

Apprentissage Automatique I MATH60629

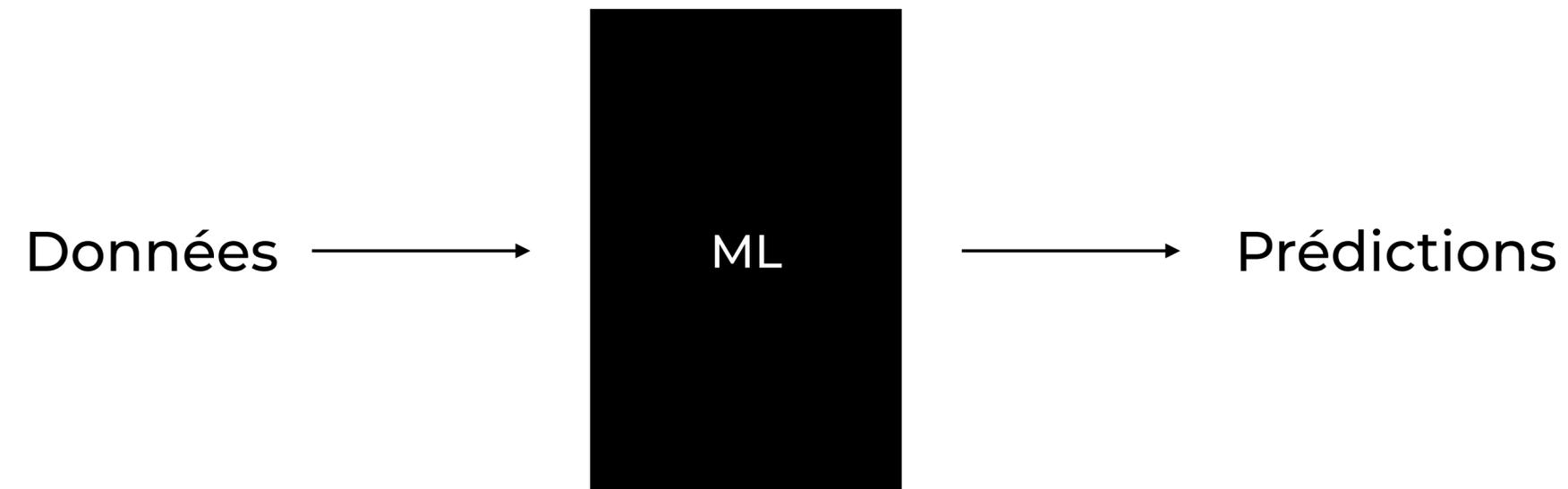
Semaine #1

Aujourd'hui

- **Introduction à l'apprentissage automatique**
- **Le plan de cours**
- **Révision mathématique (probabilité + algèbre linéaire)**

Apprentissage Automatique / Machine Learning (ML)

- Science qui étudie les aspects statistiques et computationnels de la modélisation des données pour la prédiction
- (Plutôt) Science empirique



- Tâche: Prédire si une image contient une tumeur
- Tâche: Prédire le prochain film qu'un utilisateur pourrait regarder

THIS IS YOUR MACHINE LEARNING SYSTEM?

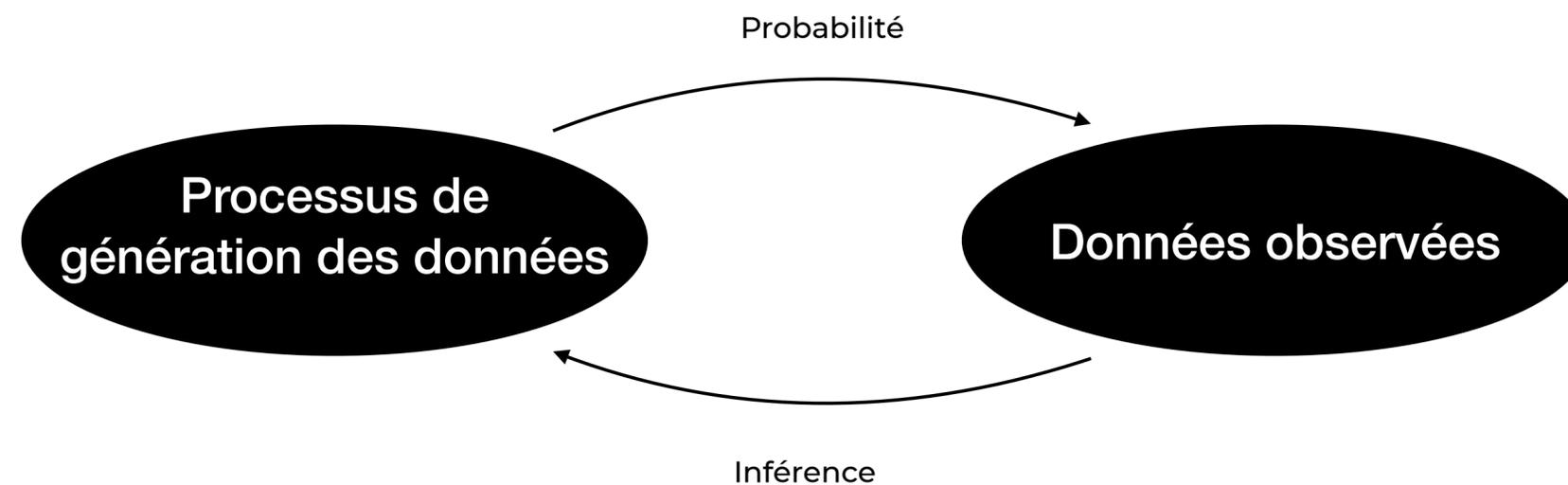
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.



