EXTRAIT (document original : 82 pages)

Session 2 : Apprendre aux ordinateurs à apprendre

Stage d'initiation à l'intelligence artificielle --- IPESUP

EXTRAIT (document original : 82 pages)

SOMMAIRE

01 **APPRENTISSAGE SUPERVISÉ**
Comment apprendre à répondre à des questions ?

02 **ALGORITHME D'APPRENTISSAGE**
Comment trouver un modèle qui colle aux données d'entraînement ?

03 **AUTRES TYPES D'APPRENTISSAGE**
Que faire quand l'apprentissage supervisé n'est pas adapté ?

04 **L'HISTOIRE DES RÉSEAUX DE NEURONES**
Les algorithmes derrière la plupart des succès de l'IA moderne

05 **DIFFÉRENTS TYPES DE RÉSEAUX**
Les différents types d'architecture

EXTRAIT (document original : 85 pages)

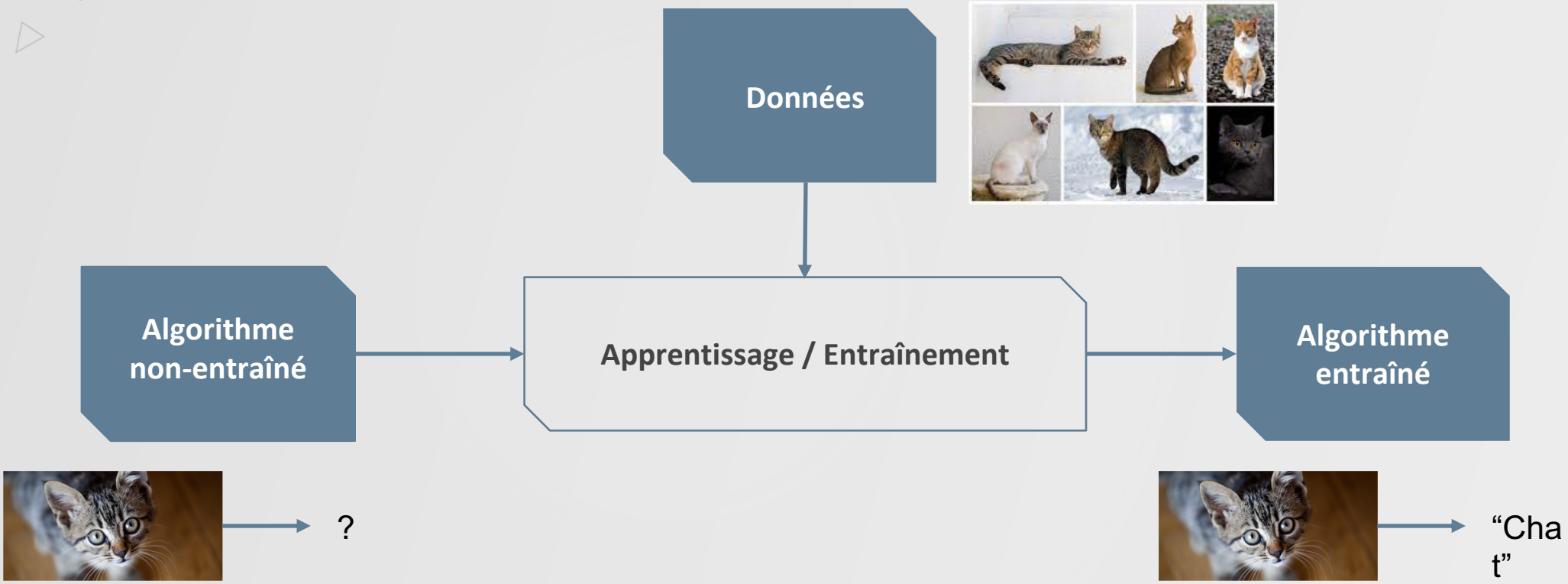
01

APPRENTISSAGE SUPERVISÉ

Comment apprendre à répondre à des questions ?

EXTRAIT (document original : 82 pages)

L'apprentissage automatique (machine learning)



