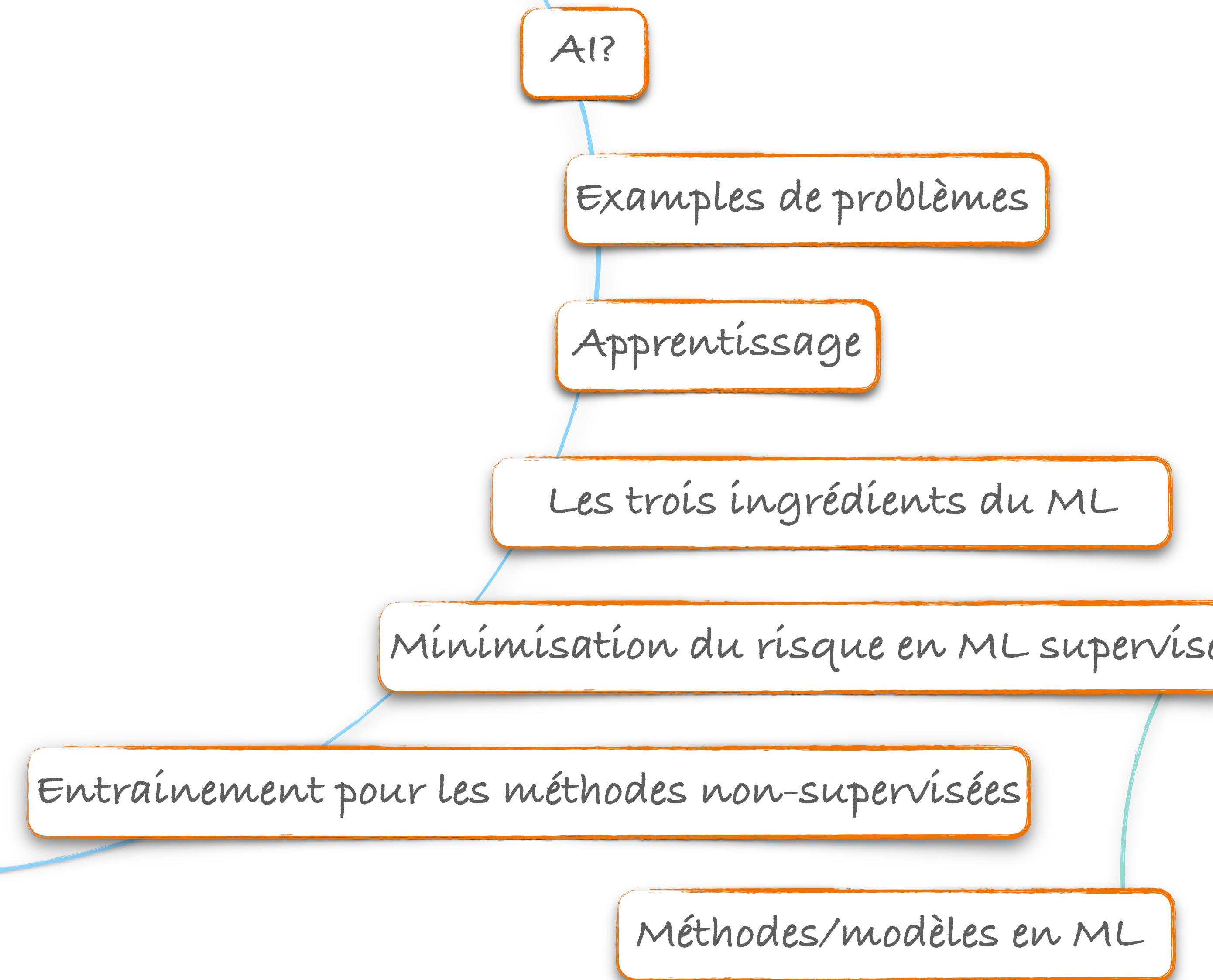
Applied QML

Lecture 2 - Machine Learning Classique

Christophe Pere

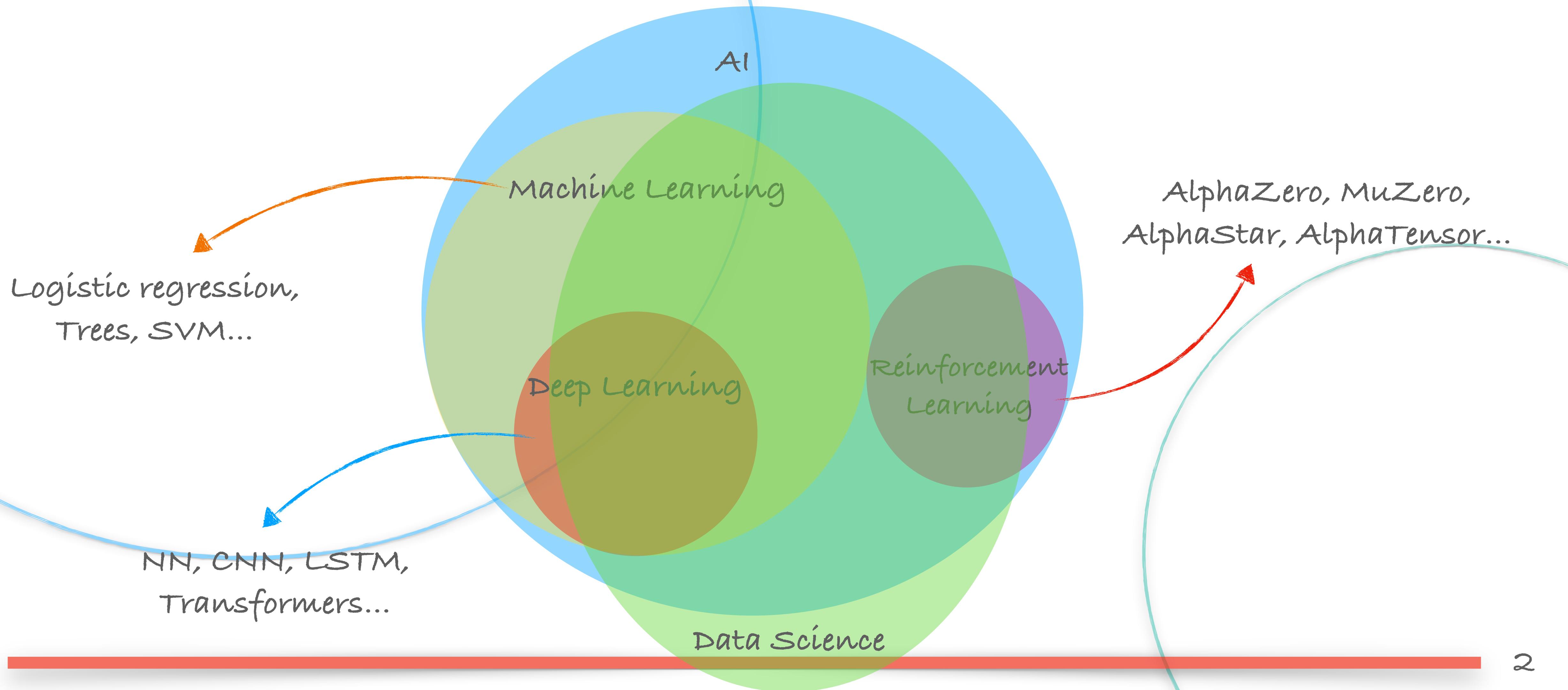
2024-01-11

Table des matières



AI?

AI?



Examples de problèmes

Exemples de problèmes

Reconnaissance d'images

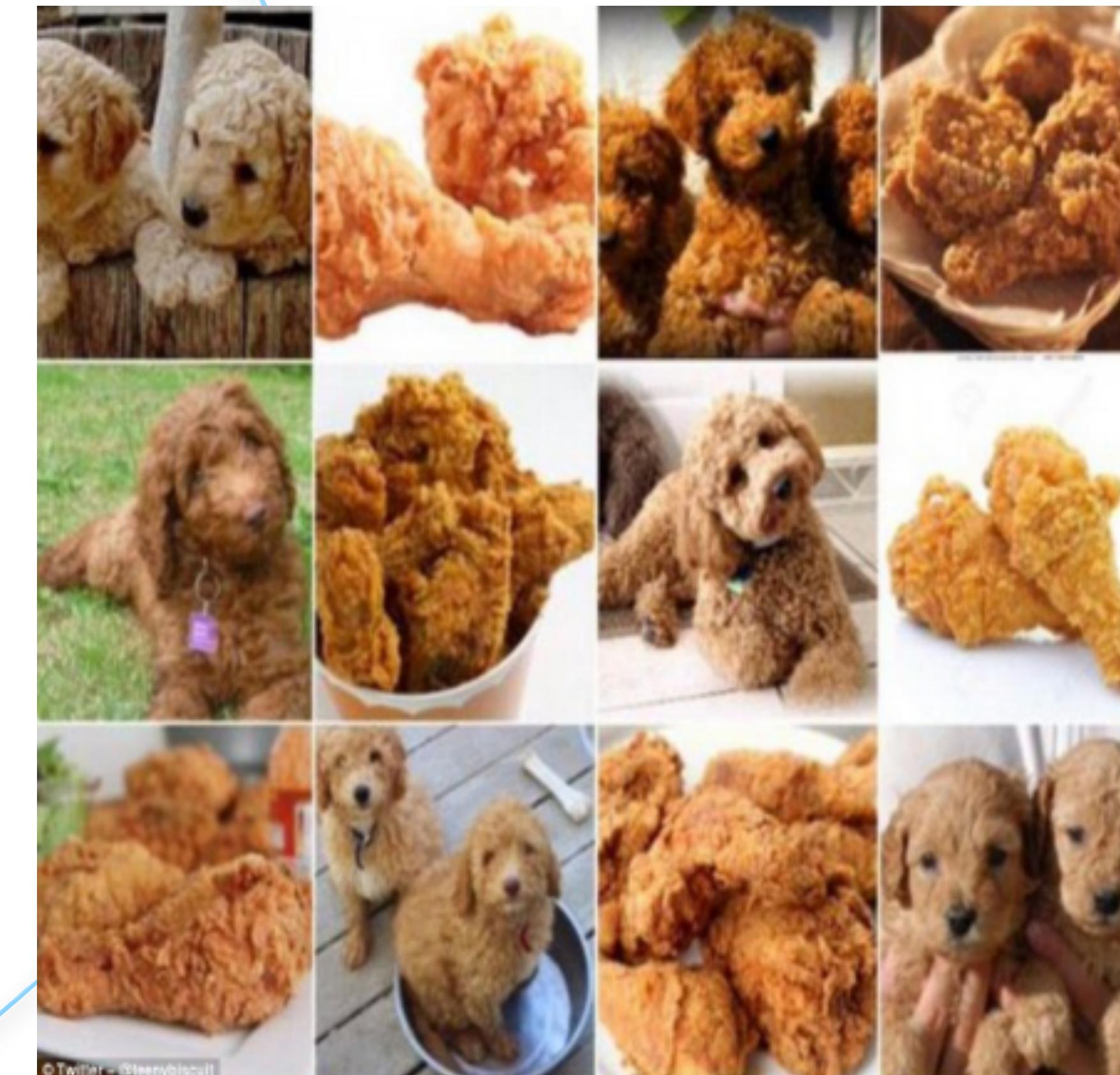
Muffin ou chihuahua?



Exemples de problèmes

Reconnaissance d'images

Poulet frit ou labrador?



Examples de problèmes

Reconnaissance d'images

Glace ou dalmatien?