“Data analysis, machine learning and data mining are various names given to the practice of statistical inference, depending on the context.”



–Larry Wasserman in “All of Statistics: A Concise Course in Statistical Inference.”

**Quel est le but de
l'apprentissage
automatique?**

ML et IA

- **Contexte historique**
 - **Quand j'ai commencé mon doctorat, peu de chercheurs osaient parler d'intelligence artificielle (IA)**
 - **Les progrès récents du ML sont clairement motivés par l'IA**
 - **Par exemple: traduction automatique, reconnaissance d'objets dans des images**
 - **Donc : veut-on créer une machine proche de l'humain? Ou peut-être une machine qui peut aider l'humain?**

- **Selon moi :**

Comprendre les données à travers les modèles prédictifs

Comprendre le monde à travers les modèles prédictifs

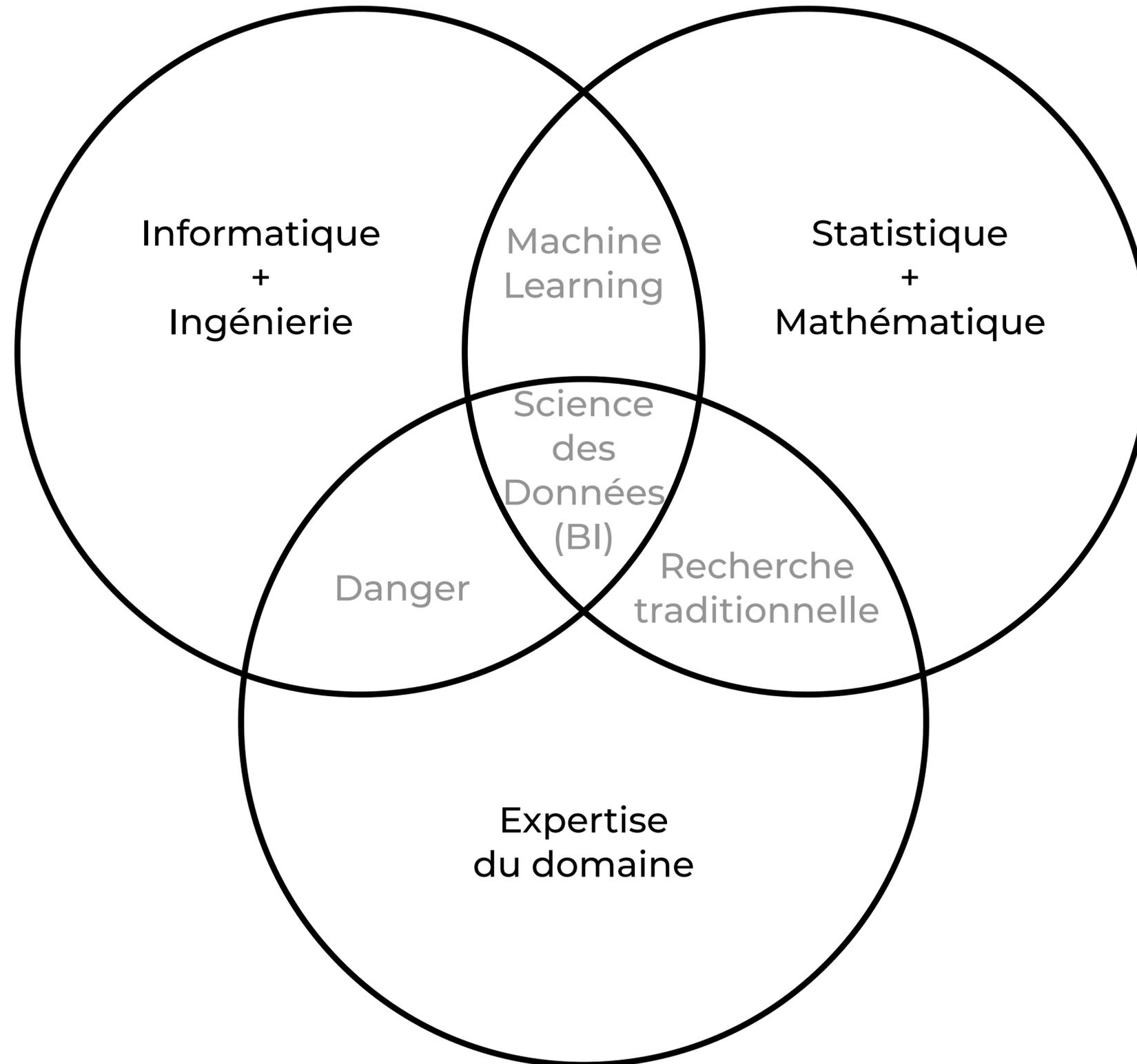
- **Dans le cours :**

Outils pour prédire des données non observées

Le ML par rapport à d'autres champs d'études

Perspective historique

- **(Moderne) Statistique: ~1900**
- **Machine Learning, IA et Data Mining: ~1960**
- **Science des données: ~2000**



Attitudes in Machine Learning and Data Mining Versus Attitudes in Traditional Statistics

Despite these differences, there's a big overlap in problems addressed by machine learning and data mining and by traditional statistics. But attitudes differ...

Machine learning

No settled philosophy or widely accepted theoretical framework.