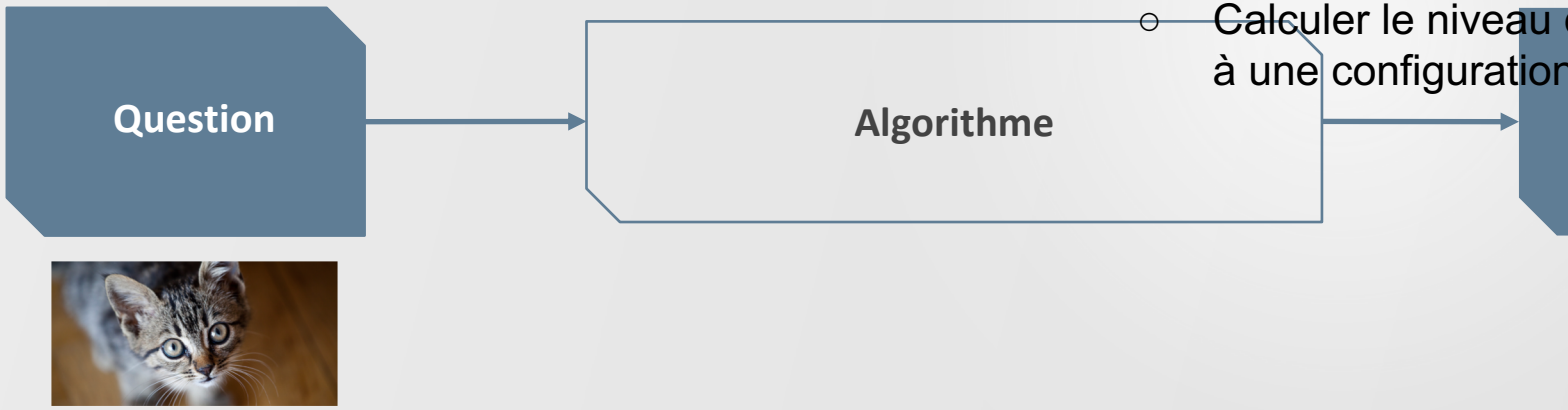
EXTRAIT (document original : 82 pages)

L'apprentissage supervisé : Questions et réponses

En **apprentissage supervisé**, l'algorithme doit répondre à des questions. Les données prennent la forme d'exemples de questions avec leurs réponses.

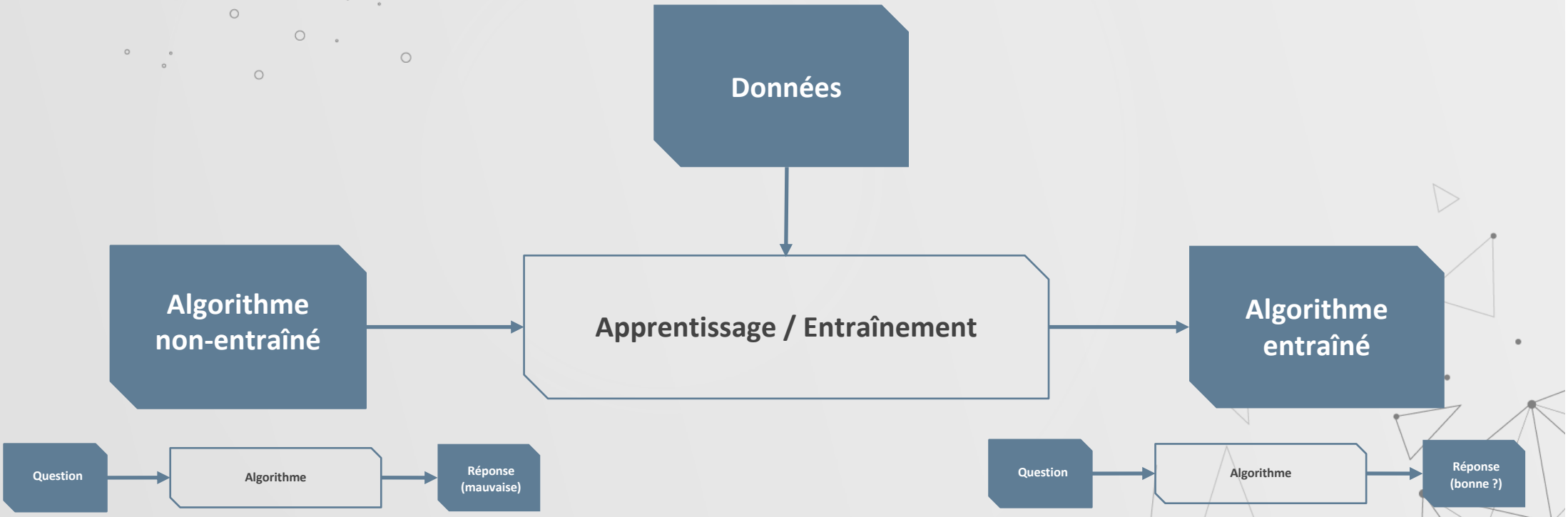
Exemples :

- Classification :
  - Dire à quelle classe (parmi "chat", "chien", "bateau", …) une image appartient
  - Donner le style musical d'une chanson
- Régression :
  - Prédire le nombres de vues d'une vidéo
  - Calculer le niveau d'énergie associée à une configuration de molécule



EXTRAIT (document original : 85 pages)

