

Intelligence artificielle

Fiche de synthèse

① APPRENTISSAGE AUTOMATIQUE

Un programme classique résout un problème dont on connaît la méthode : calculer une moyenne. Pour reconnaître un chiffre écrit à la main, on ne sait pas écrire cette méthode ; mais on dispose de très nombreux exemples.

Définition 1 (Apprentissage automatique)

L'**apprentissage automatique** consiste à écrire un algorithme qui, à partir de données et d'expériences passées, apprend à résoudre au mieux un problème.

② MODES ET TYPES DE PROBLÈMES

- **Apprentissage supervisé** : les exemples sont étiquetés par des humains, « chat » ou « chien » ; le nombre de classes est connu.
- **Apprentissage non supervisé** : l'algorithme trouve lui-même des groupes de données semblables, c'est le **partitionnement** (*clustering*).
- Un problème de **classification** donne une valeur parmi quelques possibilités, chat ou chien ; un problème de **régression** donne un nombre réel, comme le prix d'un logement d'après son quartier et sa surface.

③ ENTRAÎNER ET TESTER

Entraîner un modèle, c'est ajuster ses paramètres pour que ses prédictions se trompent le moins possible sur les exemples. On ne peut pas juger le modèle sur ces mêmes exemples : il les a en quelque sorte appris par cœur.

- Les **données d'entraînement**, environ 80 % des données, servent à ajuster le modèle.
- Les **données de test**, les 20 % restants, ne sont jamais montrées pendant l'entraînement ; le taux d'erreur sur ces données mesure honnêtement la qualité du modèle. On ne modifie plus le modèle après l'avoir testé.

Algorithmes supervisés : régression linéaire, arbres de décision, k plus proches voisins. Algorithme non supervisé : les k -moyennes, qui regroupent les données proches selon une distance.

④ LES K PLUS PROCHES VOISINS

≡ Algorithme des k plus proches voisins.

1. Choisir un nombre k de voisins et une **distance** entre les données.
2. Calculer la distance entre la nouvelle donnée et chaque donnée étiquetée.

3. Garder les k données les plus proches.
4. Attribuer à la nouvelle donnée la classe la plus représentée parmi elles : les voisins « votent ».

Avec $k = 5$, si trois voisins sont des chats et deux des chiens, la nouvelle image est classée « chat » ; avec $k = 3$, le résultat peut changer. Le réglage consiste à choisir les caractéristiques pertinentes, une bonne valeur de k , ni trop petite ni trop grande, et une distance adaptée : euclidienne, de Manhattan.

⑤ USAGES ET QUESTIONS

- **Apports** : aide au diagnostic médical, traduction, recommandations, prévision des embouteillages, assistants conversationnels, voitures autonomes.
- **Questions** : décisions automatisées en justice ou en police, biais des données, confidentialité, droit d'auteur, consommation d'énergie de l'entraînement.

Un **modèle de langage**, comme ChatGPT, Claude, Gemini ou Mistral, est un réseau de neurones entraîné sur des milliards de textes à prédire le mot suivant. Il **génère** une suite plausible sans consulter de base de faits : il peut affirmer avec assurance une chose fausse, c'est une **hallucination**. Ses connaissances s'arrêtent à ses données d'entraînement, sauf accès à des documents ou à la recherche en ligne.

📌 À retenir.

- L'apprentissage automatique apprend à partir d'exemples, supervisés ou non.
- Classification : une classe ; régression : un nombre.
- On sépare les données en entraînement, 80 %, et test, 20 %.
- Les k plus proches voisins font voter les k exemples les plus proches.