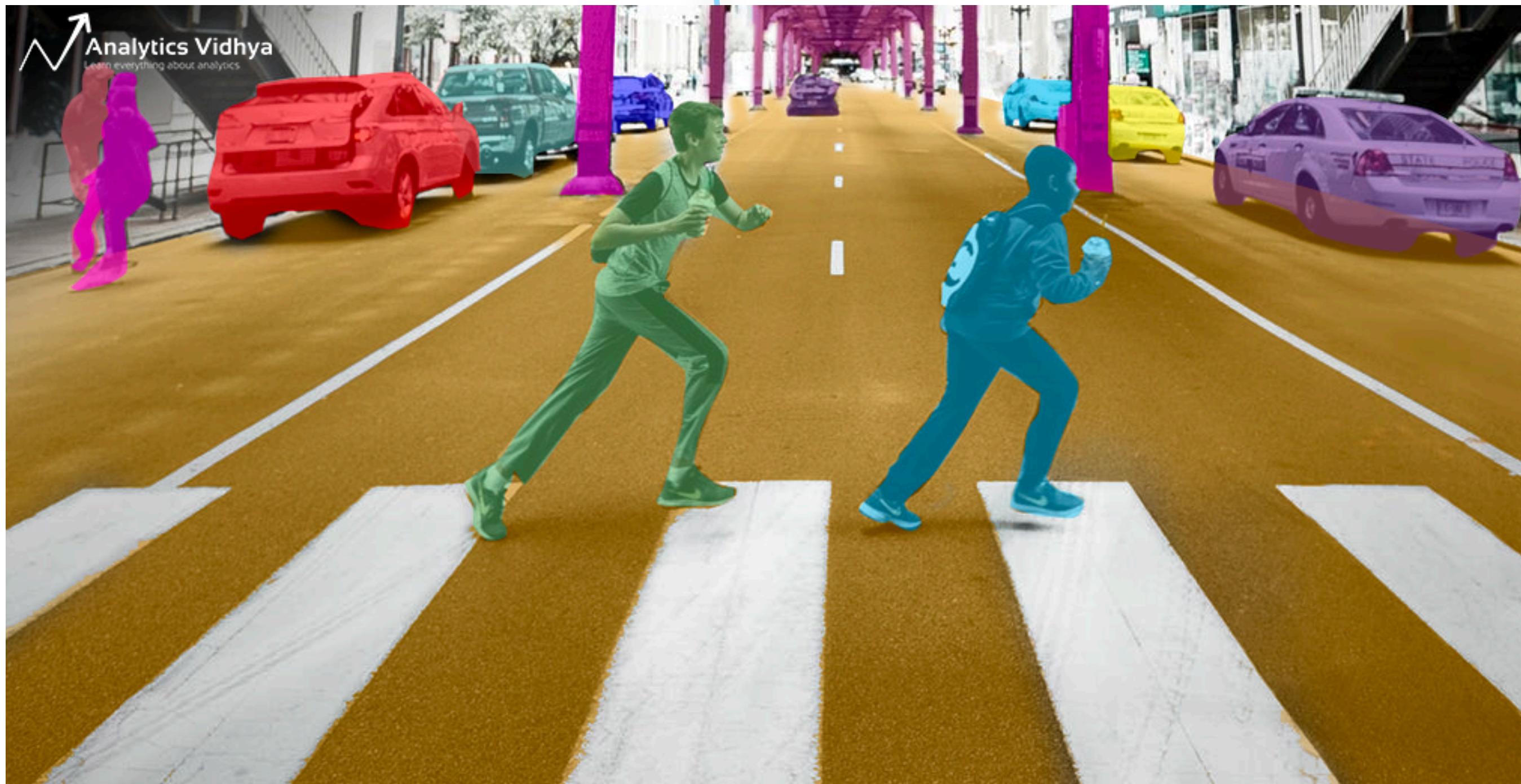
Examples de problèmes

Reconnaissance d'images

Segmentation d'images

Instance

Semantic



Examples de problèmes

Reconnaissance d'images

Segmentation d'images

Instance

Semantic



Examples de problèmes

Reconnaissance d'images

Segmentation d'images

Instance

Semantic



Exemples de problèmes

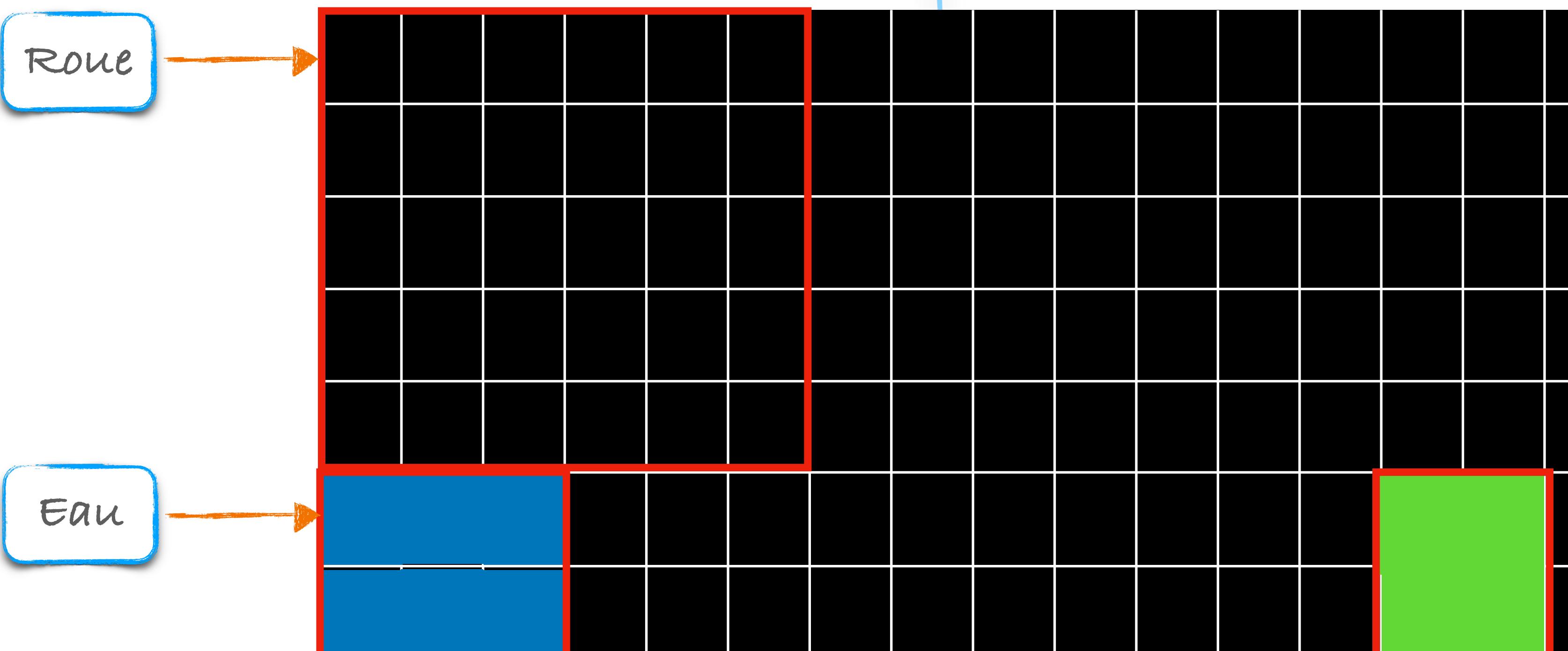
Reconnaissance d'images

Segmentation d'images

Instance

Semantic

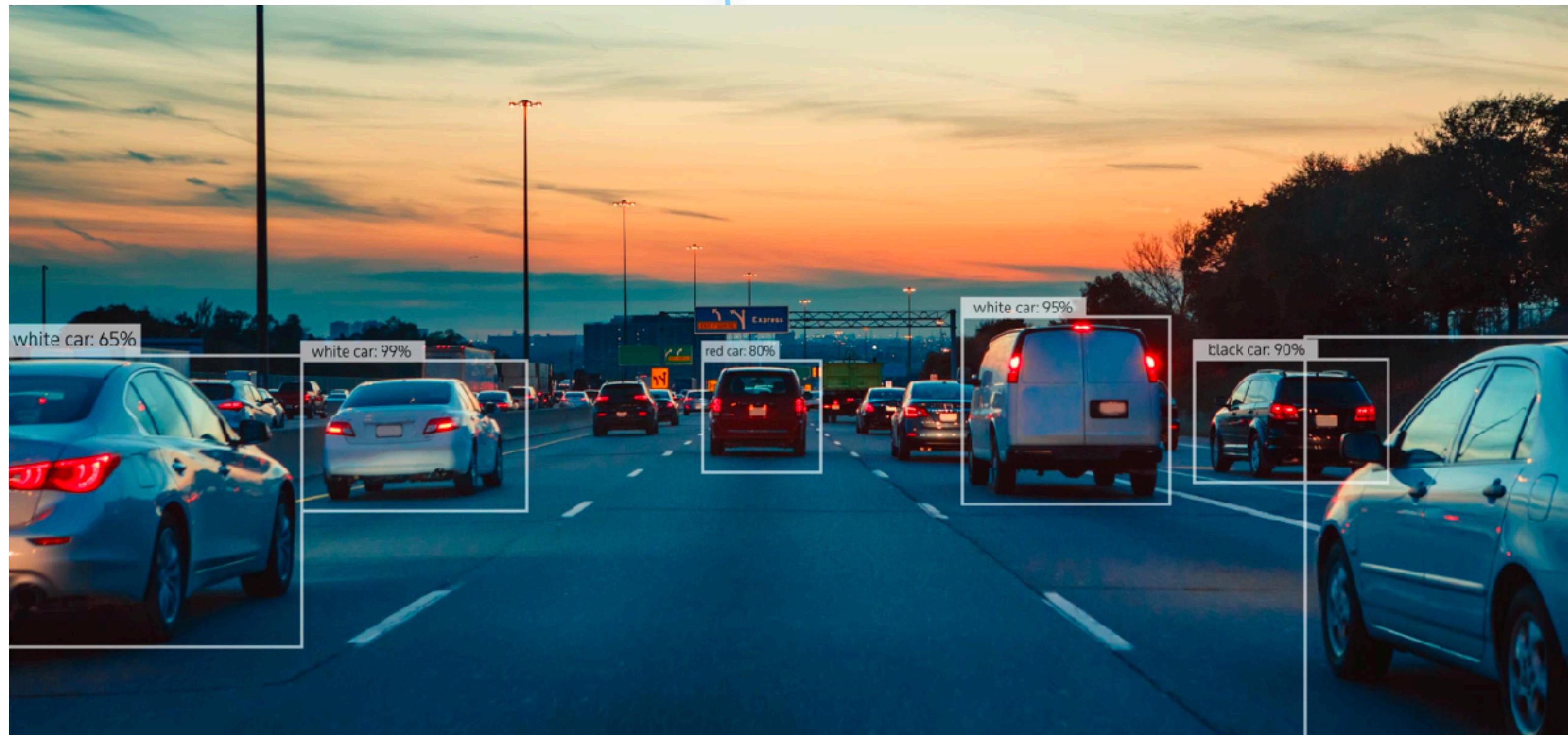
Comment coder ça pour une machine?



Examples de problèmes

Reconnaissance d'images

Reconnaissance vidéo



Exemples de problèmes

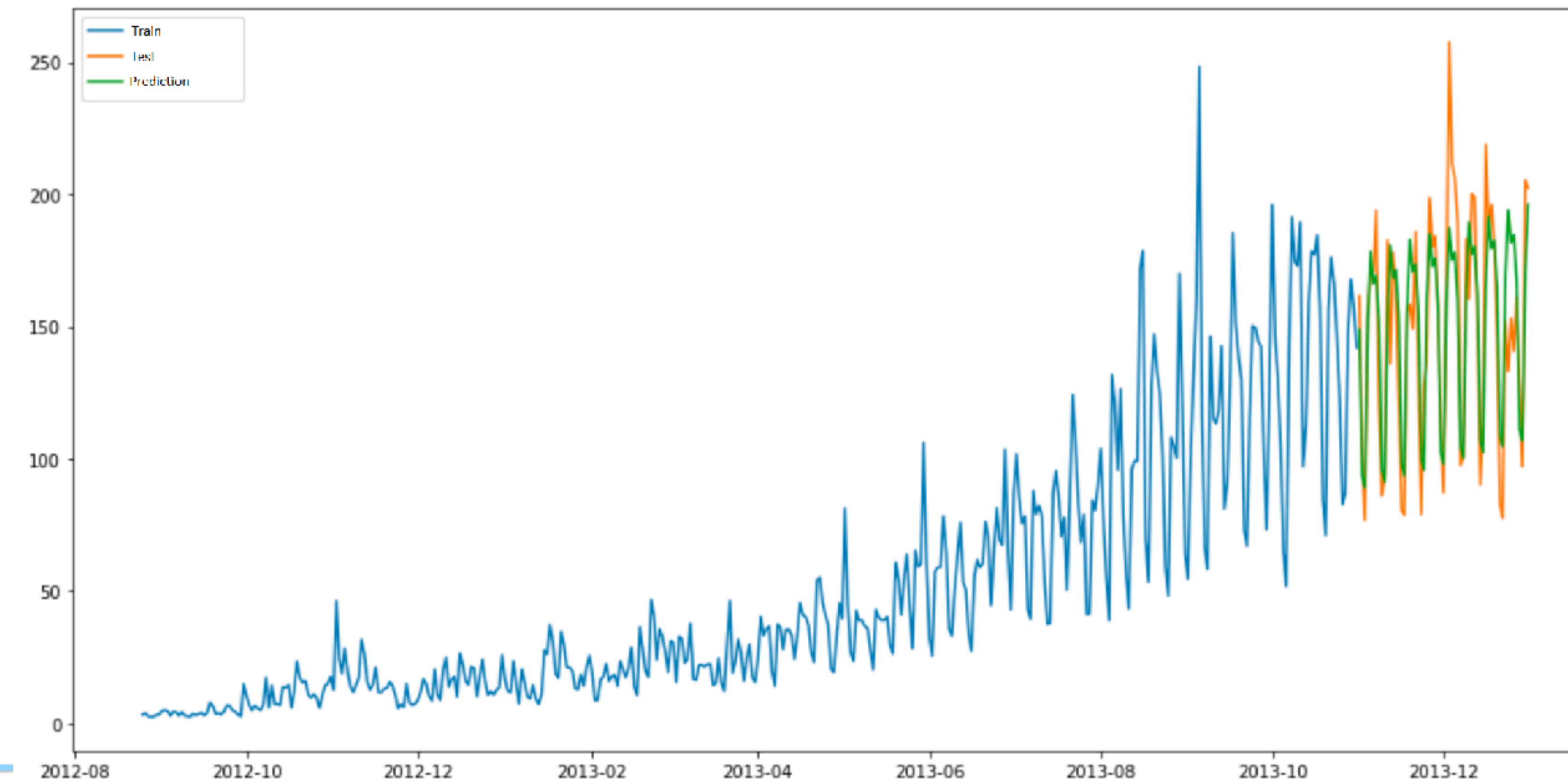
Reconnaissance d'images

Reconnaissance vidéo



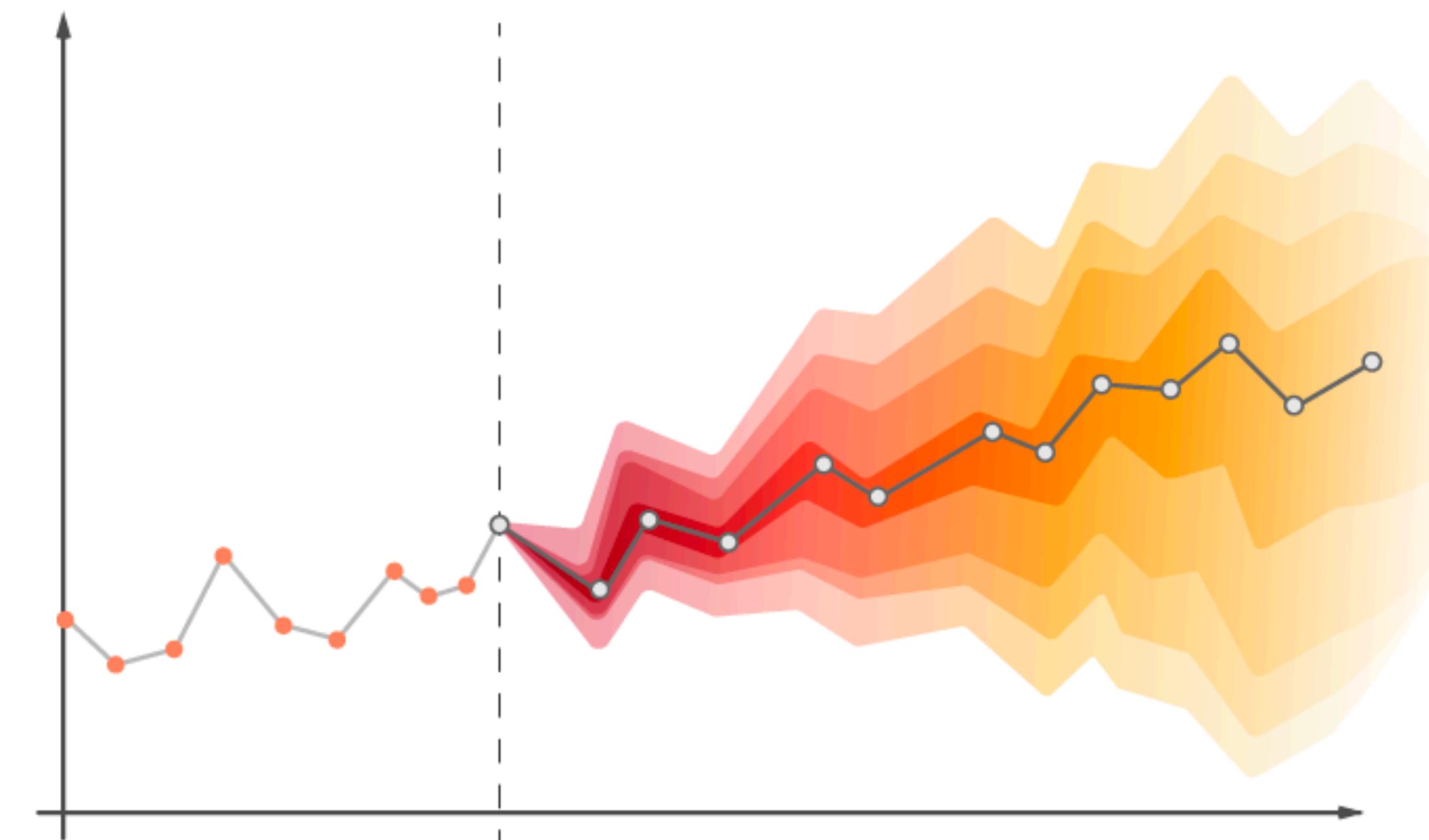
Exemples de problèmes

Prédiction de séries temporelles



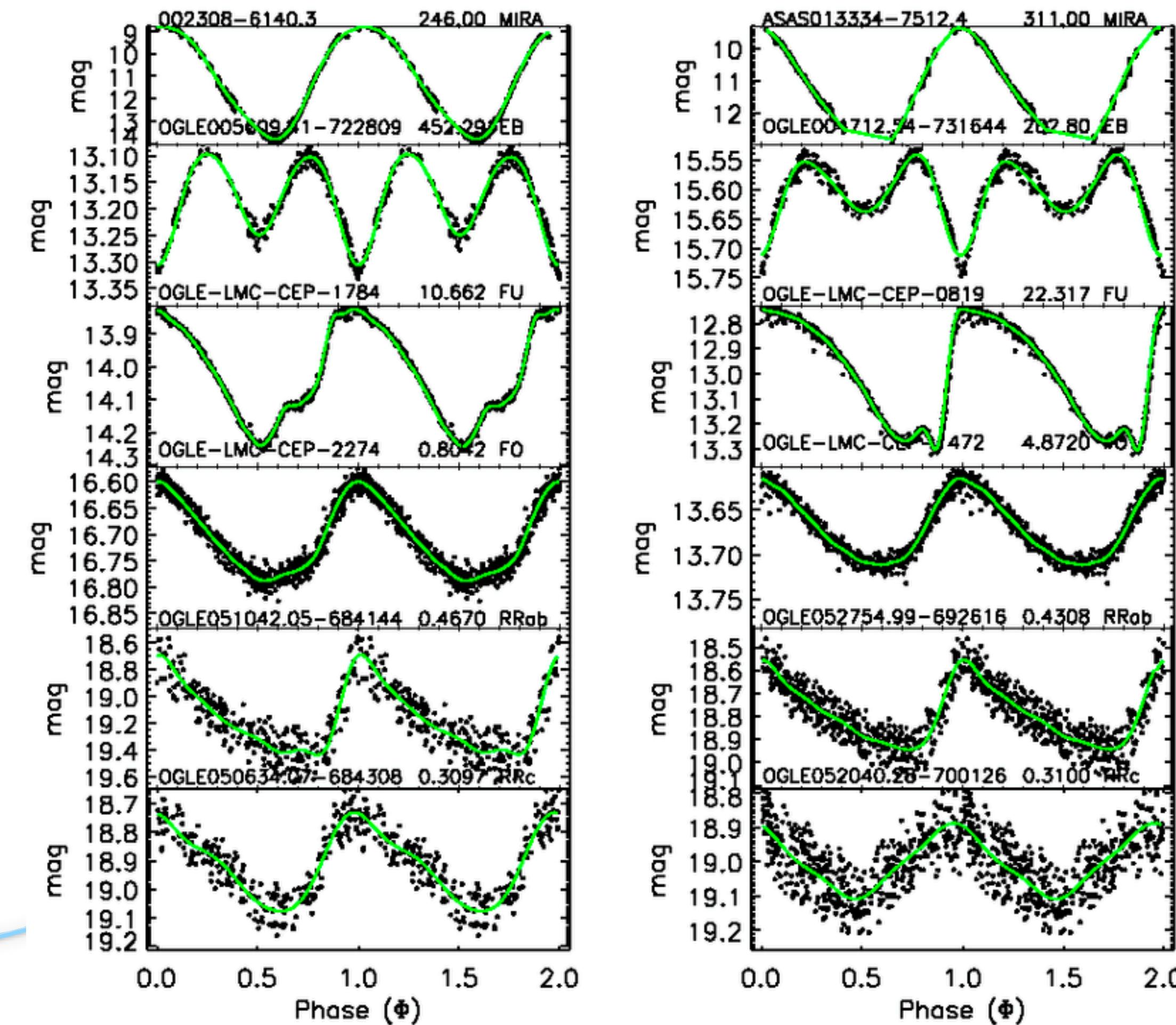
Exemples de problèmes

Prédiction de séries temporelles



Exemples de problèmes

Prédiction de séries temporelles



Exemples de problèmes

Prédiction de séries temporelles

Données de la bourse

Consommation d'énergie

Trajectoire

Dynamique d'un système

Courbes de lumière d'étoiles

Prédiction de données manquantes

Météo

...