Apprendre à répondre aux questions



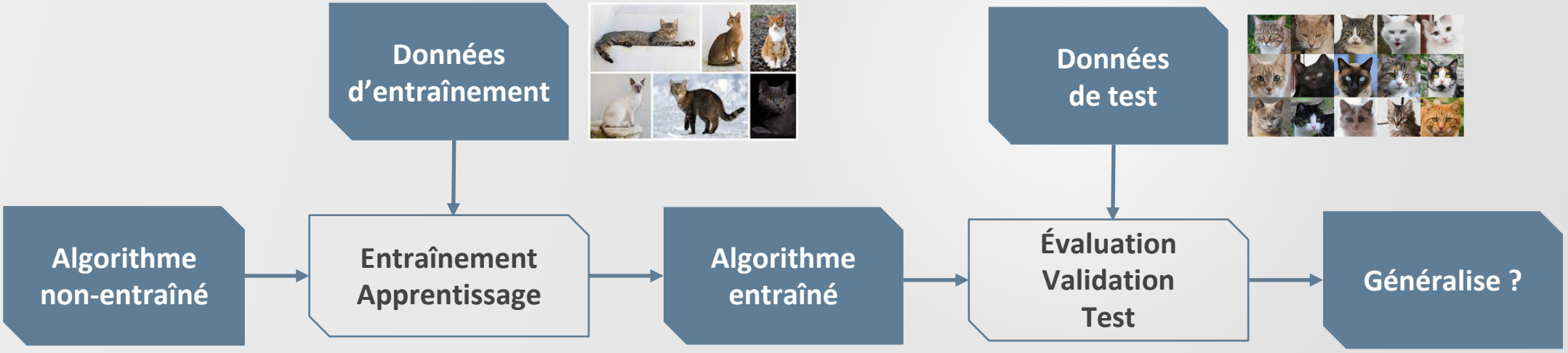
Comment savoir si l'algorithme a vraiment appris ?
On veut distinguer la mémorisation "par coeur" de la "compréhension" !

EXTRAIT (document original : 82 pages)

Évaluer la réponse aux questions

Pour évaluer si l'algorithme a bien appris, il faut utiliser de nouvelles données que l'algorithme n'a jamais vu.

On découpe donc les données en **données d'entraînement** et en **données de test**. Un algorithme qui ne fait que mémoriser les réponses des exemples d'entraînement échouera, on dit qu'il ne **généralise** pas à de nouvelles questions.



EXTRAIT (document original : 85 pages)

L'apprentissage supervisé : l'algorithme à entraîner

Question

Paramètres

Algorithme

Réponse

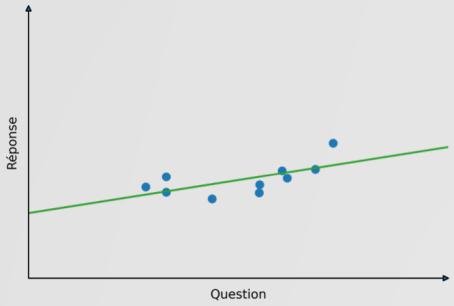
"Chat"

L'entraînement consiste à ajuster les paramètres pour que l'algorithme donne les meilleures réponses possibles.

L'algorithme est une **fonction paramétrée**.

EXTRAIT (document original : 82 pages)

Exemples de fonctions paramétrées



Linéaire / affine

Exemple : on doit prédire le prix P de revente d'un diamant en fonction de sa taille T. On peut modéliser cela par une relation affine

$$P = aT + b$$

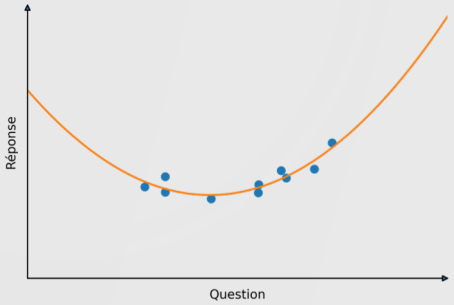
Les paramètres a et b sont à ajuster.

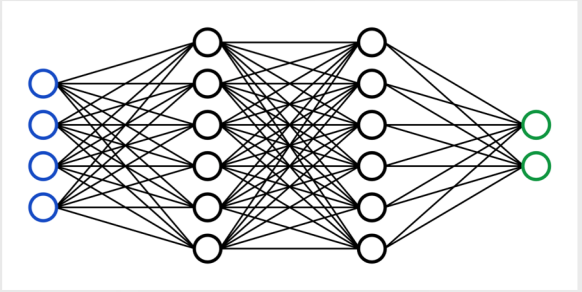
Les polynômes sont des fonctions plus compliquées, avec plus de paramètres.

Dans l'exemple précédent, cela donne :

$$P = aT^2 + bT + c$$

Polynômes



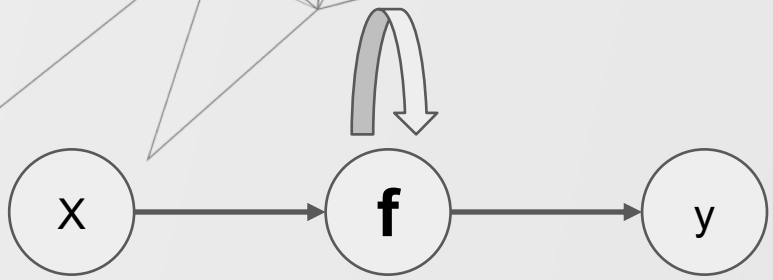


Réseaux de neurones

Chaque neurone implémente une fonction paramétrée simple (par exemple : linéaire ou affine). Les paramètres peuvent être vues comme les forces des connexions entre les neurones.