Willing to use *ad hoc* methods if they seem to work well (though appearances may be misleading).

Emphasis on automatic methods with little or no human intervention.

Methods suitable for many problems.

Heavy use of computing.

Traditional statistics

Classical (frequentist) and Bayesian philosophies compete.

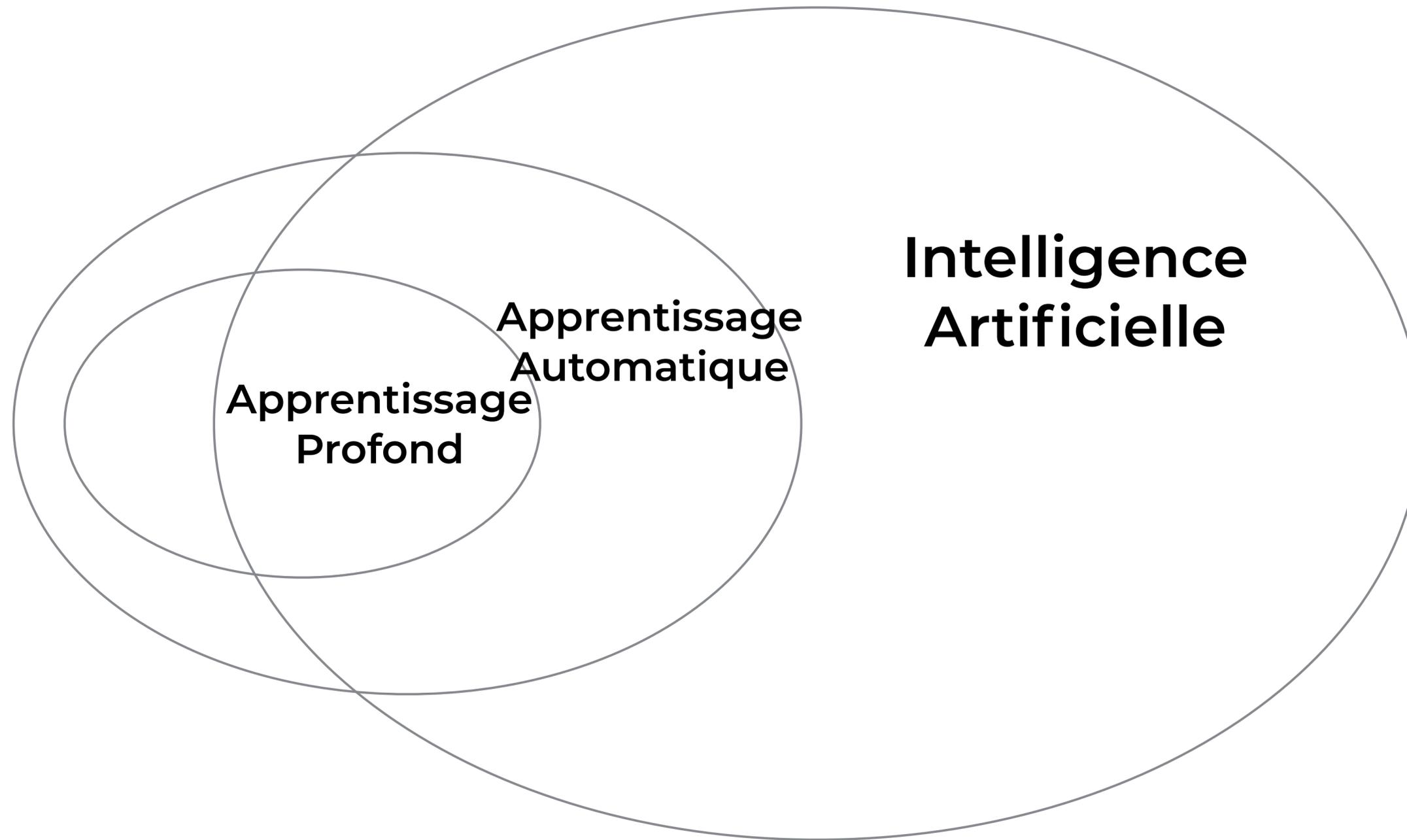
Reluctant to use methods without some theoretical justification (even if the justification is actually meaningless).

Emphasis on use of human judgement assisted by plots and diagnostics.

