ISC9000 Connaissance, Raisonnement et Prise de Décision

L'apprentissage machine, principes et conséquences pour la génération de connaissances

Roger Villemaire

Département d'informatique
UQAM

9 avril 2026

UQÀM
Université du Québec à Montréal

Plan

Plan

Apprentissage Machine/Automatique

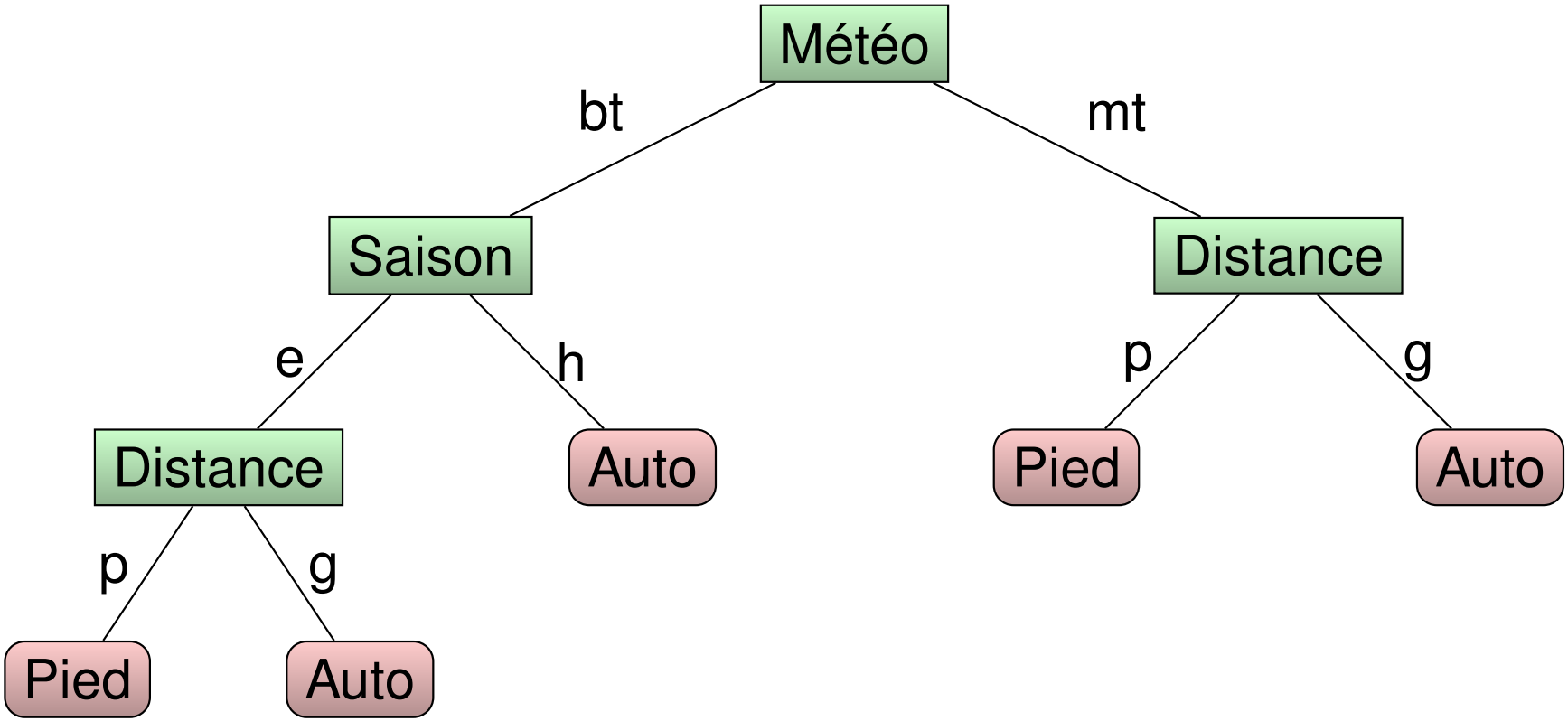
- Réaliser une fonction intelligente (relevant de la cognition humaine) en exploitant un jeux de données (conséquent).
- Il y a trois types d'apprentissage machine :
  - supervisé : le jeux de données contient la bonne réponse, ex. symptômes/traitement,
  - non-supervisé : on regroupe les données par similitude (clustering), ex. clients regroupés par types d'achat,
  - par renforcement : on reçoit une récompense/pénalité au fur et à mesure que l'on prend des décision, ex. ping-pong électronique.
- Nous allons nous limiter à l'apprentissage supervisé qui a fait des avancées prodigieuses dans les dernières décennies.