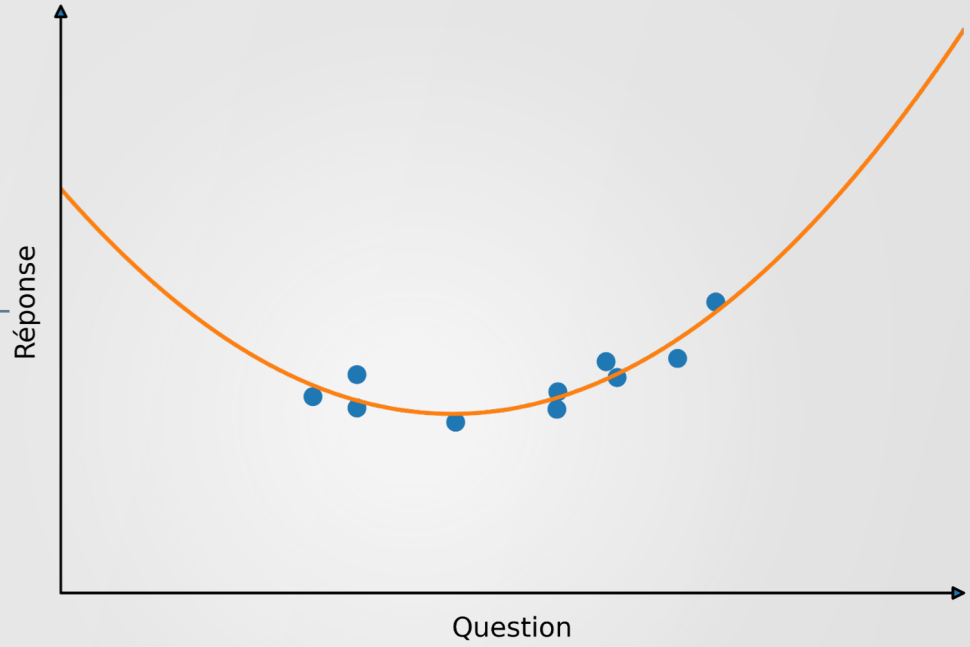
EXTRAIT (document original : 82 pages)

L’apprentissage, c’est de l’interpolation !



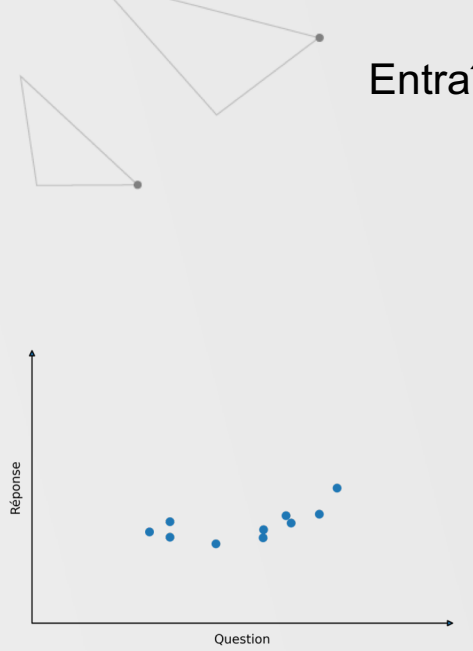
Interpolation et extrapolation

Étant donné des points et la valeur de la fonction en ces points, il s’agit d’estimer la valeur de la fonction pour des nouveaux points. Cela revient à proposer un modèle de la dépendance entre la question et la réponse (droite, parabole, …).



EXTRAIT (document original : 85 pages)

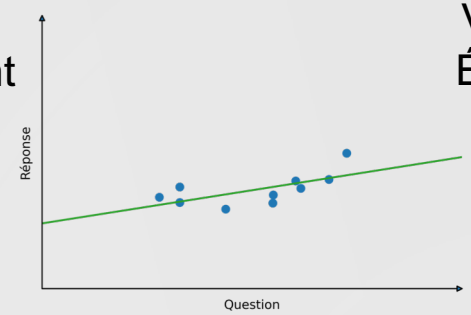
Sous-, sur-apprentissage et généralisation