Examples de problèmes

Supposition d'hypothèse

Number Game

On donne une liste
de nombres de 1 à 100

On donne les valeurs
 $4, 16, 36, 100$

Complétez la série

Vous avez 2 minutes

Exemples de problèmes

Supposition d'hypothèse

Hypothèse 1:
Nombre pair

2	4	6	8	10
12	14	16	18	20
22	24	26	28	30
32	34	36	38	40
42	44	46	48	50
52	54	56	58	60
62	64	66	68	70
72	74	76	78	80
82	84	86	88	90
92	94	96	98	100

Exemples de problèmes

Supposition d'hypothèse

Hypothèse 1:
Nombre pair

2	4	6	8	10
12	14	16	18	20
22	24	26	28	30
32	34	36	38	40
42	44	46	48	50
52	54	56	58	60
62	64	66	68	70
72	74	76	78	80
82	84	86	88	90
92	94	96	98	100

Hypothèse 2:
Multiples de 4

4	8
12	20
24	28
32	40
44	48
52	60
64	68
72	80
84	88
92	100

Exemples de problèmes

Supposition d'hypothèse

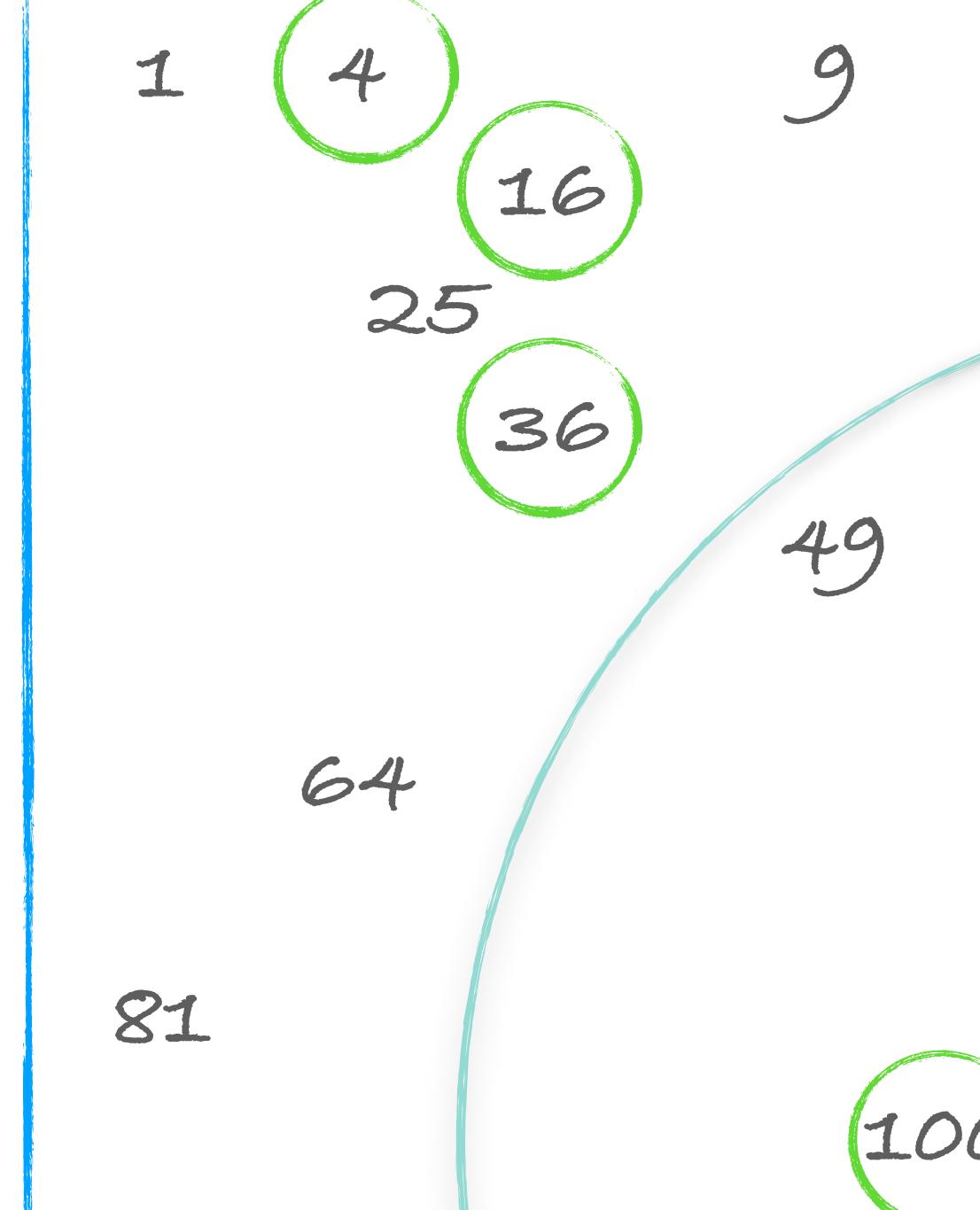
Hypothèse 1:
Nombre pair

2	4	6	8	10
12	14	16	18	20
22	24	26	28	30
32	34	36	38	40
42	44	46	48	50
52	54	56	58	60
62	64	66	68	70
72	74	76	78	80
82	84	86	88	90
92	94	96	98	100

Hypothèse 2:
Multiples de 4

12	4	8	20
24	16	28	40
32	36	44	56
44	52	64	68
52	64	76	80
64	72	84	96
72	84	88	100

Hypothèse 3:
Mise au carré



Exemples de problèmes

Supposition d'hypothèse

Les 3 hypothèses reproduisent les données

Comment choisir?

L'option qui a le moins de paramètres

Plus simple de générer H_3 à partir de H_1

Rasoir d'Occam

Exemples de problèmes

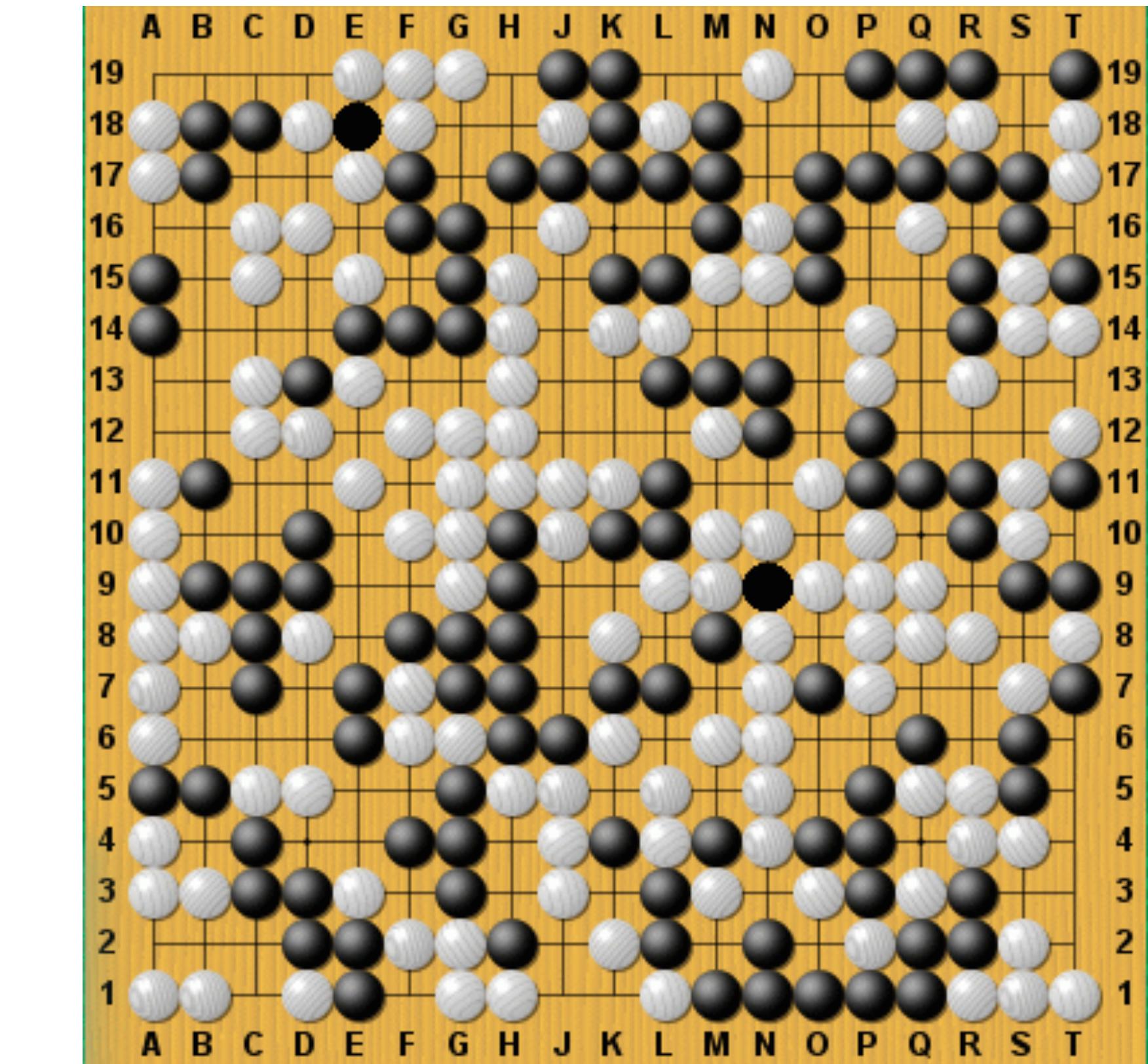
Jeux de plateaux

Deep Blue - 1996

10^{40}
Moves



alphaGo - 2016
alphaZero - 2018
MuZero - 2019



21

Exemples de problèmes

Jeux de plateaux

alphaStar - 2019



StarCraft II

OpenAI Five - 2019



Dota 2

<https://deepmind.google/discover/blog/alphastar-mastering-the-real-time-strategy-game-starcraft-ii/>

<https://openai.com/research/openai-five-defeats-dota-2-world-champions>

Apprentissage

Apprentissage

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E .

Tom M. Mitchell

Apprentissage

"L'apprentissage" est le processus par lequel un programme ou un système informatique acquiert des connaissances et améliore ses performances au fil du temps.

Machine Learning

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E .

[...] Quand l'agent est ordinateur, nous appelons ça apprentissage machine: un ordinateur observe des données, conçoit un modèle basé sur les données, et utilise ce modèle comme hypothèse sur l'environnement et une partie d'un algorithme qui peut résoudre des problèmes.

Tom M. Mitchell

Artificial Intelligence:
A Modern Approach,
4th edition, p.651

Apprentissage

Pourquoi vouloir qu'une machine "apprenne"?
Pourquoi ne pas la créer directement avec ce qu'elle doit savoir?

Limitation de la connaissance humaine.

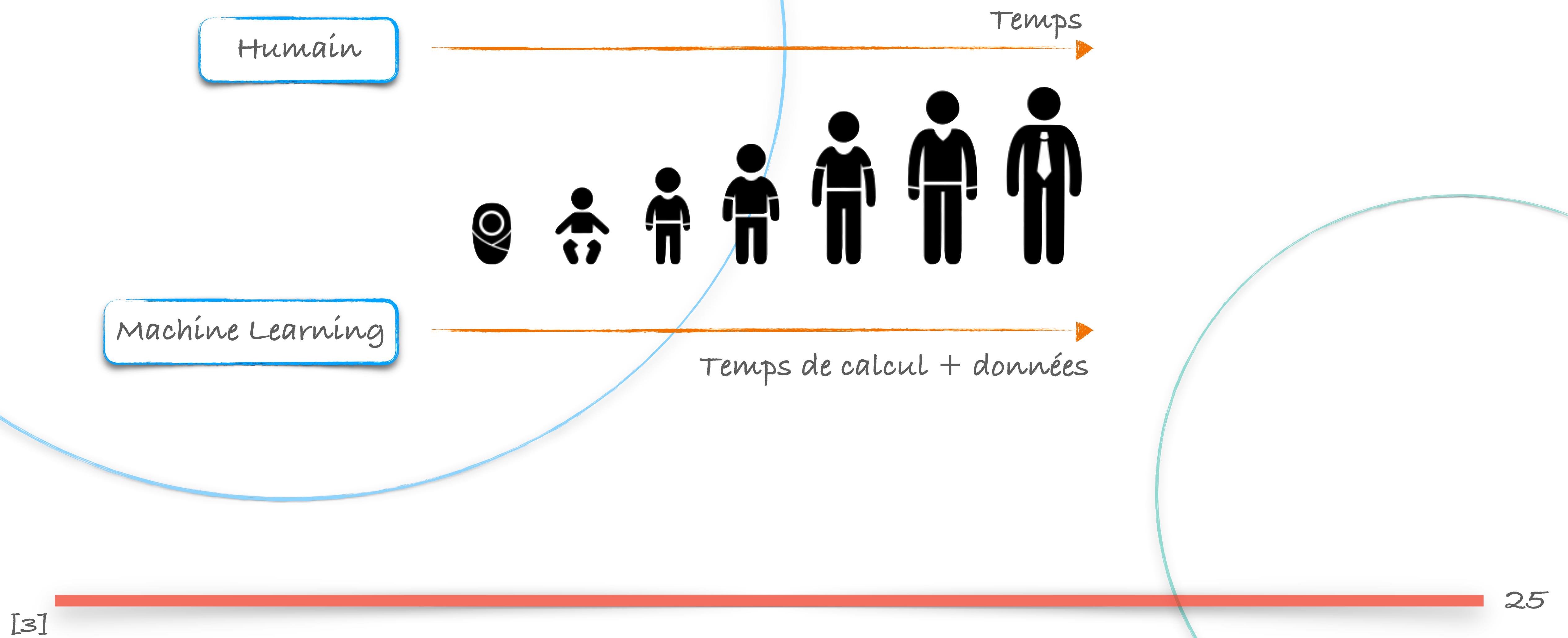
Un designer ne peut pas connaître/anticiper toutes les possibilités

Un designer peut ne pas connaître la solution à coder

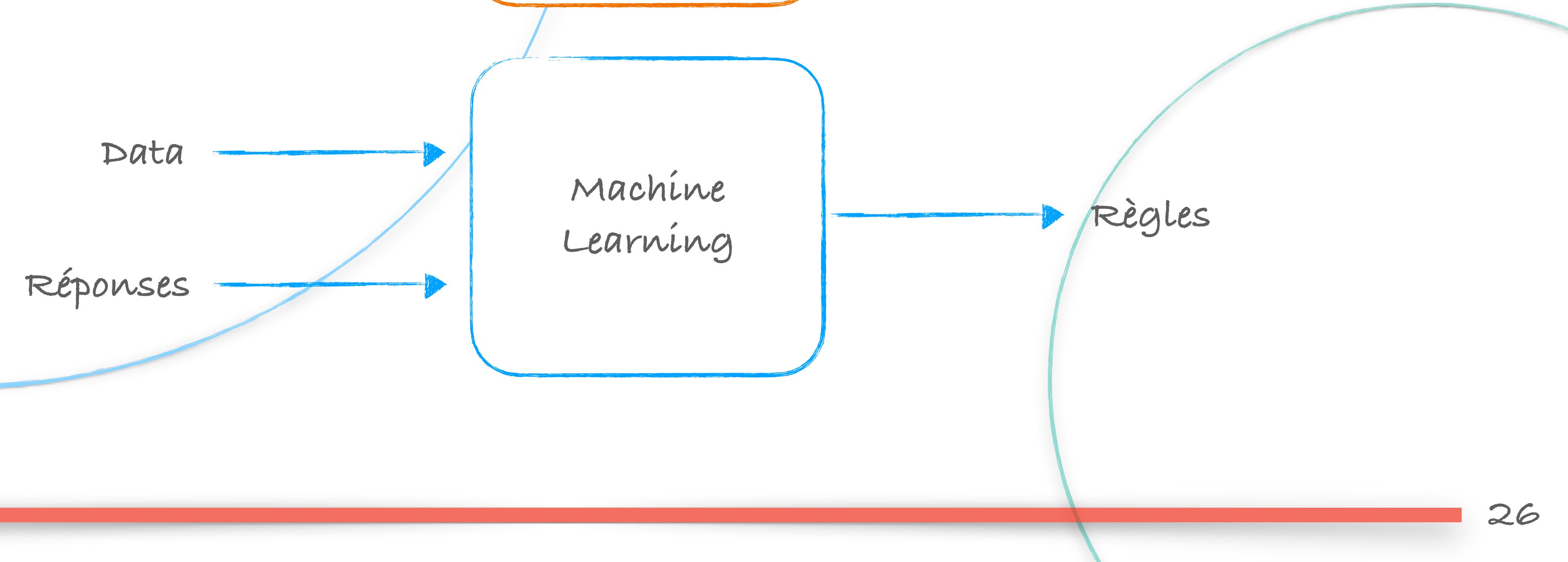
On peut mesurer/observer/récolter de l'information sur la tâche via l'environnement.

Données/dataset

Apprentissage



Apprentissage



Apprentissage

Apprentissage supervisé

Label

Apprentissage non-supervisé

Absence de Label

Set de données $\mathcal{D} = \{(x^1, y^1), \dots, (x^M, y^M)\}$

$(x^m, y^m) \in \mathcal{X} \times \mathcal{Y}$

$m = 1, \dots, M$

Distribution proba.

$p(x, y)$

Input $x \in \mathcal{X}$
Préd. $y \in \mathcal{Y}$

Set de données $\mathcal{D} = \{x^1, \dots, x^M\}$

$x^m \in \mathcal{X}$

Distribution proba.