Apprentissage machine supervisé

- Objectif : déterminer une fonction f
  - à partir d'un jeu de données étiquetées (valeurs correctes) $f(x_1) = y_1, \ldots, f(x_n) = y_n$,
- Usage : prédire de nouvelles valeurs $f(x)$.

Plan

Arbre de décision



Apprentissage supervisé pour construire un arbre de décision

- On reçoit en entrée un certain nombre de décisions correcte, par exemple
  - (Météo=bt,Saison=e,Distance=g,Décision=Auto)
- L'algorithme construit un arbre de décision qui pourra être utilisé pour trancher dans des cas futurs.

L'algorithme C4.5 [1]

- Choisir le paramètre le plus "crucial" et décomposer les données d'entraînement selon les valeurs de ce paramètre.
  - crucial = principalement celui qui offre le meilleur *gain d'information*, selon la théorie de Shannon,
  - avec une compensation pour ne pas favoriser indûment les attributs qui branchent énormément.
- Recommencer la même procédure pour construire les sous-arbres.
- Arrêter à un certain niveau et prendre la décision majoritaire.

1. Quinlan, J. Ross ; C4.5 : Programs for Machine Learning ; Morgan Kaufmann ; 1993

XAI Explainable AI

- Chaque branche d'un arbre de décision représente une règle,
- on peut donc justifier la décision,
- même si, en général, il va toujours y avoir des décisions erronées.

Plan

Apprentissage machine supervisé : méthodologie paramétrique

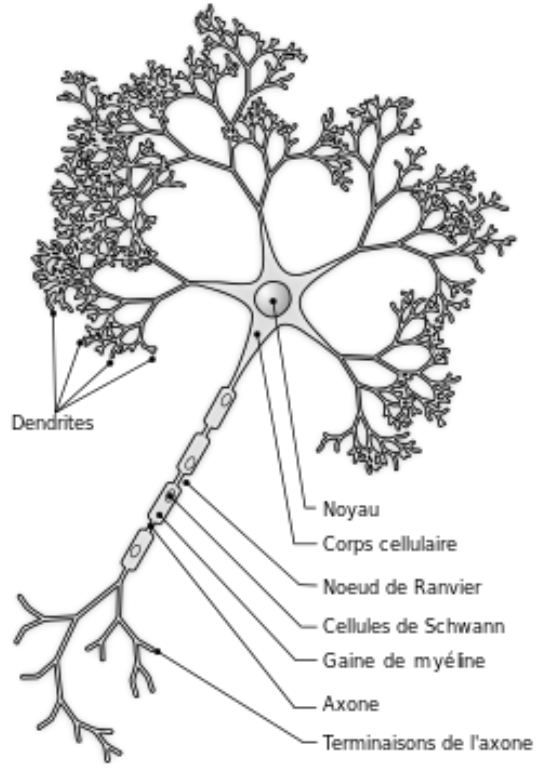
- Pour une famille de fonctions $h_{\bar{a}}(x)$, trouver des valeurs de paramètres $\bar{a}$ pour lesquels $h_{\bar{a}}(x)$ approxime bien la fonction cherchée $f(x)$:
- on divise notre jeu de données en un ensemble d'entraînement (la majeure partie) et un ensemble de test (le reste),
- itérativement, pour chaque donnée x de l'ensemble d'entraînement on compare $f(x)$ et $h_{\bar{a}}(x)$ (à l'aide d'une fonction de perte) et on corrige (légèrement) les paramètres $\bar{a}$.
- À la toute fin, on compare $f(x)$ et $h_{\bar{a}}(x)$ sur l'ensemble de test pour évaluer la qualité de la prédiction.

Correction des paramètres

- En pratique la méthode de correction de l'erreur la plus populaire est celle dite de la *descente de gradient stochastique* :
  - l'erreur (différence entre la valeur obtenue et attendue, par la fonction de perte) est corrigée
  - en distribuant une (petite) correction sur l'ensemble des poids, en proportion de l'impact sur l'erreur.
- Ce processus est réitéré sur chaque valeur de l'ensemble d'entraînement jusqu'à épuiser un certain temps ou obtenir une certaine qualité de réponse.

Neurone : fonctionnement sommaire

- Les *dendrites* conduisent l'influx nerveux venant des autres neurones au corps du neurone,
- en fonction des signaux reçus, le neurones va développer un *potentiel d'action*,
- l'unique axone décharge périodiquement le potentiel d'action par ses *terminaisons synaptiques* vers les autres neurones.
- **Note :** L'apprentissage est réalisé par le renforcement/affaiblissement des connections synaptiques.