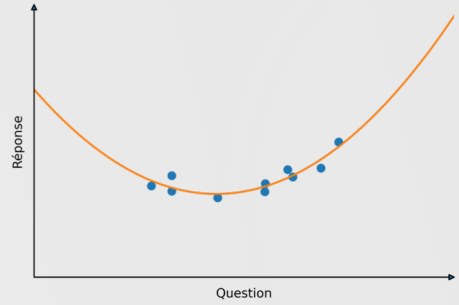


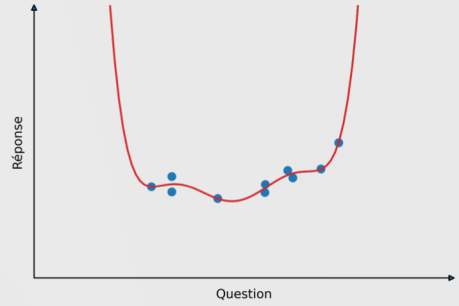
Données d'entraînement

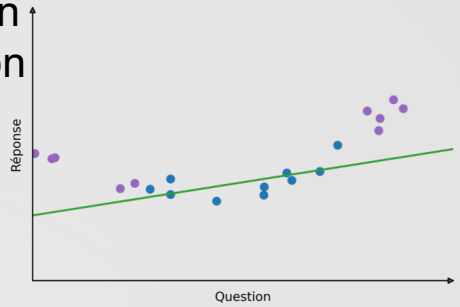
Entraînement

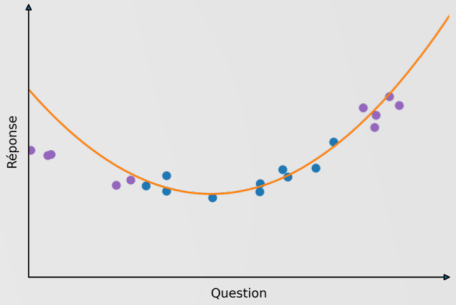
Validation
Évaluation
Test

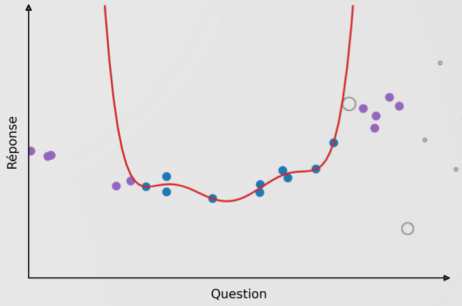












Sous-apprentissage (underfitting) : modèle trop simple

Généralisation : bon modèle

Sur-apprentissage (over-fitting) : modèle trop compliqué

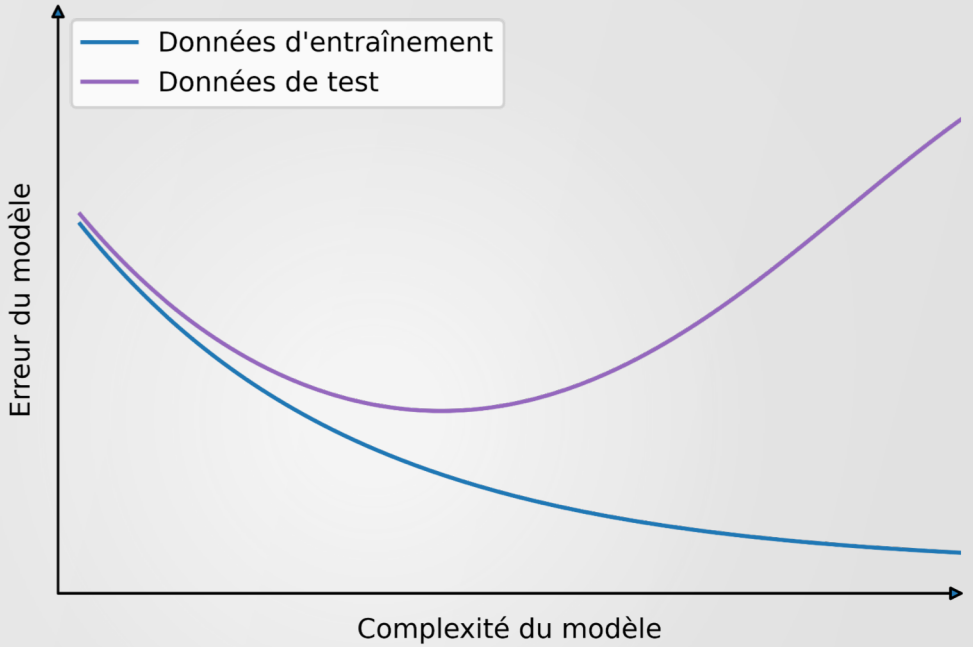
EXTRAIT (document original : 82 pages)