$p(x)$

Famille de modèles $\{f\}$

Apprentissage

Apprentissage supervisé

Apprentissage non-supervisé

Minimisation du risque

$x \longrightarrow y$

$$L(f(x), y)$$

Régression

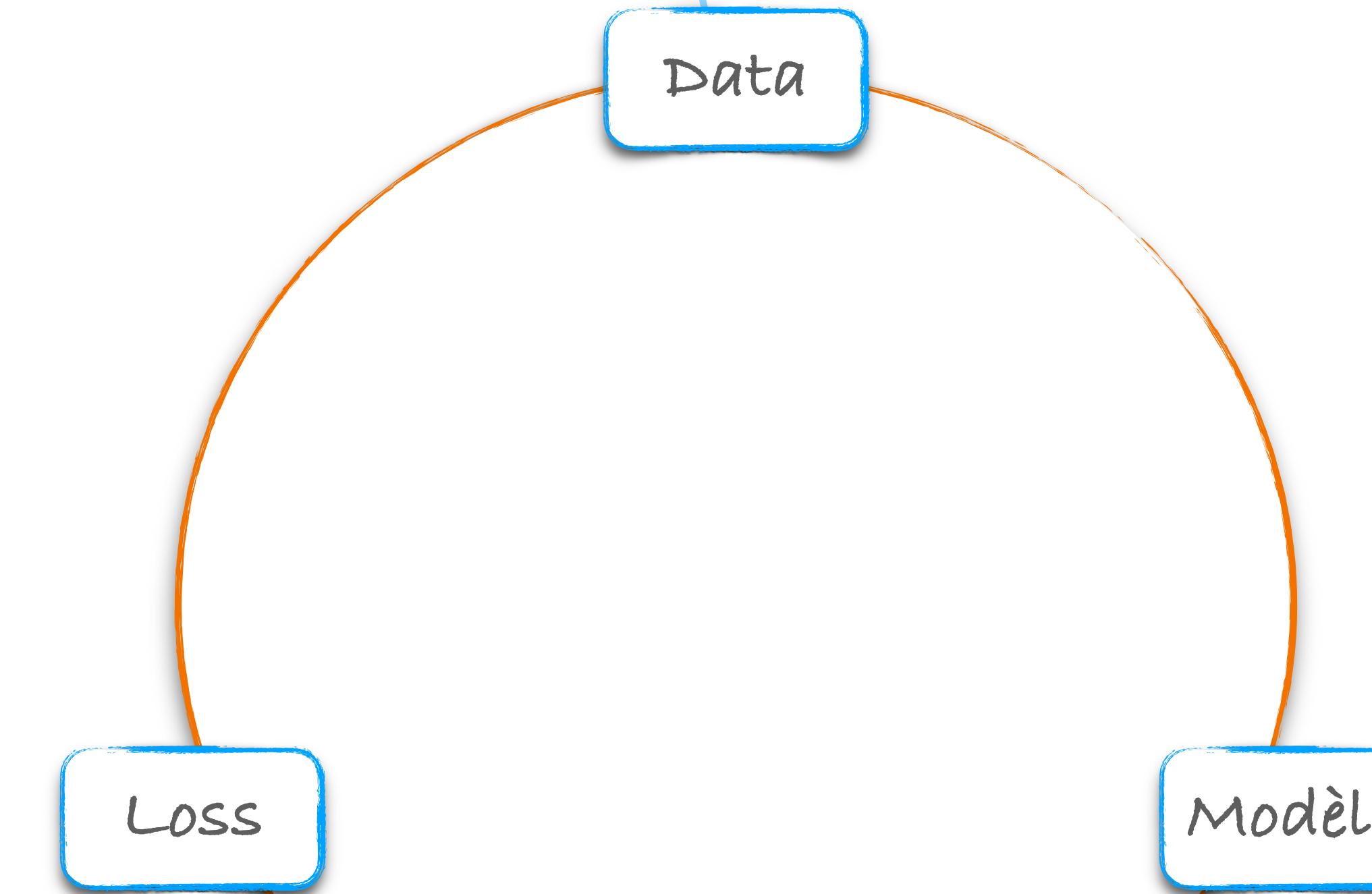
Classification

Calcul de distance

clustering

Les trois ingrédients d'un problème d'apprentissage

Les trois ingrédients d'un problème d'apprentissage



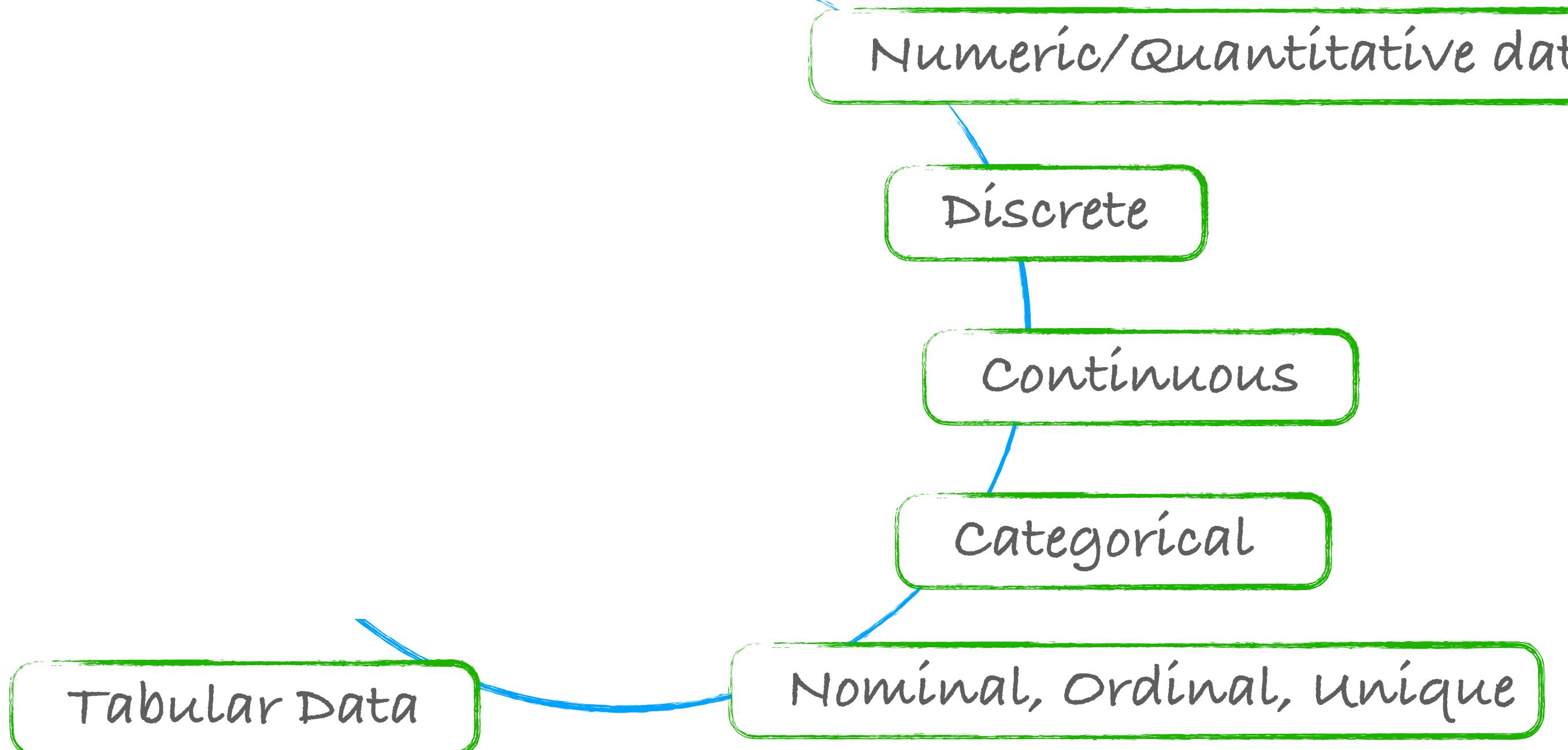
Les trois ingrédients d'un problème d'apprentissage

Data

Hypothèse

Tous les échantillons sont indépendants et distribués identiquement

Structured data



Texte

video

Son/Audio

Image

Time Series

unstructured data

Les trois ingrédients d'un problème d'apprentissage

Modèle

Déterministe

$$f: \mathcal{X} \rightarrow \mathcal{Y}$$

$$f(x) = y; \quad x \in \mathcal{X}; \quad y \in \mathcal{Y}$$

Créer des fonctions avec paramètres

$$\mathbf{w} = (w_1, w_2, \dots)^T$$

weights

Probabiliste

distribution de probabilités

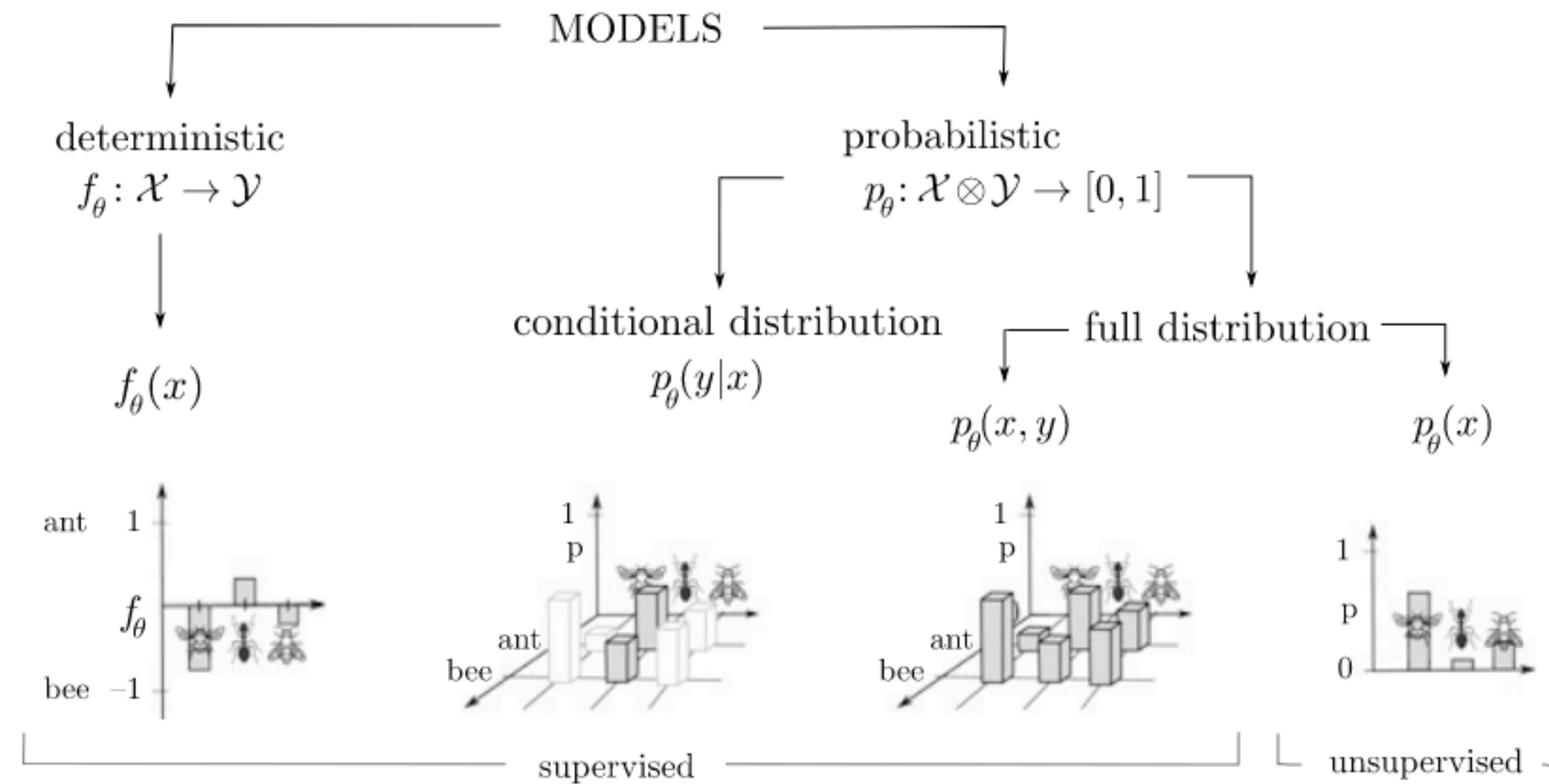
supervisé $p_\theta(x, y)$

conditionnel $p_\theta(y | x)$

Non-supervisé $p_\theta(x)$

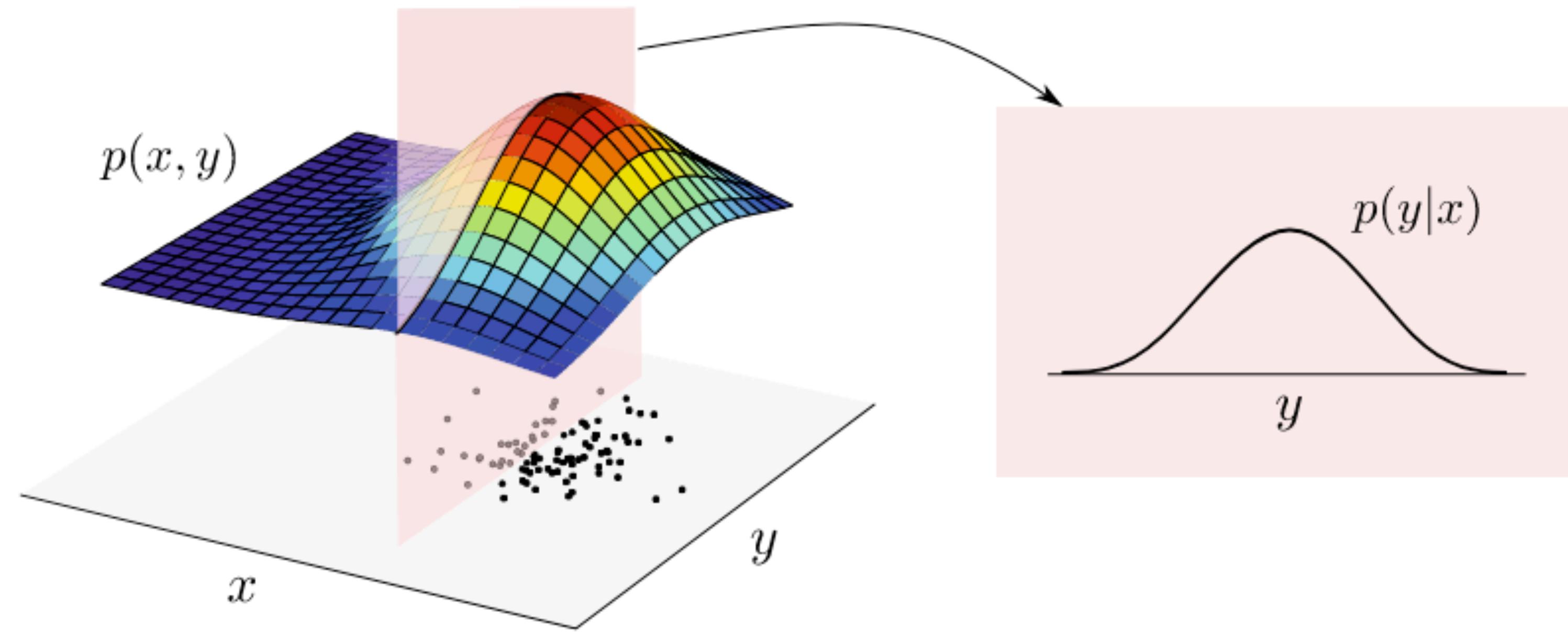
Les trois ingrédients d'un problème d'apprentissage

Modèle



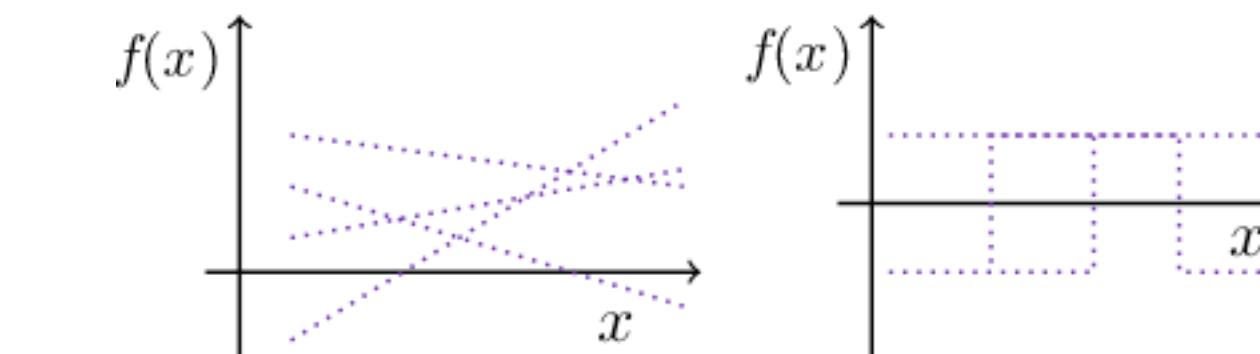
Les trois ingrédients d'un problème d'apprentissage

Modèle



$$x \rightarrow f_{\theta}(x) \rightarrow y$$

$$x \sim p_{\theta}(x)$$



Les trois ingrédients d'un problème d'apprentissage

LOSS

Comment évaluer un modèle?