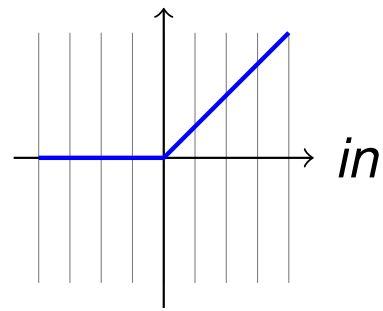


© Nicolas Rougier

Neurone Artificiel

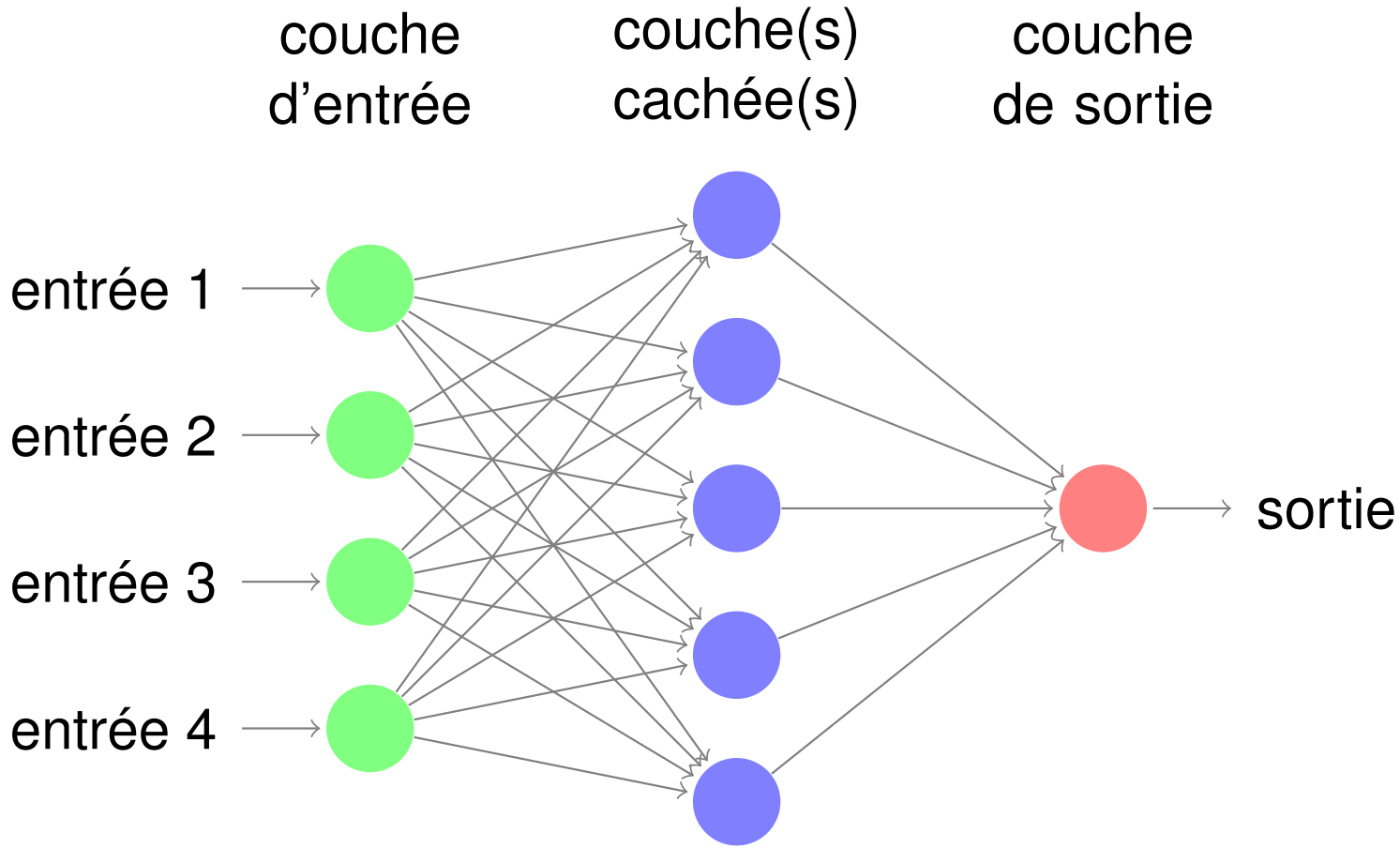
- Un neurone artificiel est formé
  - d'un nœud avec des entrées $in_1, \ldots, in_n$ de poids $w_1, \ldots, w_n$ (des nombres réels)
- Le neurone (ou nœud) calcule sa sortie (ou *niveau d'activation*) en calculant :
  - la somme pondérée de ses entrées (plus un poids de biais) $in = \sum_{j=1}^{n} w_j \cdot in_j + b$,
  - la valeur $g(in)$ de sa *fonction d'activation* g appliquée à in.
- La fonction d'activation est aujourd'hui très souvent *ReLU* ($max(0, x)$, *Rectified Linear Unit*).