Résumé : erreur en fonction de la complexité du modèle

Compromis sur la complexité

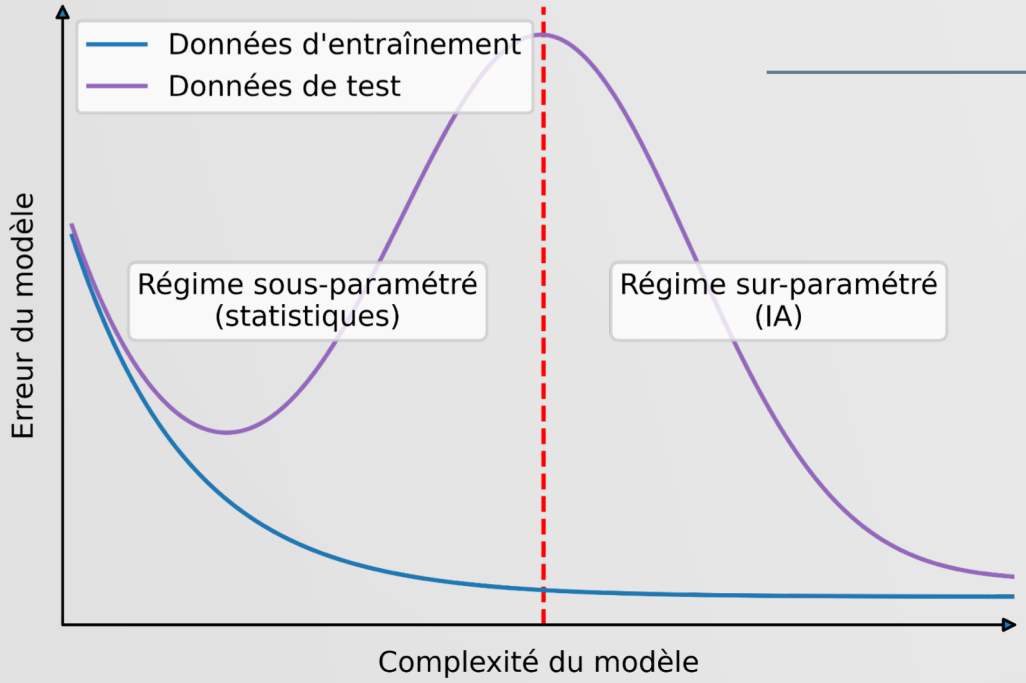
Il y a un compromis à faire sur la complexité du modèle : trop simple ou trop compliqué, et l'erreur de test est grande. Il y a une complexité optimale entre les deux.

En statistiques, on étudiait déjà cela à la fin du XIXème siècle ! Quelle est la différence avec l'IA moderne ?



EXTRAIT (document original : 85 pages)

La généralisation, c’est plus compliquée que cela...



Phénomène de la “double descente”

En considérant des modèles de plus en plus complexes, l’erreur sur les données de test peut parfois redescendre, et atteindre une erreur plus faible que la première vallée. C’est dans ce régime qu’est l’IA moderne, et c’est lié avec ses performances spectaculaires.

Ce phénomène reste encore mal compris, et fait partie des questions de recherche ouvertes en IA.

EXTRAIT (document original : 85 pages)

Différences entre IA et statistiques

Un problème d'apprentissage est plus ou moins caractérisé par 3 nombres :

- Le nombre d'exemples d'entraînement n
- Le nombre de paramètres (degrés de libertés) du modèle p
- La dimension (taille) des données d'entrée d

Le cadre est celui du régime sous-paramétré : p < n.
Les données sont souvent en petite dimension : d < 10.
Cadre statistique classique (compris)

Le cadre est souvent celui du régime sur-paramétré : p > n.
Les données sont habituellement en grande dimension : d > 10000.
Cadre IA moderne (recherche active)

EXTRAIT (document original : 82 pages)

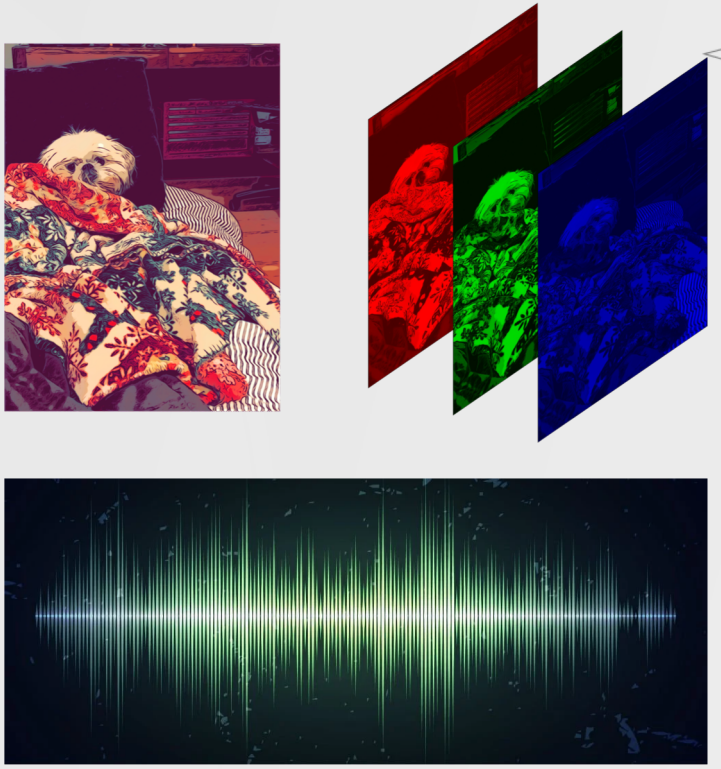
Dimension des données

Les données d'entrée (questions) prennent la plupart du temps la forme d'un tableau de nombres, ou **vecteur**.