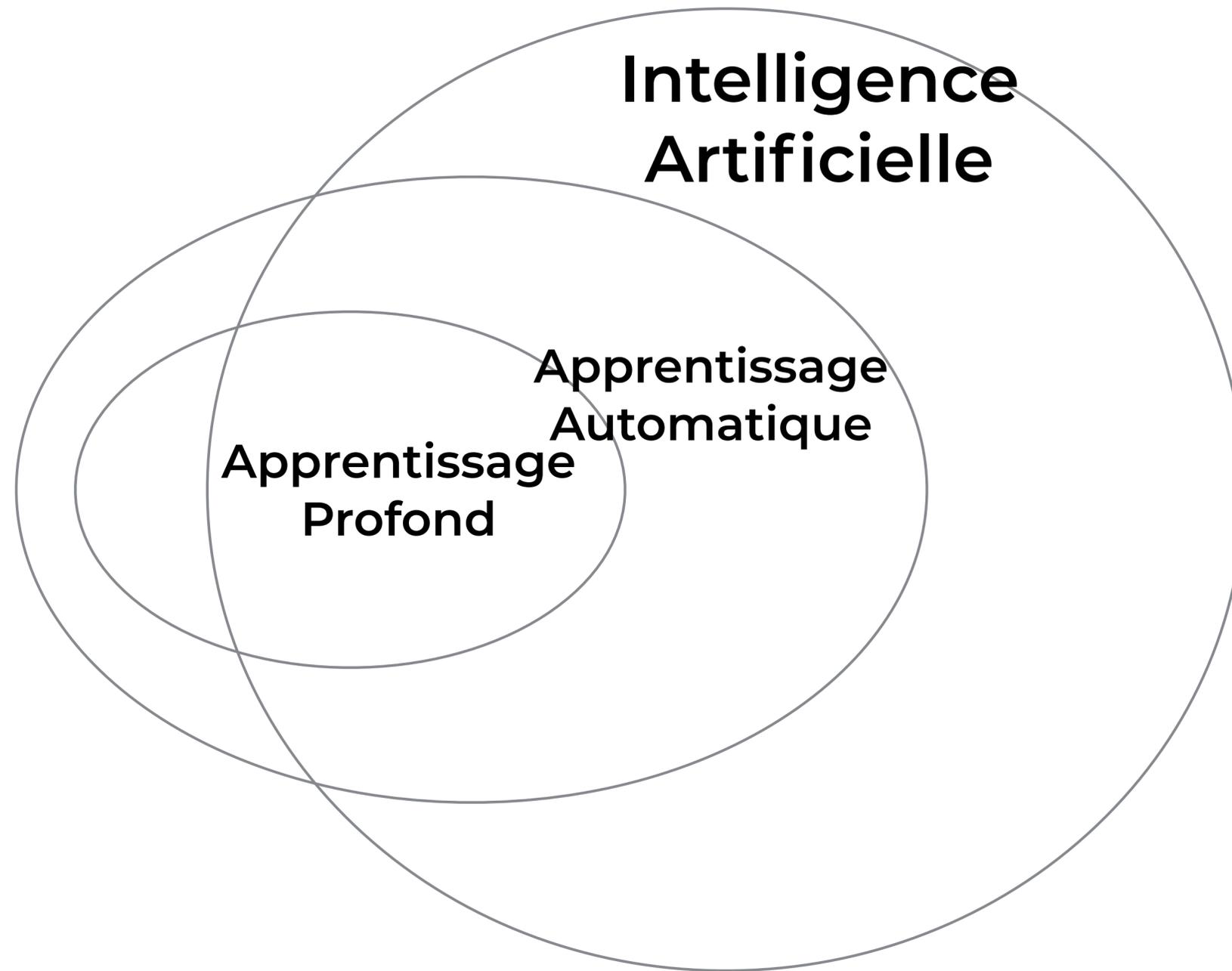
Models based on scientific knowledge.

Originally designed for hand-calculation, but computing is now very important.

2000



2024



Applications du ML



l'intelligence artificielle



l'intelligence artificielle

l'intelligence artificielle définition

l'intelligence artificielle tpe

l intelligence artificielle pdf

About 2,020,000 results (0.51 seconds)



Gmail ▾

COMPOSE

Inbox (40)

Sent Mail

Drafts (15)

All Mail

Spam (5)

Vos articles récemment vus et vos recommandations en vedette

Inspiré par votre historique de navigation

Page 2 sur 7 | Revenir



Playtex Diaper Genie Disposal System Refill, 3-Pack, Blue
★★★★☆ 79
CDN\$ 19.97 ✓Prime



MAVEA 1001122 Maxtra Replacement Filter for MAVEA Water Filtration Pitcher, 3-Pack
★★★★☆ 147
CDN\$ 19.99 ✓Prime



Kleenex Ultra Facial Tissue Flat Bundle, 70 count (Pack of 6)
★★★★☆ 19
CDN\$ 6.98 ✓Prime



Kleenex Facial Tissue Bundle, 85 Count (Pack of 10)
★★★★☆ 34
CDN\$ 10.93 ✓Prime



Dawn New Zealand Spring Scent Dishwashing Liquid 638mL
★★★★☆ 119
CDN\$ 2.47



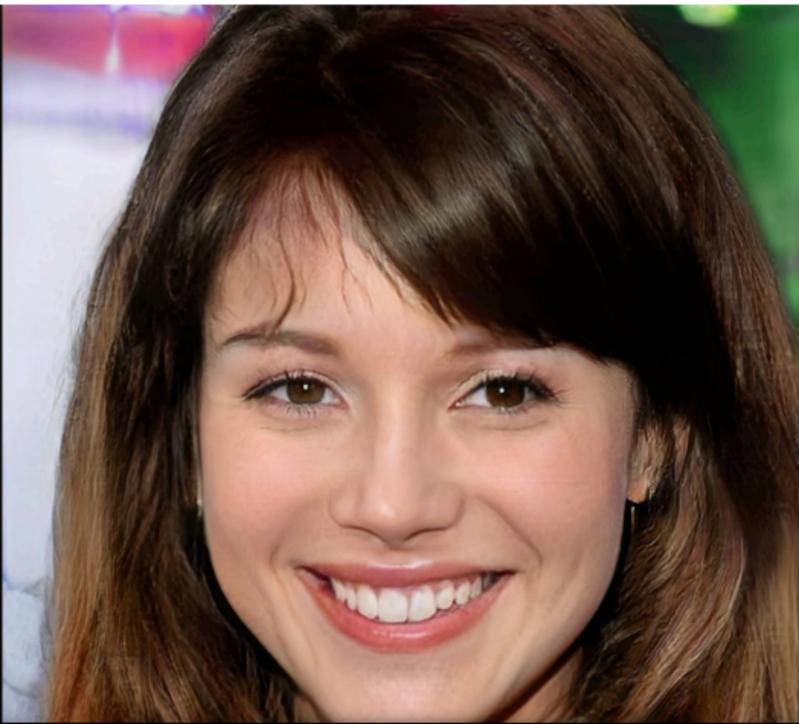
MAVEA 1001495 Maxtra Replacement Filter for MAVEA Water Filtration Pitcher, 1-Pack
★★★★☆ 147
CDN\$ 7.88 ✓Prime