Prédictions

\hat{Y}

vraies valeurs

Y

Distance
Erreur
"LOSS"

Risque
Généralisation Erreur

Les trois ingrédients d'un problème d'apprentissage

LOSS

Supervisé

Mean-squared-error

$$L(f_\theta(x), y) = (f_\theta(x) - y)^2$$

l_1

$$L(f_\theta(x), y) = |f_\theta(x) - y|$$

Hinge loss

$$L(f_\theta(x), y) = \max(0, 1 - yf(x))$$

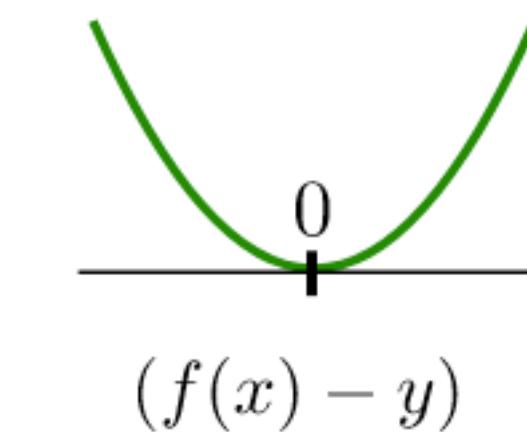
Logistic loss

$$L(f_\theta(x), y) = \log(1 + e^{-yf(x)})$$

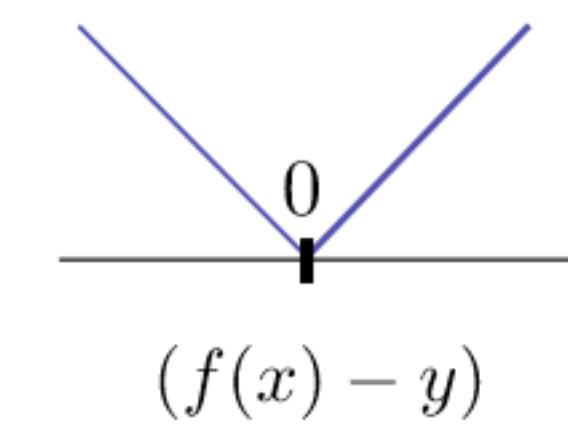
Cross entropy loss

$$L(f_\theta(x), y) = - \sum_{d=1}^D y_d \log p_d$$

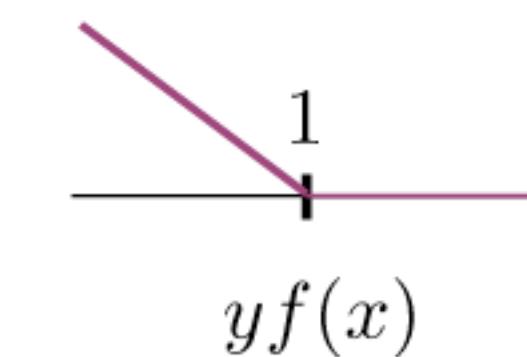
square loss



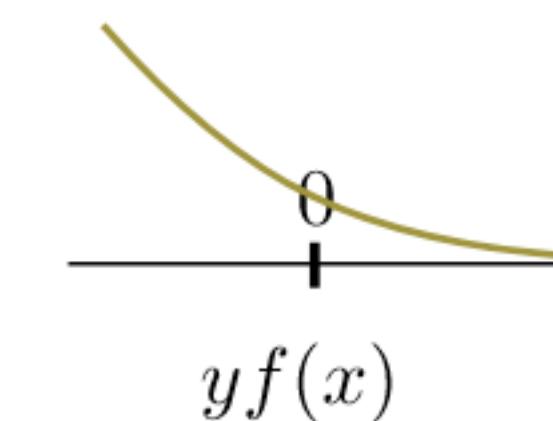
ℓ_1 loss



hinge loss



logistic loss



Les trois ingrédients d'un problème d'apprentissage

LOSS

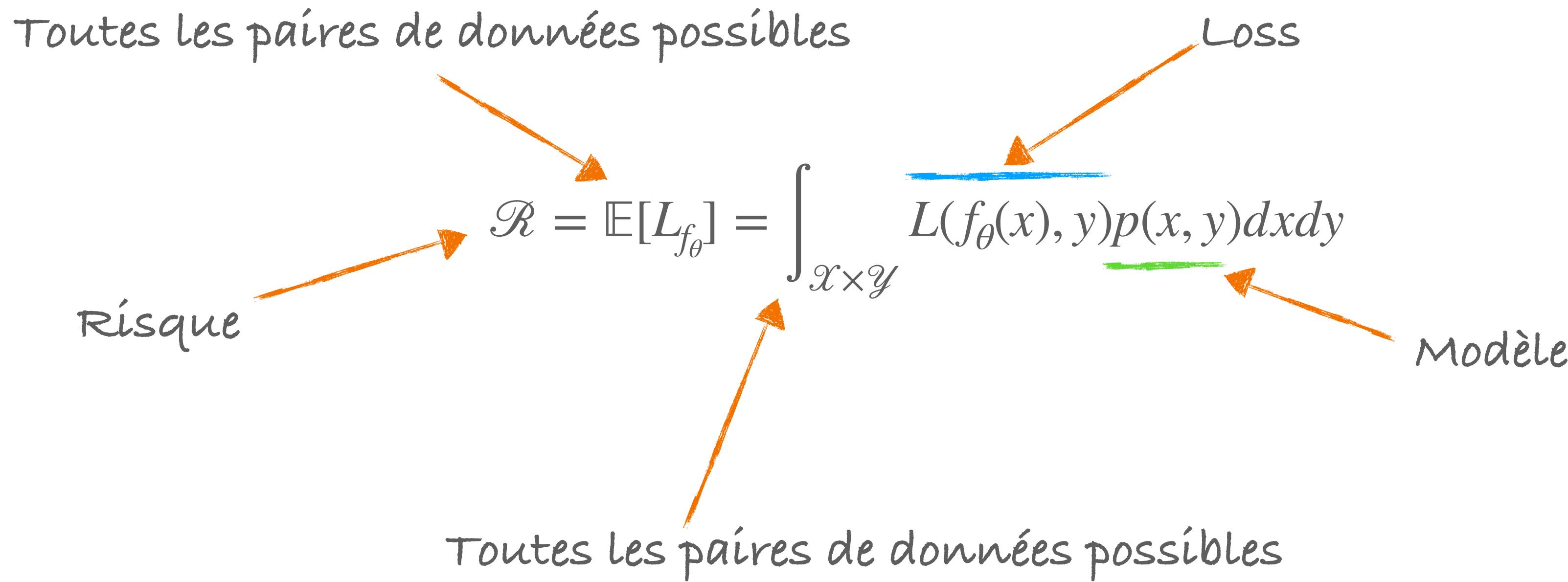
Non-Supervisé

$$L(x) = p_\theta(x)$$

Mesure la distance qu'une donnée observée appartienne à un modèle P_θ

Minimisation du risque en ML supervisé

Minimisation du risque ML supervisé



Le risque évalue la performance du modèle sur l'intégralité du domaine de données

Minimisation du risque ML supervisé

Toutes les paires de données possibles

Risque

Toutes les paires de données possibles

\mathcal{R}

$$\mathcal{R} = \mathbb{E}[L_{f_\theta}] = \int_{\mathcal{X} \times \mathcal{Y}} L(f_\theta(x), y) p(x, y) dx dy$$

Loss

Modèle

Impossible à calculer
dans un cas réel

Estimation

Risque empirique

Le risque évalue la performance du modèle sur l'intégralité du domaine de données

Minimisation du risque ML supervisé

$$\mathcal{R} = \mathbb{E}[L_{f_\theta}] = \int_{\mathcal{X} \times \mathcal{Y}} L(f_\theta(x), y) p(x, y) dx dy$$

impossible à calculer
dans un cas réel

Estimation

$$\hat{\mathcal{R}}_{f_\theta} = \hat{\mathbb{E}}[L_{f_\theta}] = \frac{1}{M} \sum_{m=1}^M L(f(x^m), y^m)$$

Comment minimiser?

Choix d'un modèle dans
une famille

Principe
d'optimisation

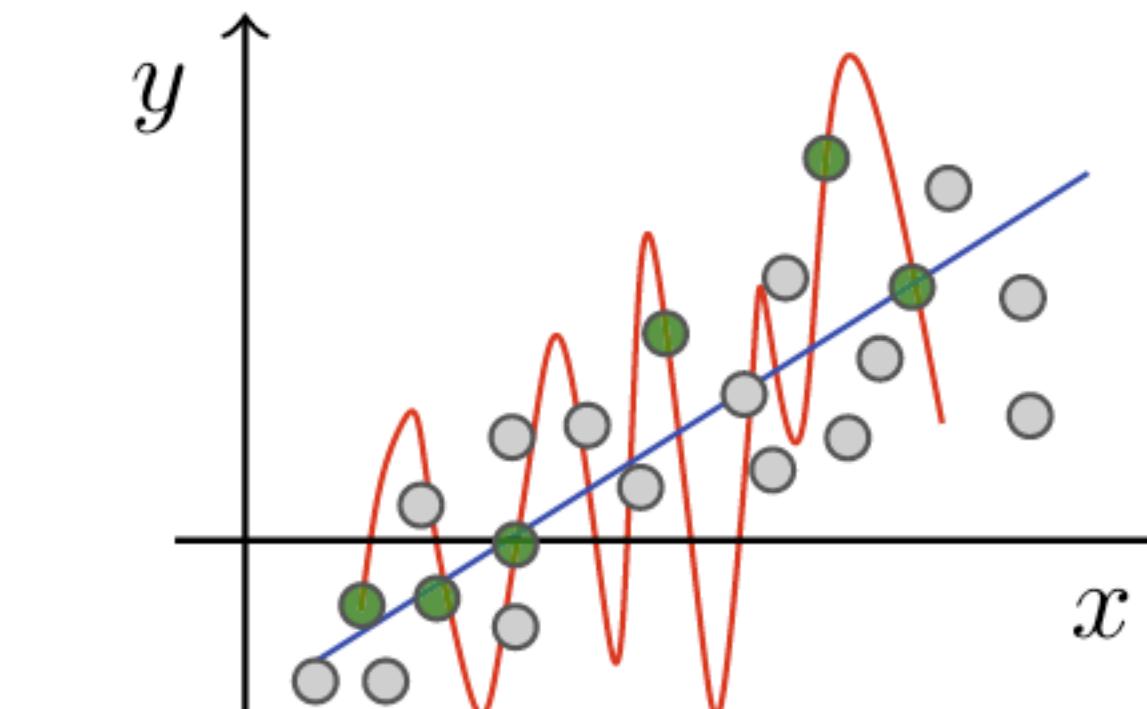
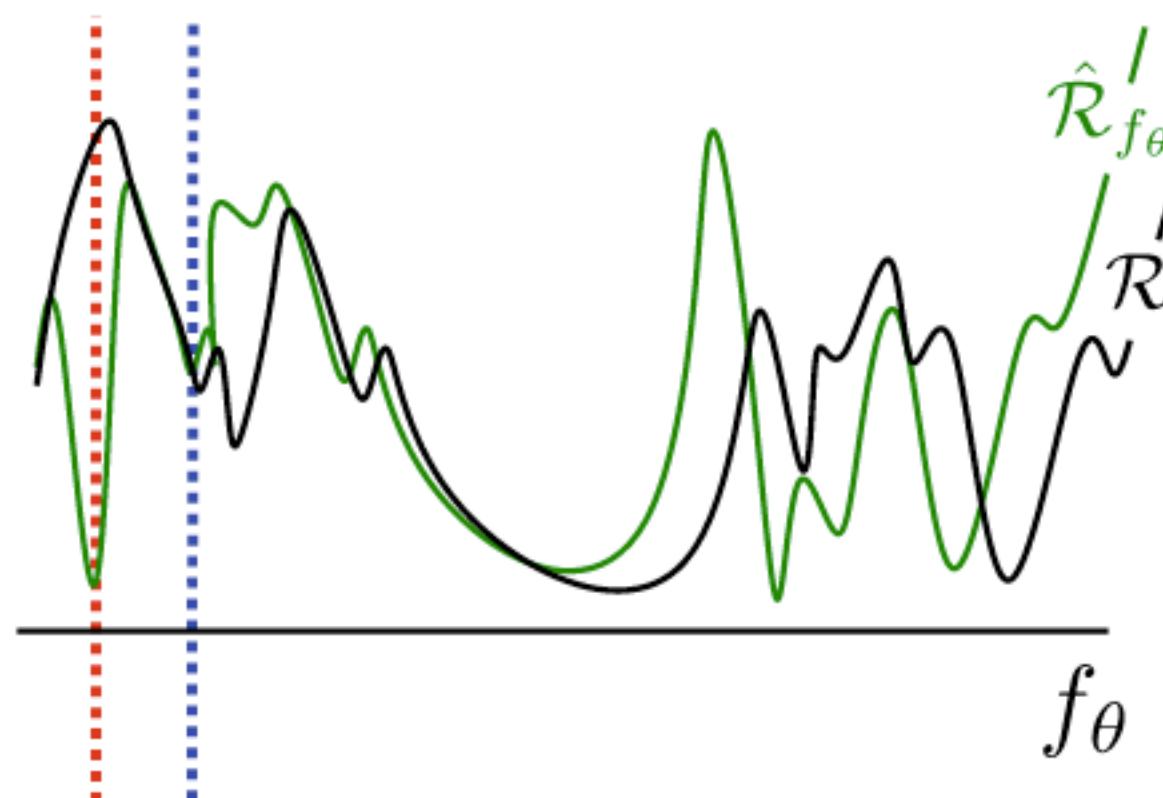
$$\theta^* = \min_{\theta} \hat{\mathcal{R}}_{f_\theta}$$

Minimisation du risque ML supervisé

overfitting

Lorsqu'un modèles performe bien sur les données d'entraînement mais performe pauvrement sur les données de test

Le modèle a appris les données et non les règles de généralisation



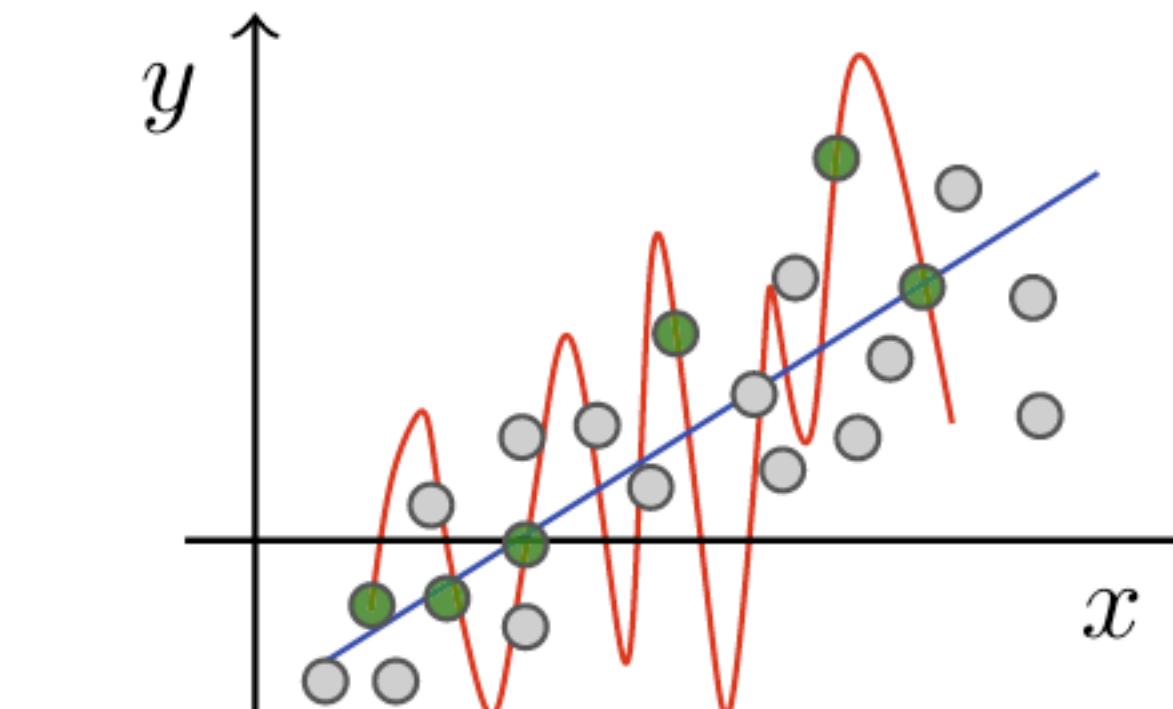
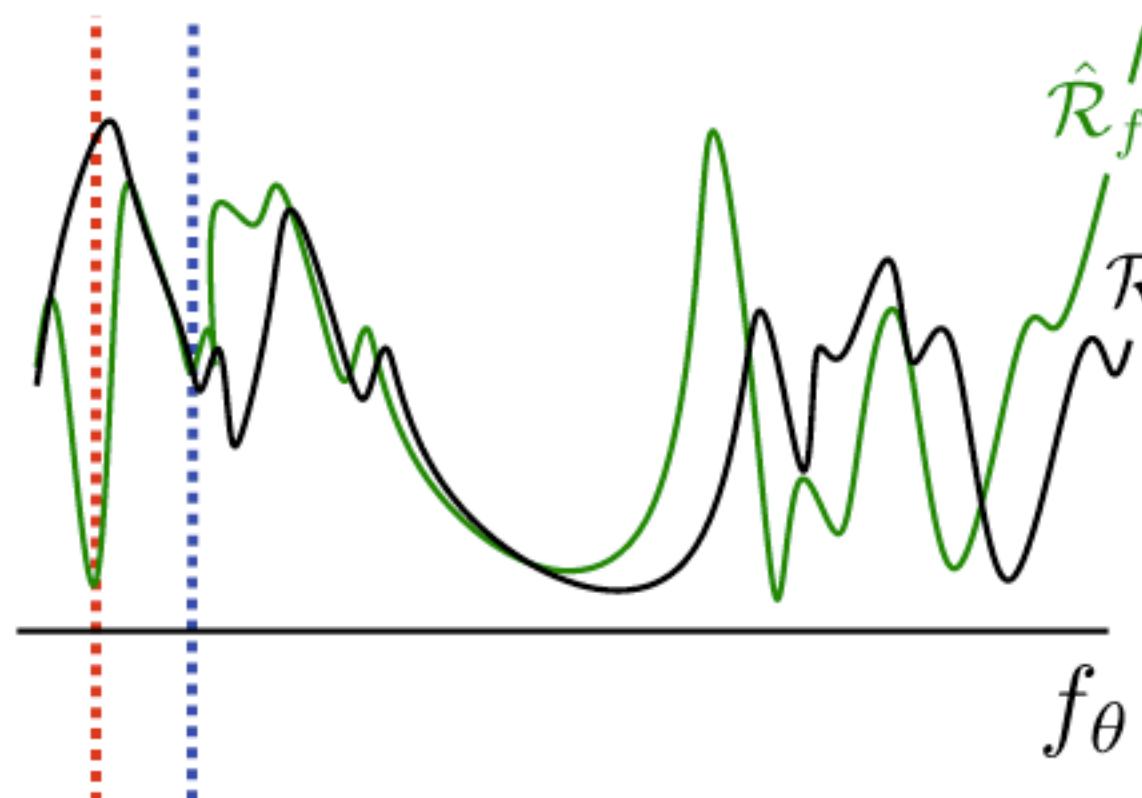
Minimisation du risque ML supervisé

overfitting

Lorsqu'un modèle forme bien sur les données d'entraînement mais forme pauvrement sur les données de test

plus gros challenge des modèles supervisés

Le modèle a appris les données et non les règles de généralisation



Minimisation du risque ML supervisé

Overfitting

Comment éviter?