- Pour une image, il s'agit de l'intensité de rouge, vert et bleu de chaque pixel (généralement 1 000 000 de pixels).
- Pour du son, il s'agit de la différence de pression de l'air à chaque instant (généralement 44 100 échantillons par seconde).

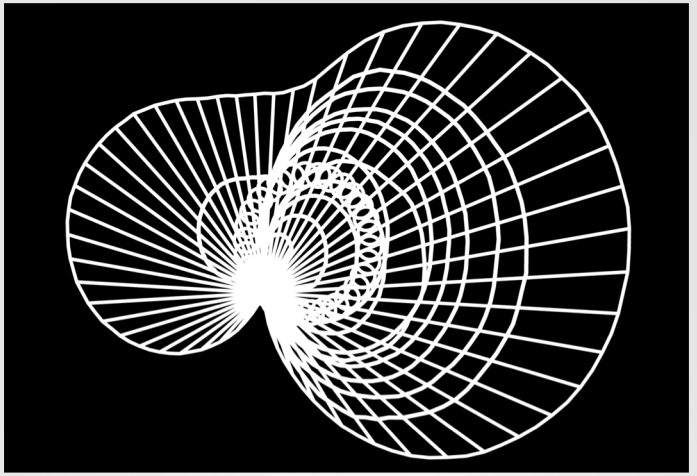
La **dimension** est la taille de ce tableau. L'IA traite souvent des données en haute dimension.

Exemple de données qui ne sont pas naturellement sous cette forme ?



EXTRAIT (document original : 82 pages)

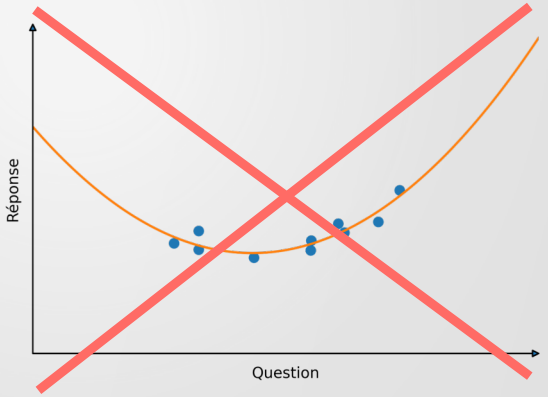
La malédiction de la dimension



En grande dimension

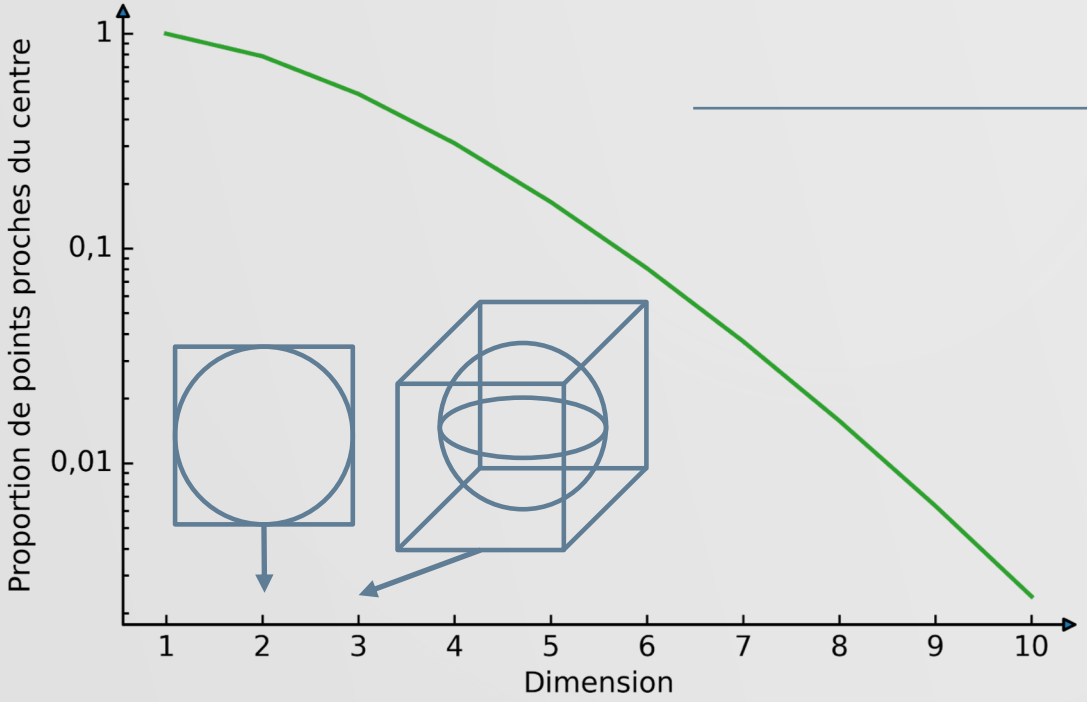
En grande dimension, les notions de distances, de points proches, de volume, ne correspondent plus à notre intuition. Elles ont un comportement très différent de leurs équivalents en petite dimension.