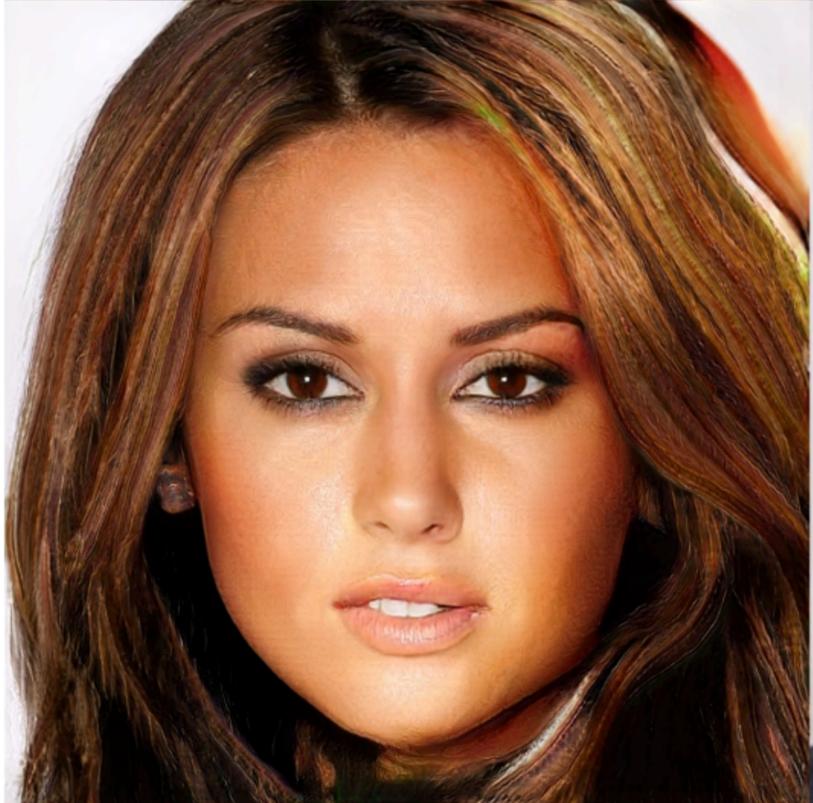
AmazonBasics Mini DisplayPort (Thunderbolt) to VGA Adapter
★★★★☆ 8
CDN\$ 20.99 ✓Prime



Medela Breastmilk Bottle Set 5oz.
★★★★☆ 8
CDN\$ 43.50



Toutes générées par ordinateur



Progressive Growing of GANs for Improved Quality, Stability, and Variation
Karras et al., ICLR'18



2014



2015



2016



2017



2018

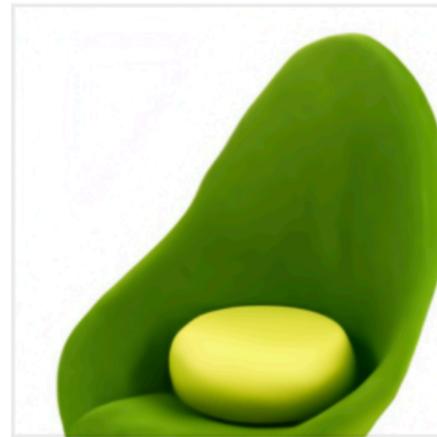
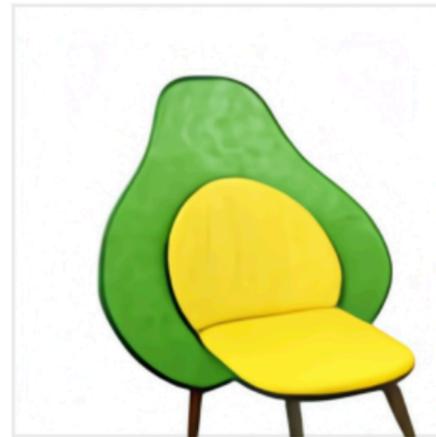




TEXT PROMPT

an armchair in the shape of an avocado. . . .

AI-GENERATED
IMAGES



- **Médecine: personnalisée, précision, automatiser les diagnostics**
- **Ingénierie: simulation de systèmes complexes, entretien préventif, automatiser la prise de décision**
- **Finance: mieux quantifier l'incertitude, analyser de grandes quantités de données, décision court-terme**
- **Marketing: pour comprendre et quantifier l'expérience utilisateur, améliorer l'efficacité des publicités**
- **Bien d'autres: changements climatiques, découverte de nouveaux médicaments change, créativité**
- **Votre domaine d'intérêt...**

Risques

- Technologie déjà impressionnante qui continuer à s'améliorer
- Dual-utilité (comme la majorité des technologies)
- Biais, coûts énergétiques élevés, factualité, évaluation compliquée