Régularisation

$$C(\theta) = \hat{\mathcal{R}}_{f_\theta} + g(f_\theta)$$

Fonction de coût

Deux façons

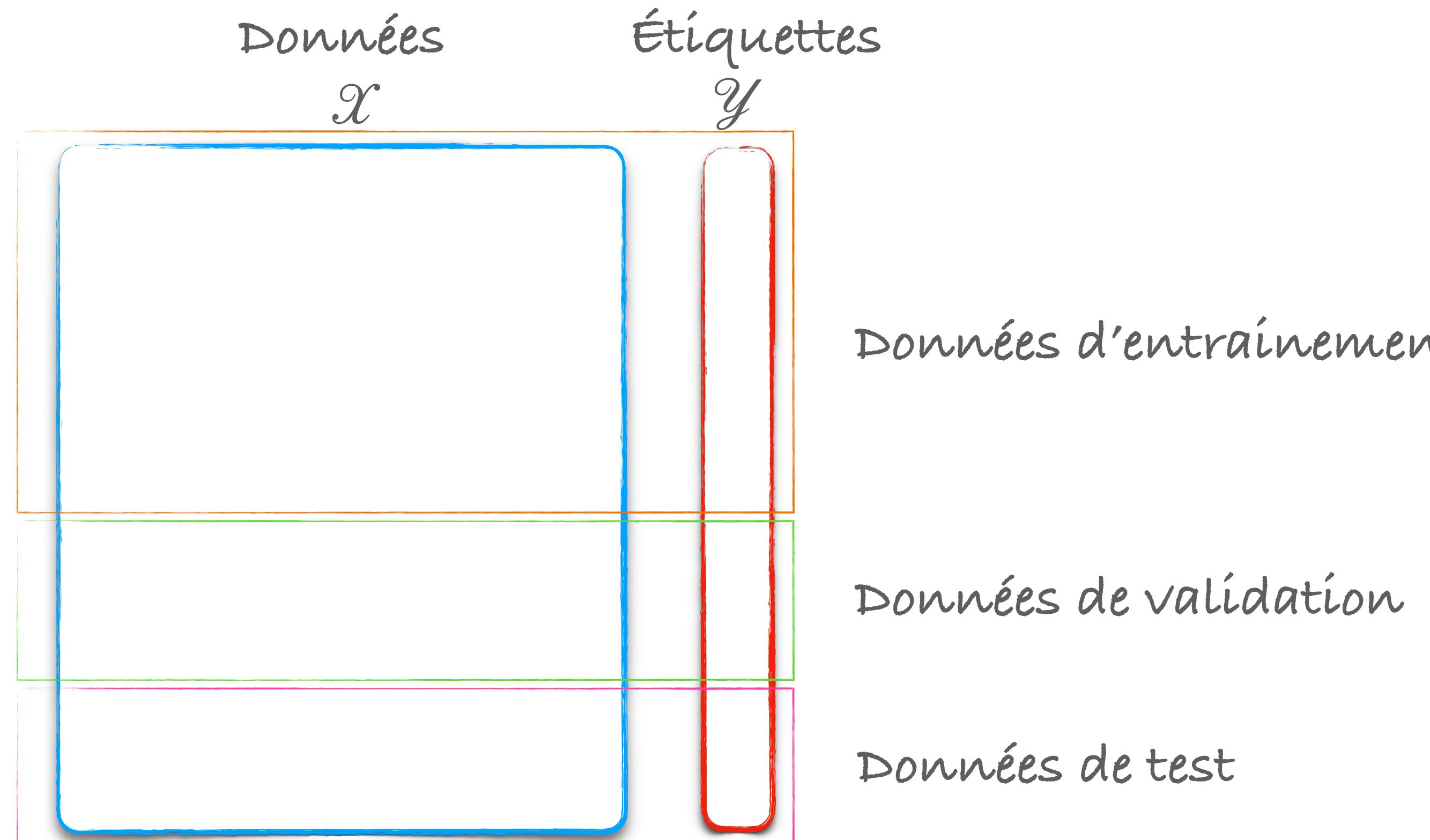
$$g_{l_1}(\theta) = \sum_i |\theta_i| \quad \text{Terme de régularisation } l_1$$

$$g_{l_2}(\theta) = \sum_i \theta_i^2 \quad \text{Terme de régularisation } l_2$$

Minimisation du risque ML supervisé

Overfitting

Mesure



Évaluation des performances

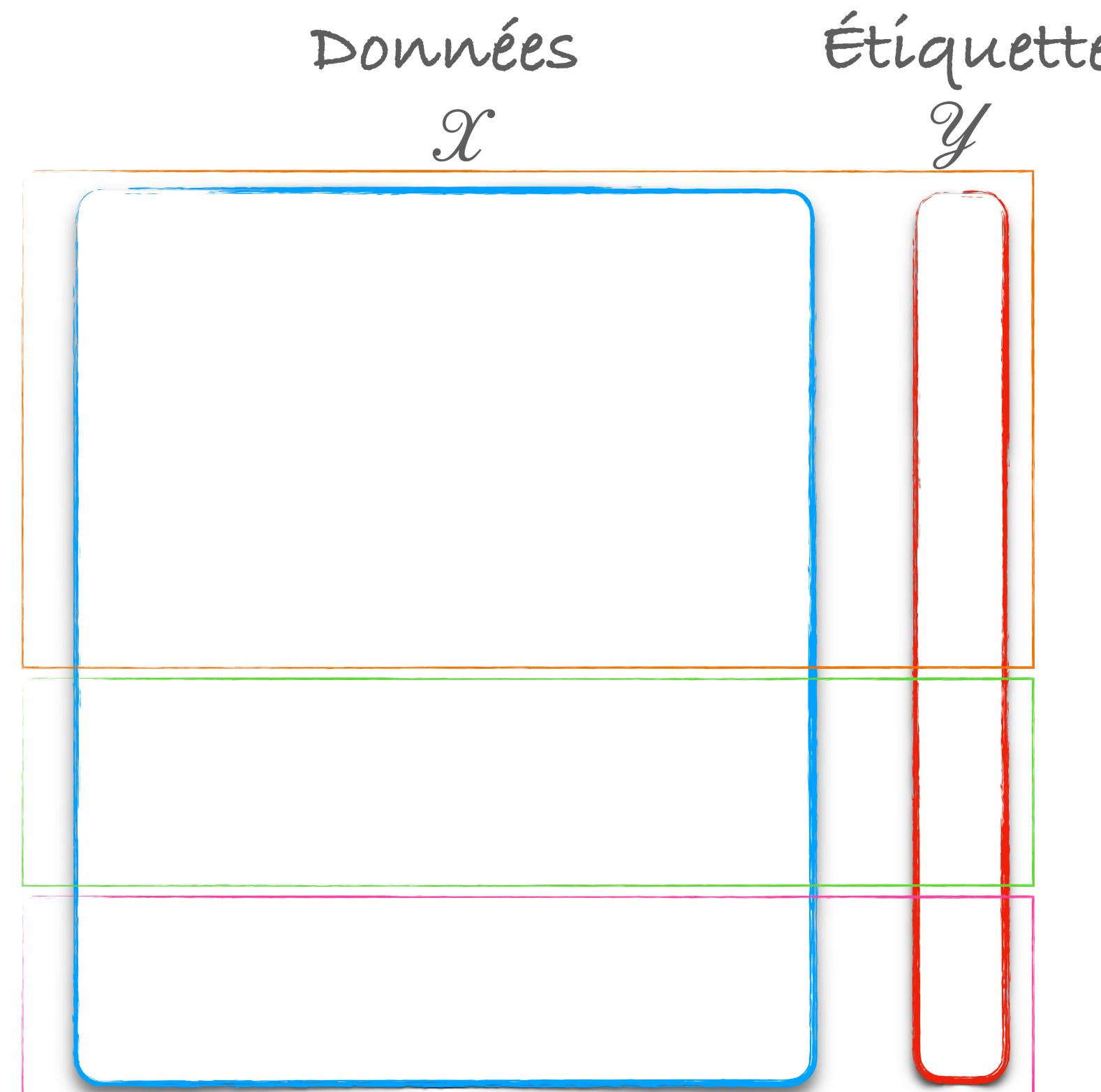
Évaluation pour changer
les hyperparamètres

Évaluation de la généralisation
ces données ne sont utilisées que
quand le modèle est complètement
entraîné

Minimisation du risque ML supervisé

Overfitting

Mesure



Modèle généralisé?

Juste données d'entraînement

Limites de généralisation

Erreur de généralisation

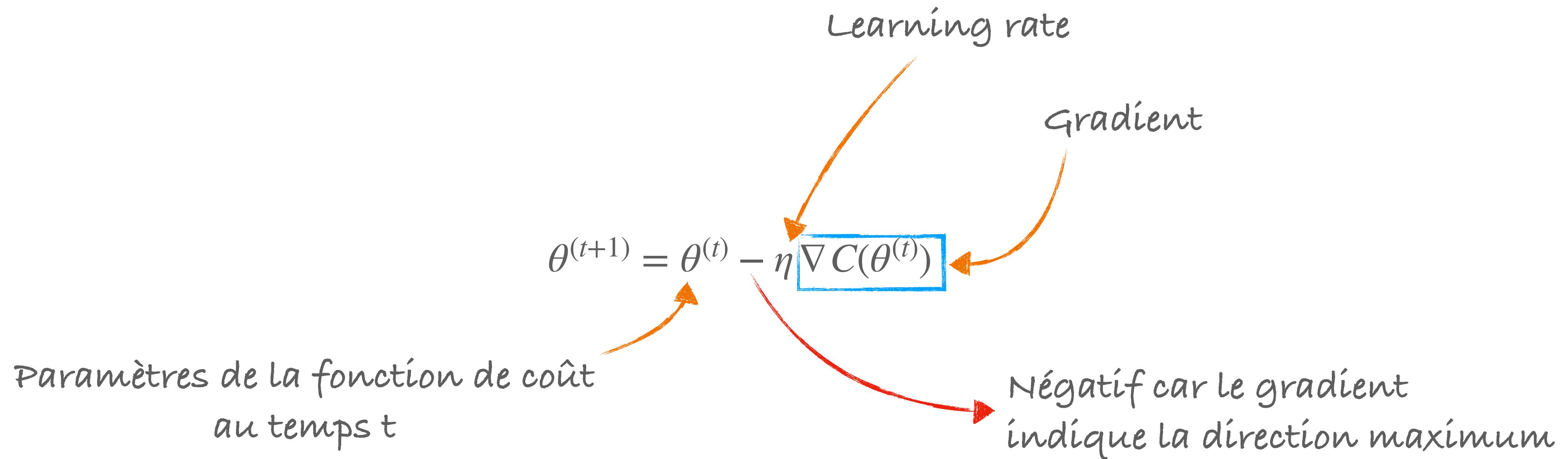
+

Expressivité

Chance qu'un modèle overfit

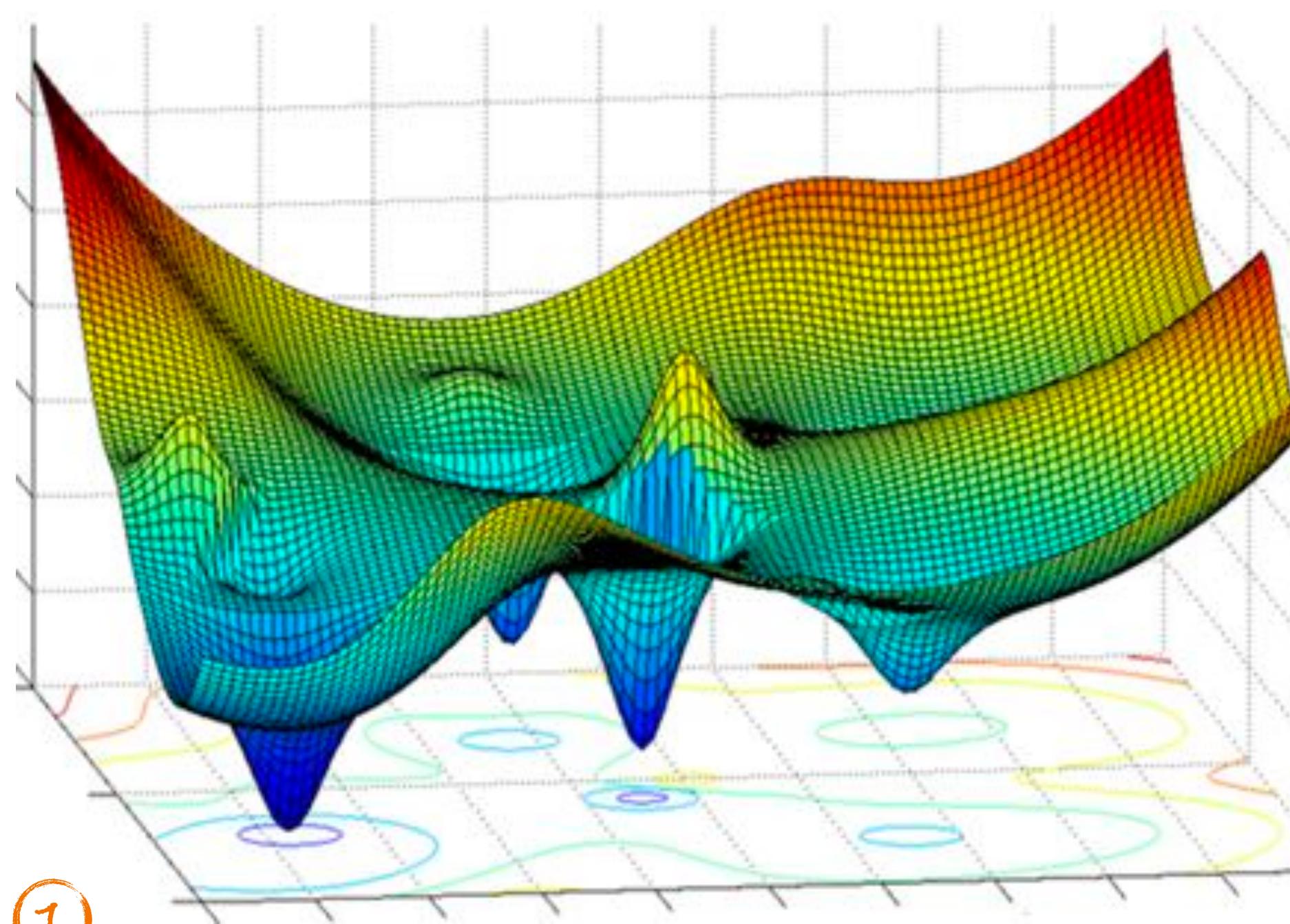
Minimisation du risque ML supervisé

Optimisation

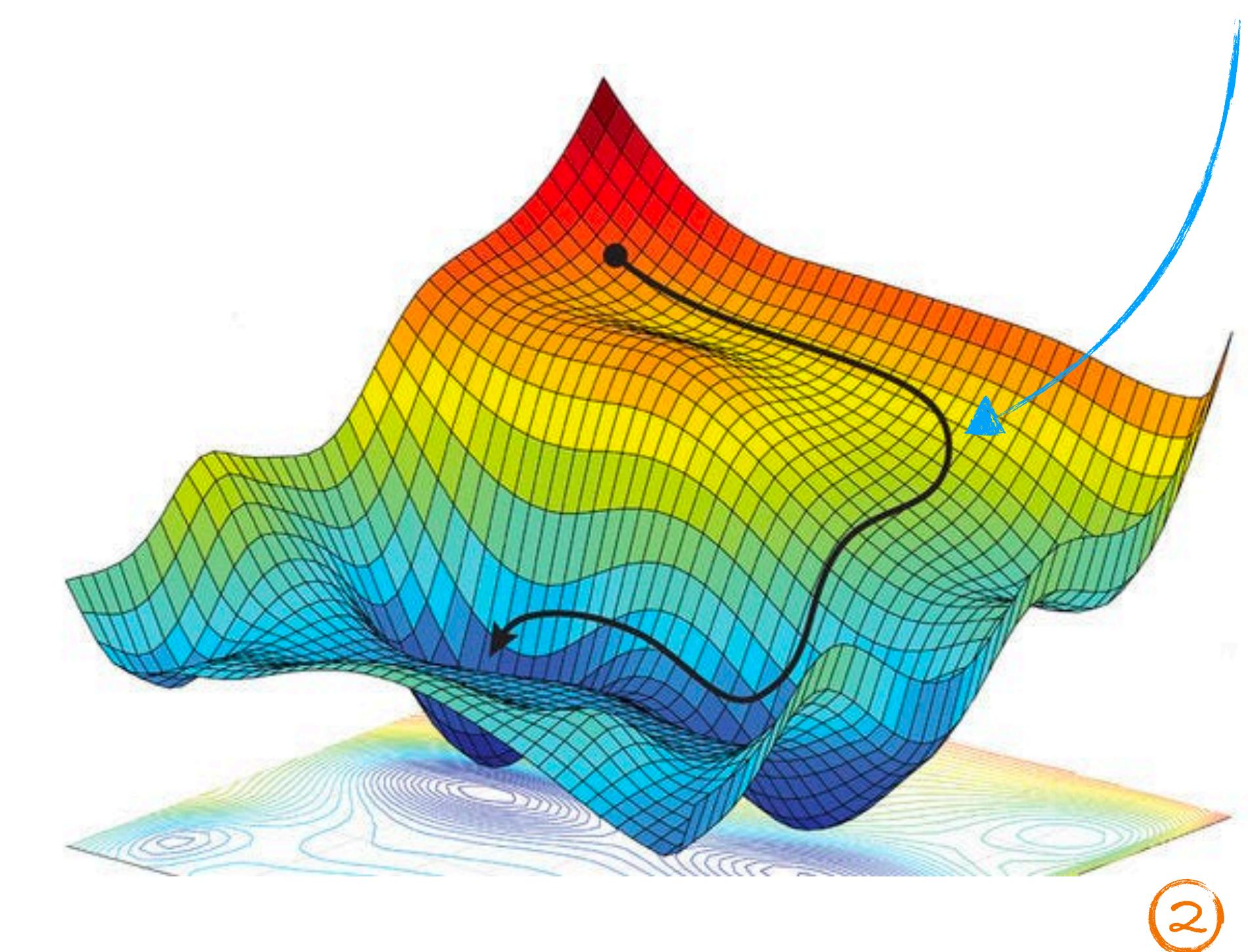


Minimisation du risque ML supervisé

Gradient



"Landscape"