Par exemple, dans une base de données d'images, il n'y a jamais deux images proches ! Notre intuition ne fonctionne donc plus



EXTRAIT (document original : 85 pages)

La malédiction de la dimension



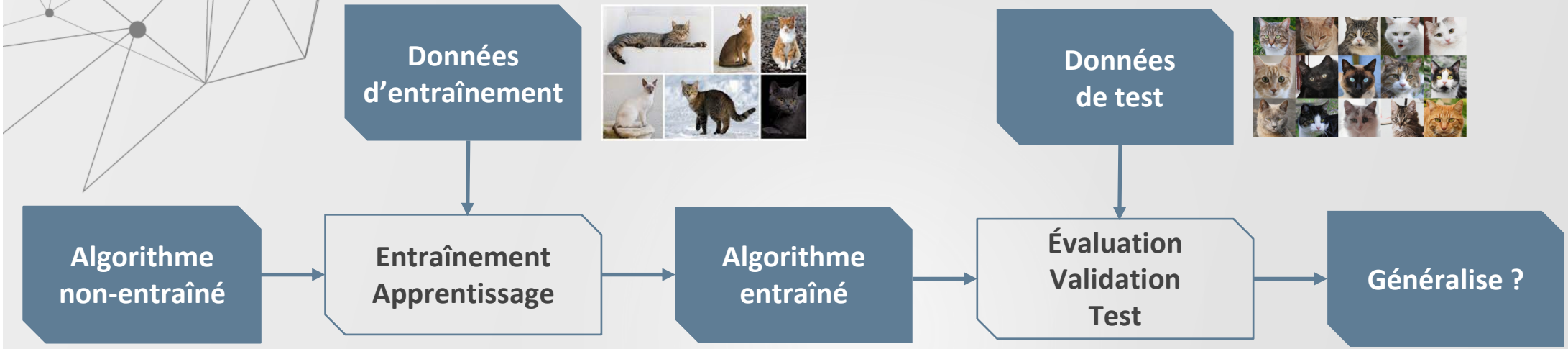
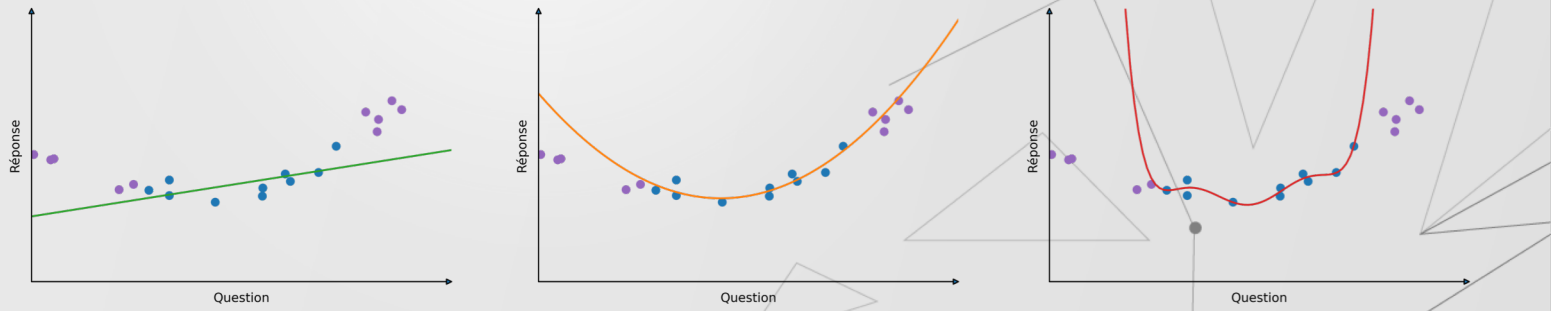
En grande dimension

Prenons des points au hasard dans un carré et comptons la proportion qui est proche du centre.

Nous pouvons passer de la dimension 2 à la dimension 3 en regardant un cube. De même, on peut poser cette question en dimension supérieure.

EXTRAIT (document original : 85 pages)

Résumé de l'apprentissage supervisé



Données en **grande dimension** : problème difficile, intuition ne fonctionne plus.