Le cours

Logistique

- **Plan de cours:** http://www.cs.toronto.edu/~lcharlin/courses/60629/index_fr.html
- Le lien est disponible à partir de ZoneCours.
- (Sinon par Google avec mon nom ou le sigle du cours)

Pédagogie inversée

- Chaque semaine:
 1. Préparation **avant le cours** (chez vous):
 - Matériel hebdomadaire (60–90 minutes)
 - Capsules et/ou lectures
 2. Temps en class (ensemble):
 - Sommaire, Questions-réponses, exercices (120 minutes)

Quelques conseils pour la pédagogie inversée

- **En classe: Soyez prêts**
 - **Regarder les capsules, lisez les références avant le cours**
 - **Venez avec des questions**
- **Capsules: Écoute active (p. ex., prenez des notes, n'hésitez pas à ré-écouter, bâtissez petit à petit un modèle mental de la matière)**

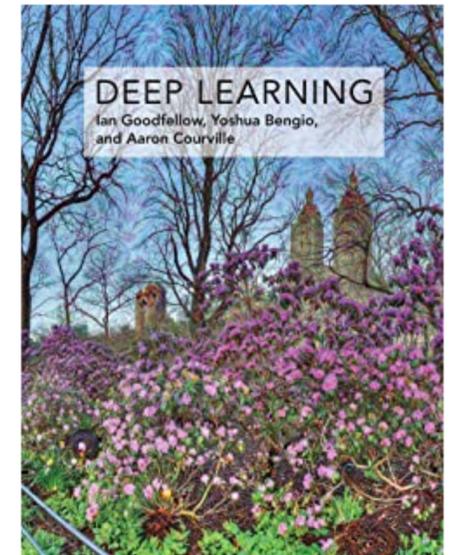
Le cours par rapport aux autres cours à Montréal

- HEC
 - Cours de doctorat à l'origine
 - Similaire au cours en *sciences des données* (MSc)
 - Requier une meilleure compréhension des principes computationnels
 - Requis pour
 - Machine Learning II: Deep Learning (MATH 60630A)
 - Trustworthy Machine Learning (MATH 80630)
- Autres cours de ML à Montréal (U.Montreal, Polytechnique, Mcgill)
 - Plus appliqué (similar to COMP-551@McGill)

**Bref retour sur certaines
notions essentielles en
algèbre linéaire et probabilité**

- Basé sur les chapitres 2 et 3 du livre “Deep Learning”

<http://www.deeplearningbook.org/>



Algèbre linéaire

- **Scalaire:** une valeur.

$$\mathbf{a} \in \mathbb{R}, \mathbf{a} \in \mathbb{N} \quad \mathbf{a} = 3$$

- **Vecteur:** un tableau à une dimension.

$$\mathbf{a} \in \mathbb{R}^D, \mathbf{a} \in \mathbb{N}^D \quad \mathbf{a} = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix}$$

- **Matrice:** un tableau à deux dimensions.

$$\mathbf{A} \in \mathbb{R}^{D_1 \times D_2}, \mathbf{A} \in \mathbb{N}^{D_1 \times D_2} \quad \mathbf{A} = \begin{bmatrix} 3 & 4 & 2 \\ 1 & 2 & 9 \end{bmatrix}$$

Indexation

- Indexer les éléments d'un vecteur: a_i

$$a = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix} \leftarrow a_1$$

Convention:
L'élément zéro est
le premier élément.

- Indexer les éléments d'une matrice: A_{ij}

$$A = \begin{bmatrix} 3 & 4 & 2 \\ 1 & 2 & 9 \end{bmatrix}$$

↑
 A_{12}

Opérations de base

- Transposé

$$\begin{array}{l} \mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \\ \mathbf{a}^\top = [a_0 \quad a_1 \quad a_2] \end{array} \quad \left| \quad (\mathbf{A}_{ij})^\top = \mathbf{A}_{ji} \right.$$

- Addition

- Vecteurs et matrices avec les mêmes dimensions

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \quad \mathbf{a} + \mathbf{b} = \begin{bmatrix} a_0 + b_0 \\ a_1 + b_1 \\ a_2 + b_2 \end{bmatrix} \quad \left| \quad (\mathbf{A} + \mathbf{B})_{ij} = \mathbf{A}_{ij} + \mathbf{B}_{ij} \right.$$

Opérations de base

- Multiplications par un scalaire

$$\alpha \mathbf{a} = \begin{bmatrix} \alpha \mathbf{a}_0 \\ \alpha \mathbf{a}_1 \\ \alpha \mathbf{a}_2 \end{bmatrix}$$

- Multiplication par un vecteur

- Produit scalaire

$$\mathbf{a}^\top \mathbf{a} = \sum_i \mathbf{a}_i \mathbf{a}_i$$

- Résultat est un scalaire.