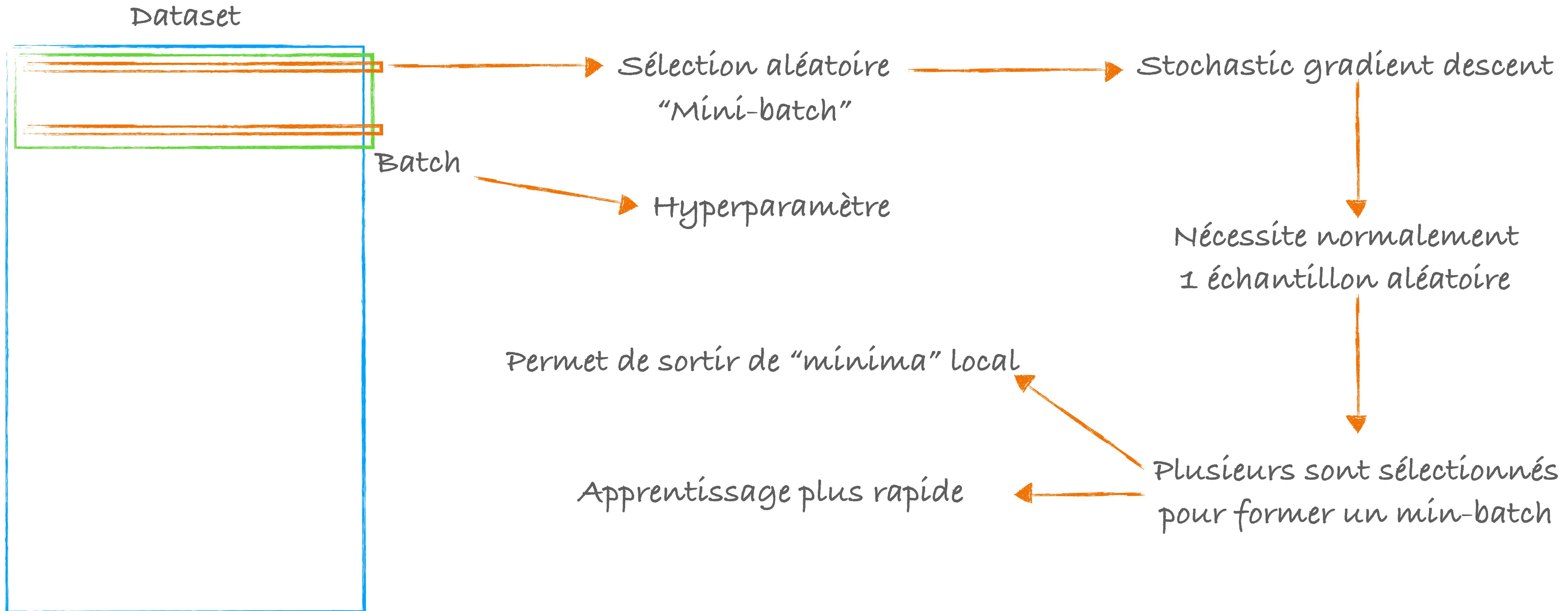
Descente de gradient

① <https://www.fromthegenesis.com/gradient-descent-part-2/>

② <https://towardsdatascience.com/coding-deep-learning-for-beginners-linear-regression-gradient-descent-fcd5e0fc077d>

Minimisation du risque ML supervisé

Gradient



Entrainement pour les méthodes non-supervisées

Entraînement pour les méthodes non-supervisées

Apprentissage non-supervisé

L'objectif est d'analyser et de grouper des données qui n'ont pas d'étiquette

Trois grands types

Clustering

Calcul distances dans le cluster et entre cluster
Basé sur les similarités et différences

Règles d'association

Règles prédéterminées ex: relations entre des produits

Réduction dimensionnelle

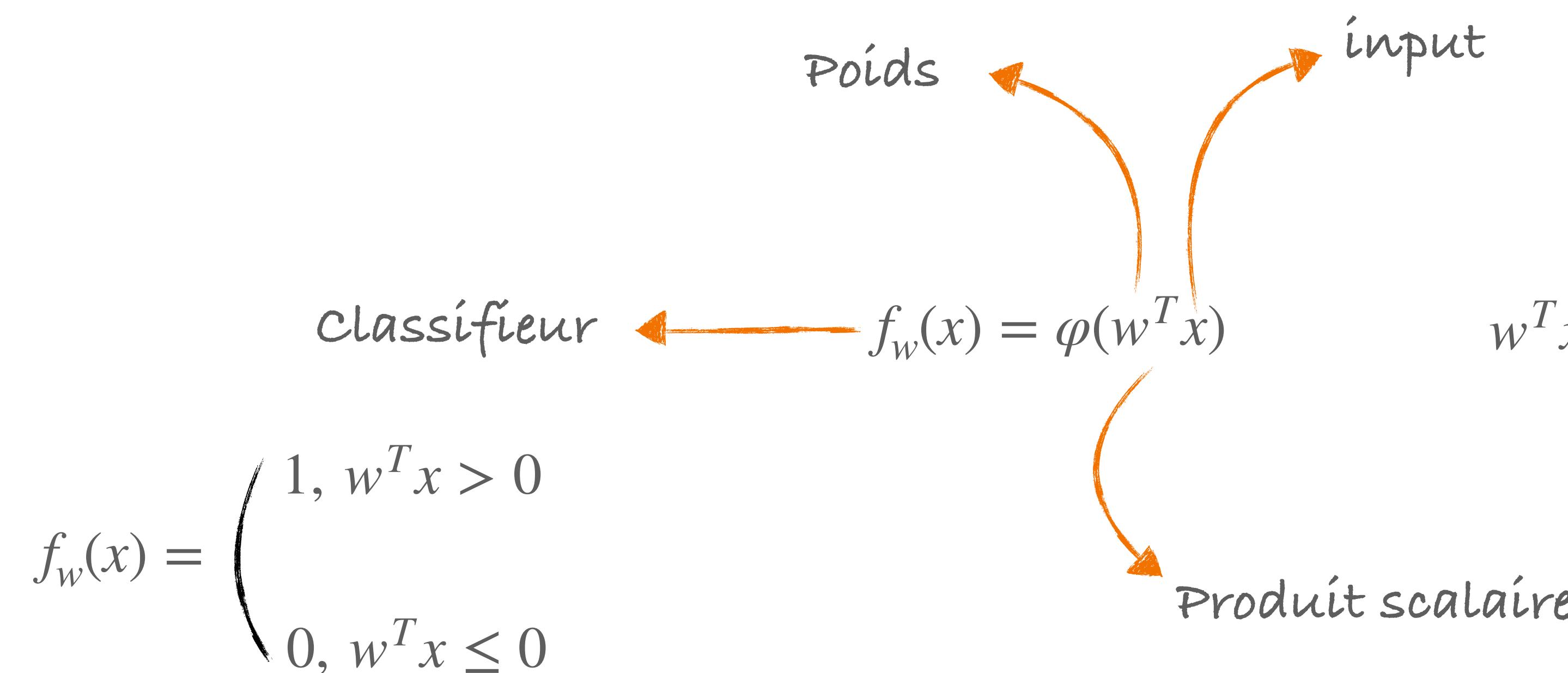
PCA
SVD
Autoencoder

Méthodes/modèles en ML

Non-exhaustif

Méthodes/modèles en ML

Modèles linéaires

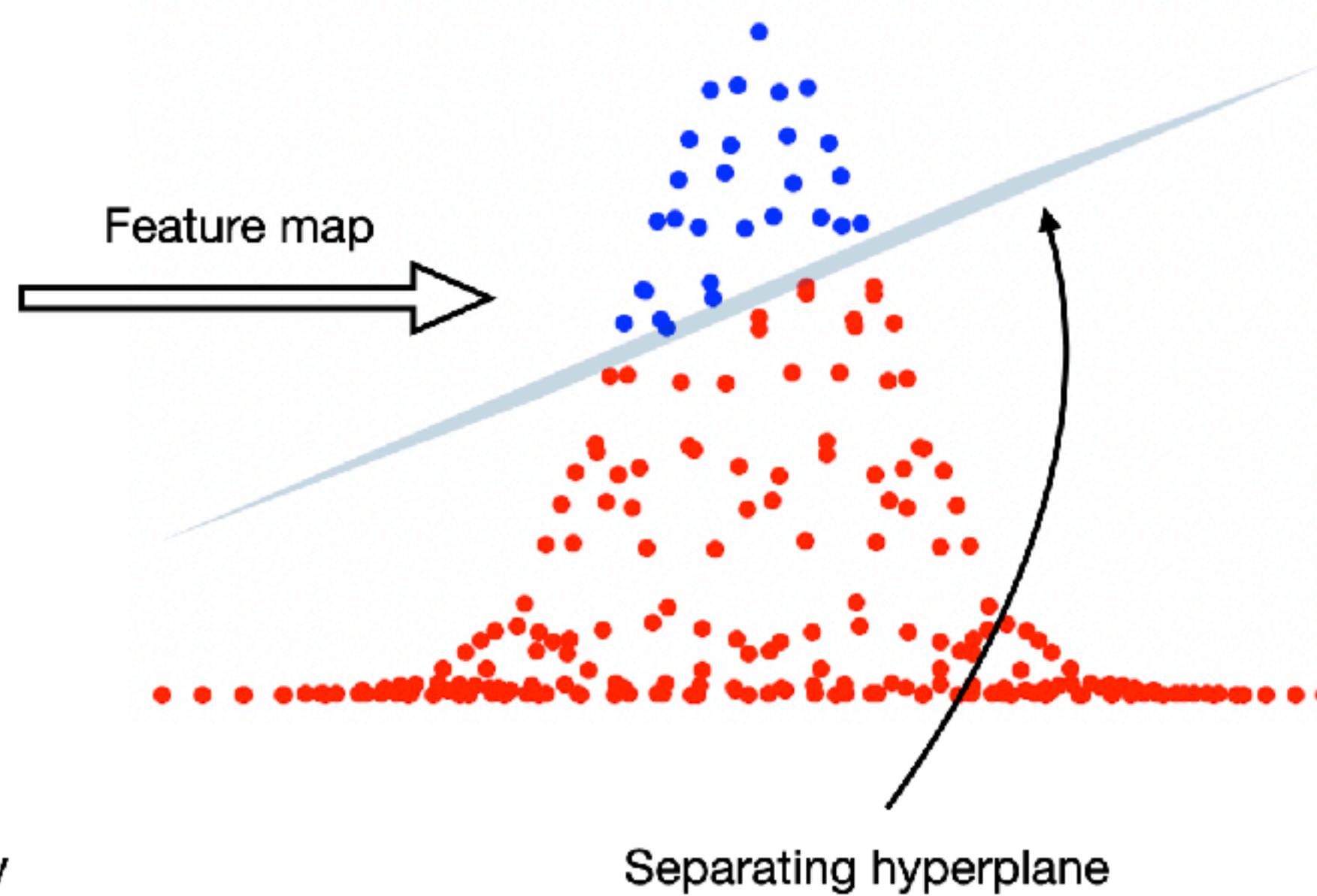
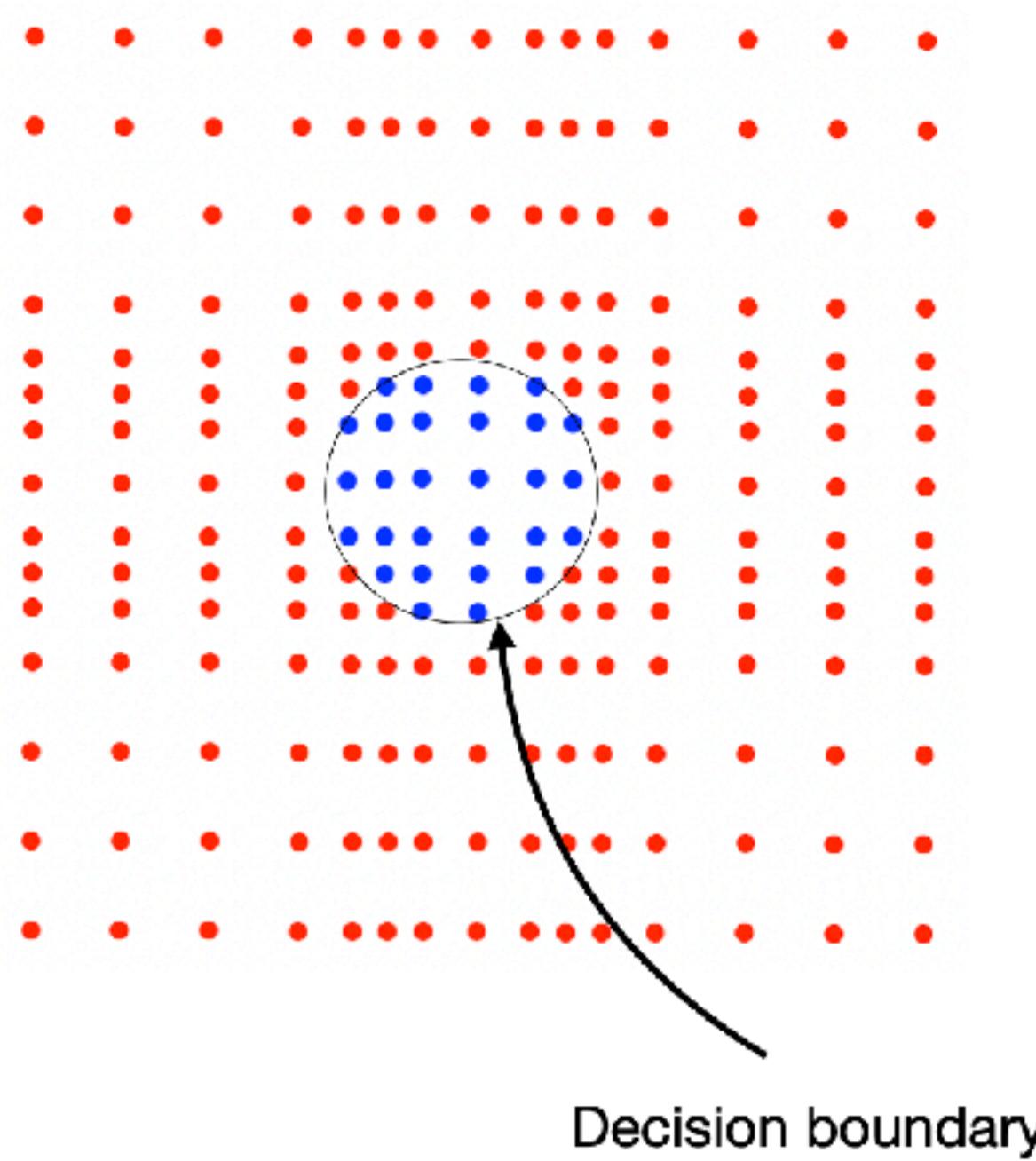


$$w^T x = w_0 x_0 + w_1 x_1 + \dots + w_N x_N + b$$

Formule générale pour
une dimension N

Méthodes/modèles en ML

Kernels



Kernel entre les points

Résoudre des problèmes basés sur la mesure de similarité des données

Approche linéaire pour résoudre des problèmes non-linéaires

Application d'une "feature map" pour projeter dans un espace à m dimension une donnée de n dimension

$$m \gg n$$

Le produit scalaire donne la similarité entre deux points

$$X_1, X_2 \in X \text{ and a map } \phi : X \rightarrow \mathbb{R}^m$$

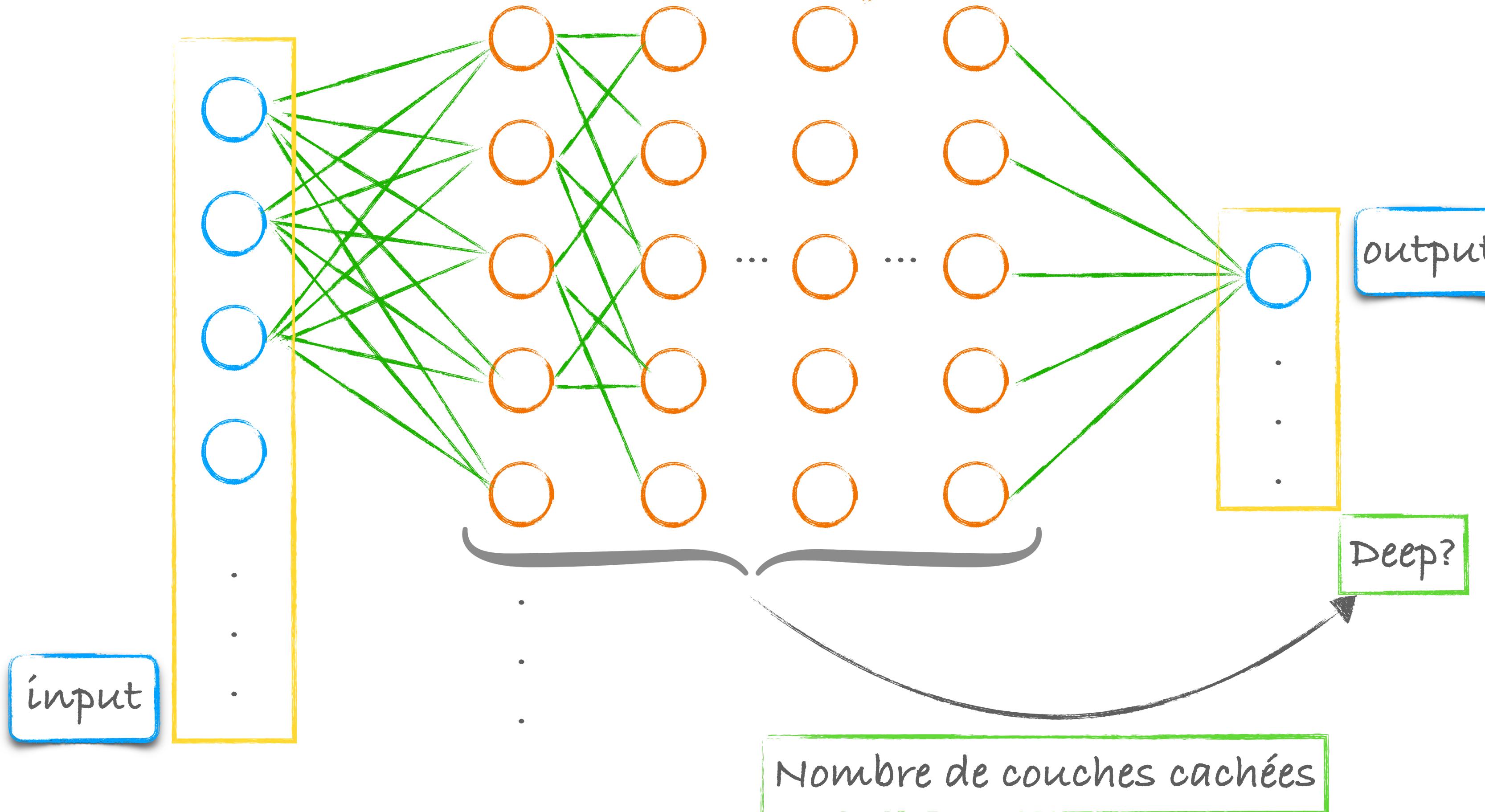
$$k(X_1, X_2) = \langle \phi(X_1), \phi(X_2) \rangle$$