$ReLU(0, in)$



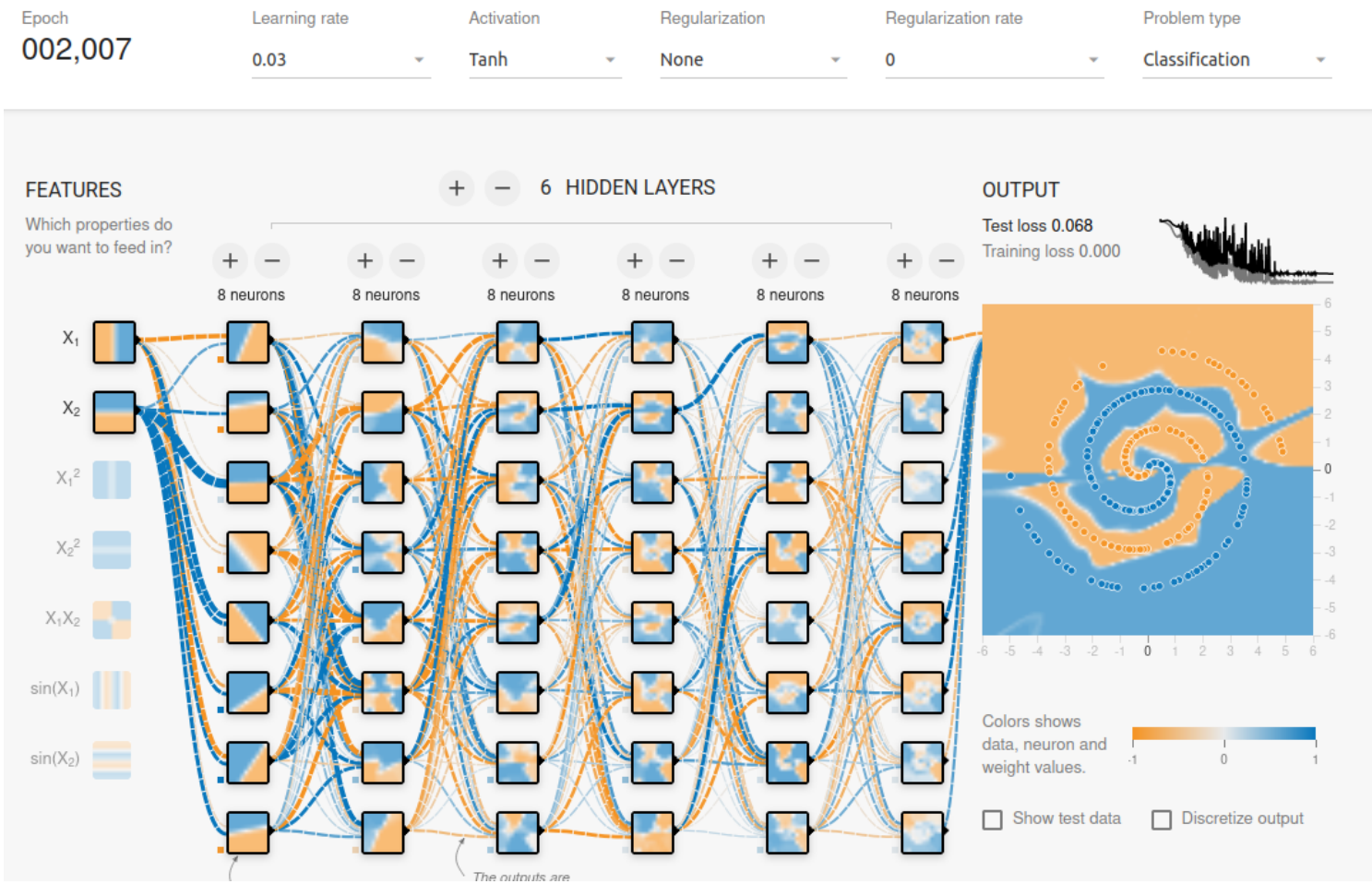
Réseaux multi-couches

- La plupart des applications utilisent un réseau (feed-forward) structuré en couches :



Playground TensorFlow

- playground.tensorflow.org/



UQÀM
Université du Québec à Montréal

Interprétation

- Plus on avance dans le réseau, plus la fonction apprise est complexe,
- il s'agit donc de composer des fonctions de plus en plus complexes jusqu'à obtenir un bon résultat.
- "Apprendre" signifie donc "performer".
- La perte de test (test loss) est sur les données non-utilisées dans l'entraînement, ceci mesure donc la capacité à généraliser.