- Multiplication par élément:

$$\mathbf{a} \odot \mathbf{a} = \begin{bmatrix} \mathbf{a}_0 \mathbf{a}_0 \\ \mathbf{a}_1 \mathbf{a}_1 \\ \mathbf{a}_2 \mathbf{a}_2 \end{bmatrix}$$

- Produit d'Hadamard

Opérations de base

- **Produit matriciel (produit scalaire):**

$$C_{ij} = \sum_k A_{ik} B_{kj}$$

- **Le nombre de colonnes de A doit être égal au nombre de rangées de B (non commutatif)**

$$\mathbf{A} \in \mathbb{R}^{D_1 \times D_2}, \mathbf{B} \in \mathbb{R}^{D_2 \times D_3}$$

- **Distribution: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$**
- **Association: $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$**
- **Produit des transposes: $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$**

Inverse

- Dénotons l'inverse d'une matrice par A^{-1}
- Une matrice est inversible si et seulement si:
 - Elle est carrée $D_1 = D_2$
 - Ses colonnes sont linéairement indépendantes
 - Aucune colonne n'est la combinaison d'autres colonnes.
- Les inverses sont notamment utiles pour résoudre des systèmes d'équations:

Une matrice carrée,
non inversible est appelée
singulière

$$Ax = b \quad x = A^{-1}b$$

Normes

- Norme L^p . Taille d'un vecteur (ou matrice)

$$\| \mathbf{a} \|_p = \left(\sum_i |\mathbf{a}_i|^p \right)^{1/p}$$

- Quelques normes souvent utilisées en ML:

- Norme Euclidienne ($p=2$)

$$\| \mathbf{a} \|_2 = \sqrt{\left(\sum_i |\mathbf{a}_i|^2 \right)}$$

- Produit scalaire:

$$\mathbf{a}^\top \mathbf{b} = \| \mathbf{a} \|_2 \| \mathbf{b} \|_2 \cos \theta_{\mathbf{a}\mathbf{b}}$$

- Norme de Frobenius (matrice):

$$\| \mathbf{A} \|_2 = \sqrt{\left(\sum_i \sum_j |\mathbf{a}_{ij}|^2 \right)}$$

Matrices spéciales & Vecteurs spéciaux

- Identité. Dénotée I_n .
 - Des zéros partout sauf sur la diagonale
- Symétrique: $A = A^T$
- Vecteur unitaire: $\|a\|_2 = 1$
- Vecteurs orthogonaux: $a^T b = 0$
- Vecteurs orthonormaux: unitaires & orthogonaux
- Matrice orthonormale: rangées et colonnes orthonormales

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A^T A = A A^T = I$$

- **Sauter décomposition en vecteurs propres, SVD, pseudo-Inverse, déterminants (Sections 2.7–2.11).**

- **Passons aux probabilités**
- **Chapitre 3 du livre “Deep Learning”**
 - **J’ai adapté certaines diapos venant du livre**
 - **Merci à Ian Goodfellow pour ces diapos**

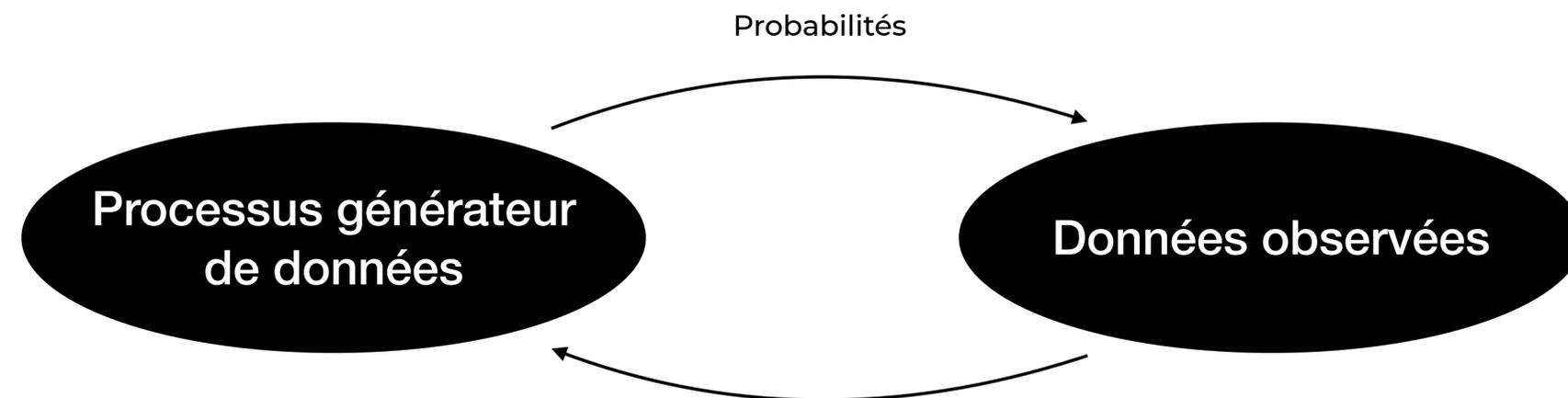
Les probabilités pour quoi faire?

- Pour modéliser l'incertitude

P. ex., à quelle heure vais-je rentrer ce soir?

- Les probabilités offrent un formalisme pour comprendre le processus qui “génère les données” (L. Wasserman)

P. ex. lancer un dé à six côtés



L'exemple du

- Génère des données en lançant un dé.
- Que peut-on dire sur un lancer?
 - 6 résultats possibles. (**univers**)
 - Chaque résultat (p. ex. 1). (**issue, évènement élémentaire**)
 - Les sous-ensembles des évènements (p. ex., <3). (**évènement**)
 - Les évènements sont équiprobables. (**distribution uniforme**)

Variabliques aléatoires

- Une variable aléatoire (RV) est "une variable dont la valeur est déterminée après un tirage aléatoire"
- Par exemple:
 - Lancer de dé (X)
 - Un lancer $\in \{1, 2, 3, 4, 5, 6\}$. (x)
- Une fonction de probabilité (P) assigne une valeur réelle à chaque issue: $P(x) \geq 0, \forall x \in X$

$$P\left(\bigcup x\right) = 1$$

Variabes aléatoires discrètes

- Une RV est discrète si elle peut prendre un nombre fini de valeurs¹ $P(x = x_i) \geq 0, \forall i$