Produit scalaire

Méthodes/modèles en ML

Réseau de neurones

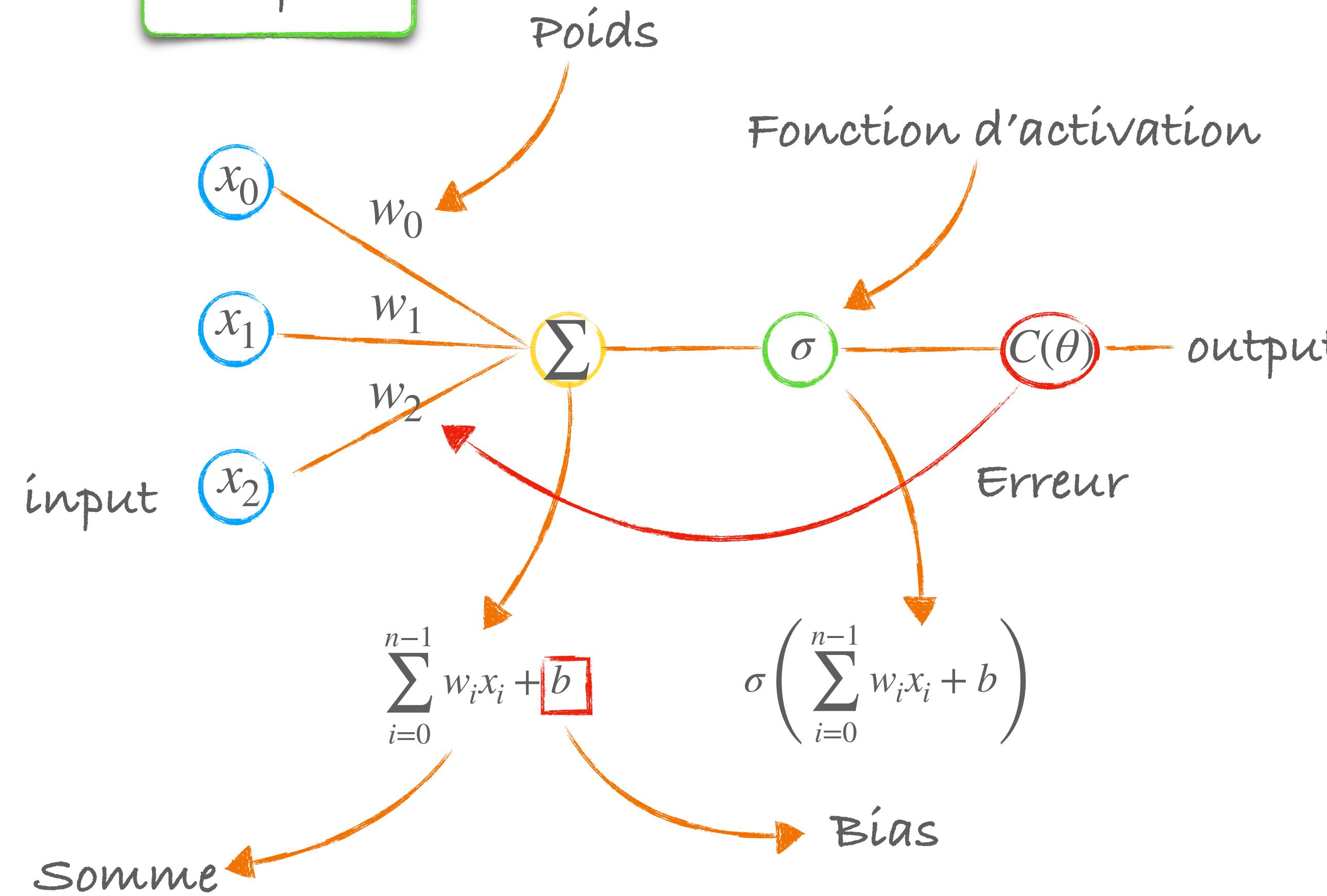
Neurone/perceptron



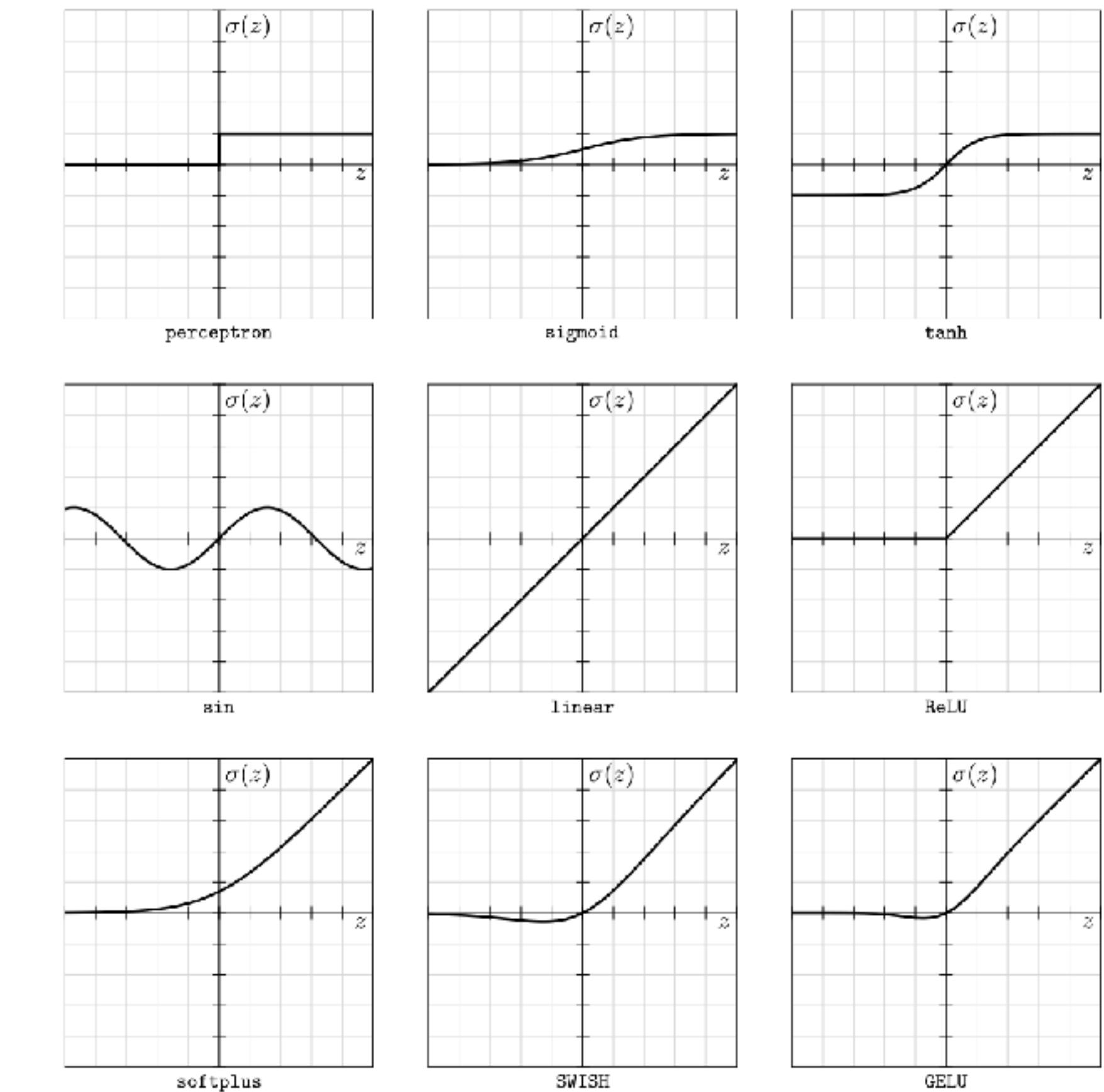
Méthodes/modèles en ML

Réseau de neurones

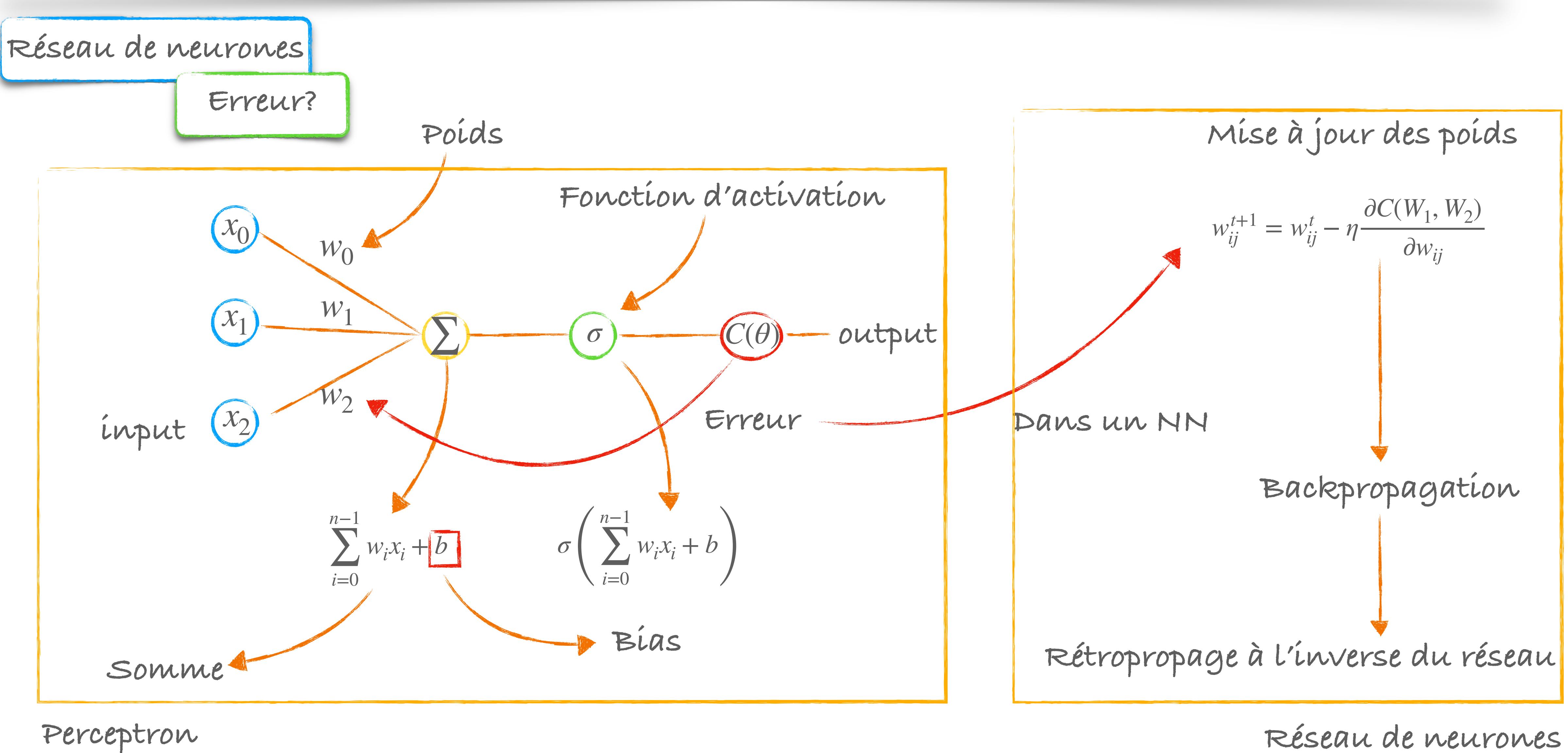
Perceptron



Fonctions d'activation les plus communes

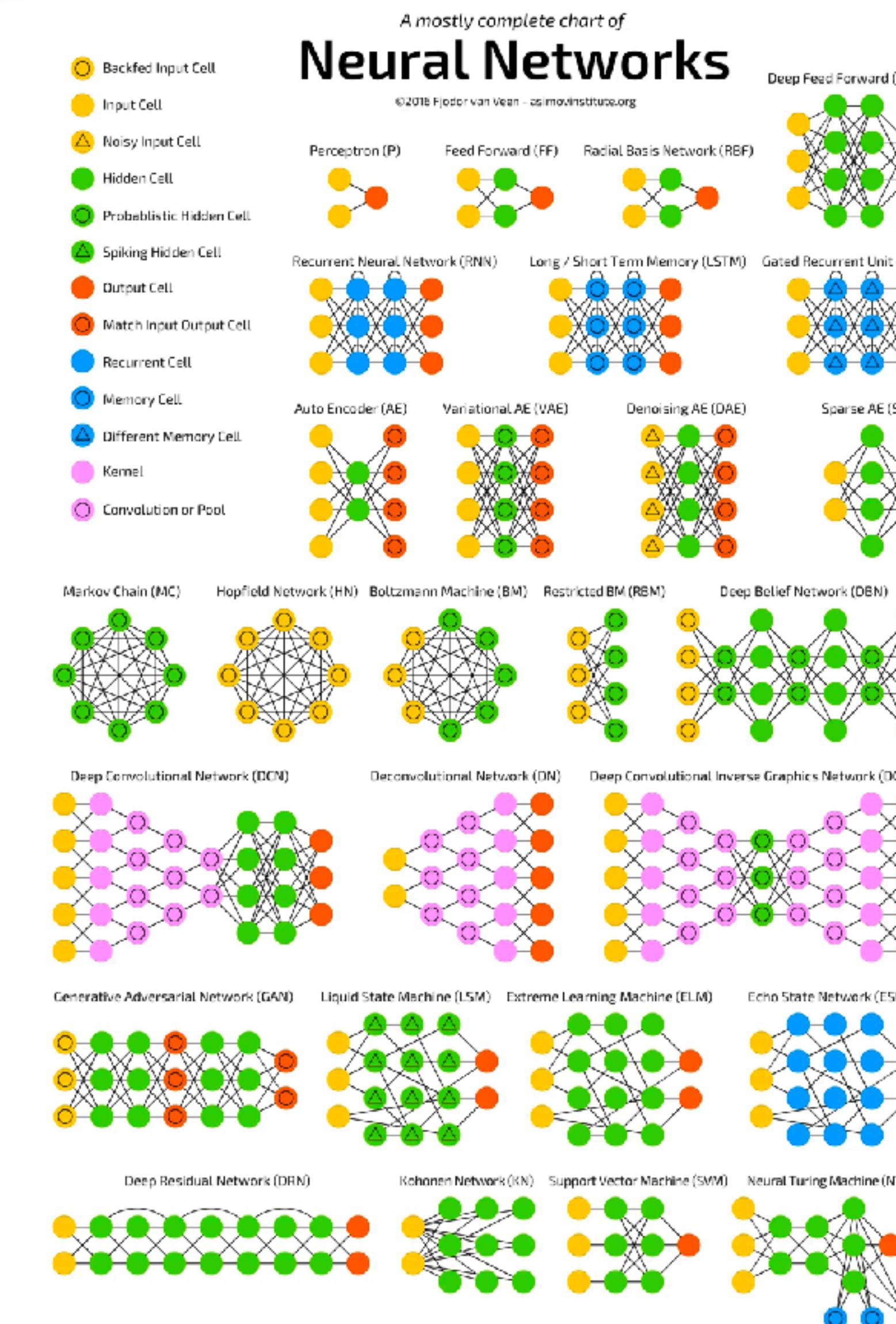


Méthodes/modèles en ML



Méthodes/modèles en ML

Types de réseaux



No-free lunch theorem

No-free lunch theorem

Le théorème "no free lunch" (NFL) pour l'apprentissage automatique supervisé est un théorème qui implique essentiellement qu'aucun algorithme d'apprentissage automatique n'est universellement le plus performant pour tous les problèmes.

Ce théorème vient des hypothèses sous-jacente de la corrélation entre données et étiquettes estimée pour chaque modèle.

No-free lunch theorem

Traduction

un modèle spécifique pour
un problème spécifique

Références

- [1] Schuld & Petruccione, 2021, Machine Learning with Quantum Computers, Springer, Chap 2
- [2] Togootogtokh & Amartuvshin, 2018, Deep Learning Approach for very similar objects Recognition Application on chihuahua and Muffin Problem, <https://arxiv.org/pdf/1801.09573.pdf>
- [3] Tenenbaum, J. (1999). A Bayesian framework for concept learning. Ph.D. thesis, MIT.
- [4] Roberts & Yaida, 2021, The Principles of Deep Learning Theory, <https://arxiv.org/pdf/2106.10165.pdf>

- [3] <https://www.google.com/url?sa=i&url=https%3A%2F%2Fhotcore.info%2Fbabki%2Fhuman-development-life-cycle.htm&psig=AOvVaw1xjQ83A2Wb82sqWowboCES&tust=1698628293826000&source=images&cd=vfe&opi=89978449&ved=OCB1QjRxqFwOTCLCYmdKJm0IDFQAAAAAAadAAAAABAL>