Exemples

- Reconnaissance de l'écriture (codes postaux) :
  - entrées : matrice de pixels,
  - sortie : un nombre $[0, 9]$.
- Voiture qui se conduit elle-même :
  - entrées : matrice de pixels de l'image vidéo (éventuellement radar, détecteurs etc.),
  - sortie : instructions pour opérer le volant (tourner à droite, gauche) et contrôler l'accélération, le freinage, etc.
- IA générative :
  - entrées : les mots vus jusqu'ici,
  - sortie : le prochain mot.

Plan

Apprentissage Profond (Deep learning) [2]

- Nombre très important de couches, et de nœuds !
- Architecture : on peut choisir le type d'interconnexion entre chaque couche.
- Nécessite beaucoup de données et de temps pour l'entraînement. Un Grand Modèle de Langage (LLM, Large Language Model) prend des mois à entraîner.
- Retour en force des RNA dans les années 2000, car on a les données (Bigdata) et la puissance de calcul nécessaire (GPU).

2. Yann LeCun ; Cours apprentissage profond, Collège de France www.college-de-france.fr/site/yann-lecun/course-2015-2016.htm

Données Massives (Big Data)

- D'énormes quantités de données sont aujourd'hui disponibles :
  - textes en ligne : Wikipédia, journaux, livres, ces acétates !
  - par notre contribution en ligne : Google, Facebook, X (ex-Twitter), etc.
  - par des projets ciblés, par exemple :
    - ImageNet www.kaggle.com, etc.
- Idée : L'intelligence humaine sera reproductible bien plus grâce à d'énormes quantités de données qu'avec de nouveaux algorithmes.

Plongements

- On peut représenter les entrées par un vecteur (suite de nombres réels) quelconque mais si des termes “similaires” sont représentés par des vecteurs proches, l’apprentissage sera plus facile.
- Ceci vient du fait que deux vecteurs d’entrées proches seront envoyés sur des sorties proches, par un neurone artificiel. On devrait donc avoir besoin de moins de couches pour obtenir le même résultats.
- Ex. Si ’mardi’ et ’mercredi’ sont proches, il sera plus facile de répondre à la question “Est-ce que X est un jour de congé ?”.

Pré-entraînement

- Trouver un bon plongement est encore un problème d'apprentissage machine qui pourra même être effectué avec un jeu de données non-étiquetées !
  - Ex. On entraîne d'abord avec la tâche de prédire un mot masqué.
- De façon générale, on peut d'ailleurs post-entraîner un réseau déjà pré-entraîné.

Grands modèles de langage (LLM Large Language Models)

- Les grands modèles de langage (ChatGPT, Gemini, etc.) utilisent les méthodes de l'apprentissage profond pour *prédire* le prochain mot approprié en fonction des mots précédents.
- Un LLM n'effectue pas de recherche d'information à proprement parler, il ne fait que générer du texte à partir de l'invite (prompt) et possiblement du résultat de recherches sur l'internet effectuée à partir du texte de l'invite.
- Le corpus d'entraînement utilise des textes en format numérique, Wikipédia etc.

Limites de l'apprentissage machine

- L'apprentissage machine donne d'excellents résultats, avec suffisamment de données et de temps d'entraînement.
- Il reste qu'il n'y a jamais de garantie autre qu'expérimentale.
- Le contexte est aussi d'importance : un taux de succès de 75% est excellent pour un système de recommandation, mais évidemment largement insuffisant pour la voiture qui se conduit elle-même !
- Il ne faut aussi pas perdre de vue la nature de la fonction de perte qui permet d'optimiser le modèle : les *hallucinations* des LLM sont formées de mots qui vont très bien ensemble, même si l'énoncé est erroné !

Plan

Conclusion

- L'apprentissage machine est particulièrement adapté aux questions pour lesquelles on ne sait pas établir des algorithmes explicites.
- L'impact en est donc important en reconnaissance de l'écriture, de la parole, d'images et en production/traduction de textes.
- On tente d'ailleurs toujours d'en étendre la portée, comme on le verra à l'école d'été.
- Il reste, qu'il n'y a jamais de garantie intrinsèque de qualité, et que plusieurs méthodes, comme les réseaux neuronaux artificiels, sont des boîtes noires qui n'offrent pas de justification pour leurs résultats.