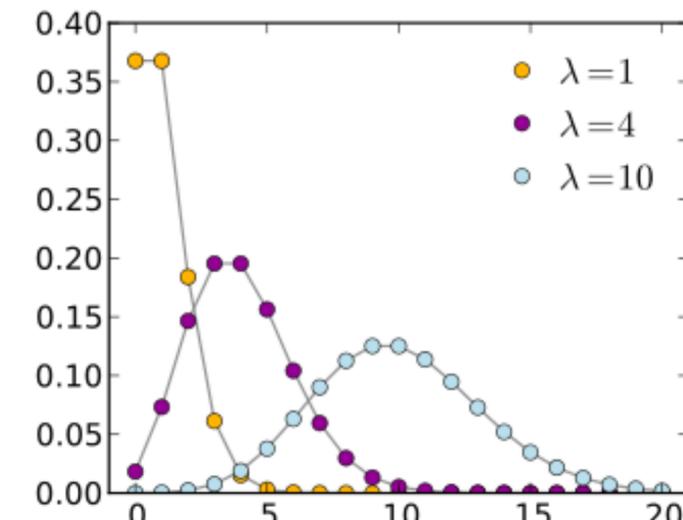
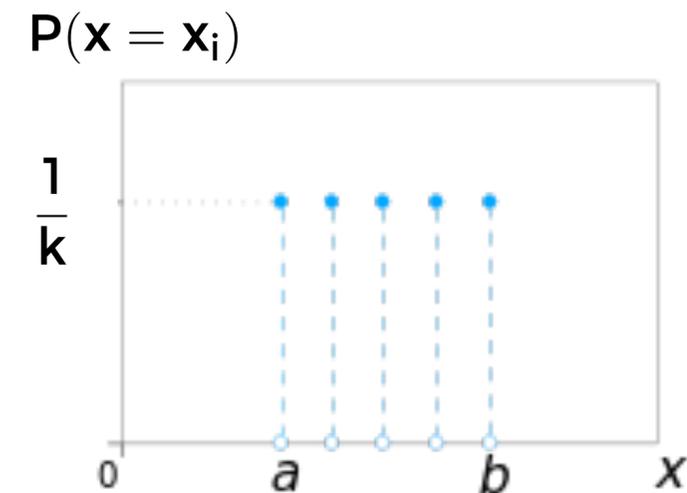
$$\sum_i P(x = x_i) = 1$$

- P. ex., distribution uniforme:

$$P(x = x_i) = \frac{1}{k}, \forall i$$

- P. ex., distribution de Poisson:

$$P(x = x_i; \lambda) = \frac{\lambda^{x_i} \exp^{-\lambda}}{x_i!}$$



Les images sont de:
wikipedia.org

RV continues

- Une RV est continue ssi: $f(x) \geq 0, \forall x \in X$

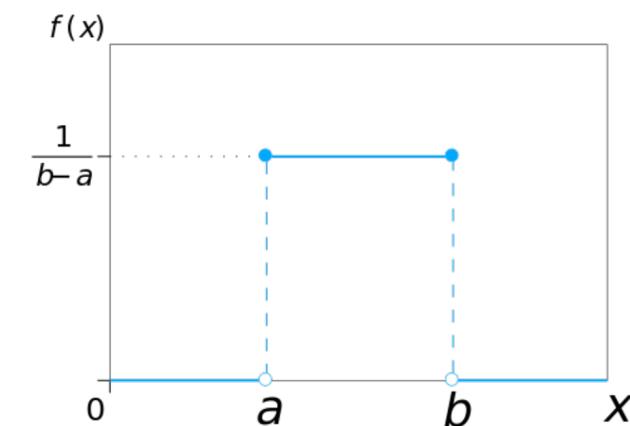
$$\int_{-\infty}^{\infty} f(x) dx = 1$$

$$P(a < x < b) = \int_a^b f(x) dx$$

- $f(x)$ est une fonction de densité (PDF)
- P. ex., distribution uniforme (continue):

$$u(x; a, b) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

- P. ex., Distribution normale



L'image est de:
wikipedia.org

Quelques propriétés utiles

(Démontrées avec des variables discrètes pour simplifier les choses)

- **Somme:** $P(X) = \sum_Y P(X, Y)$
- **Produit:** $P(X, Y) = P(X | Y)P(Y)$
- **Décomposition:** $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | X_1, \dots, X_{i-1})P(X_1)$
- **Si X et Y sont indépendants:** $P(X, Y) = P(X)P(Y)$
- **Loi de Bayes:** $P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)}$

Moments

- **Espérance:** $\mathbb{E}[\mathbf{X}] = \sum_i \mathbf{P}(\mathbf{x} = \mathbf{x}_i) \mathbf{x}_i$ $\mathbb{E}[\mathbf{aX}] = \mathbf{a}\mathbb{E}[\mathbf{X}]$
- **Variance:** $\sigma^2 = \mathbf{E}[(\mathbf{X} - \mathbf{E}[\mathbf{X}])^2]$
- **Covariance:** $\mathbf{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{E}[(\mathbf{X} - \mathbf{E}[\mathbf{X}])(\mathbf{Y} - \mathbf{E}[\mathbf{Y}])]$
- **corrélation:** $\rho(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{Cov}(\mathbf{X}, \mathbf{Y})}{\sigma_x \sigma_y}$