

L'APPRENTISSAGE AUTOMATISÉ: CONNEXIONS AVEC NOS CONTENUS!

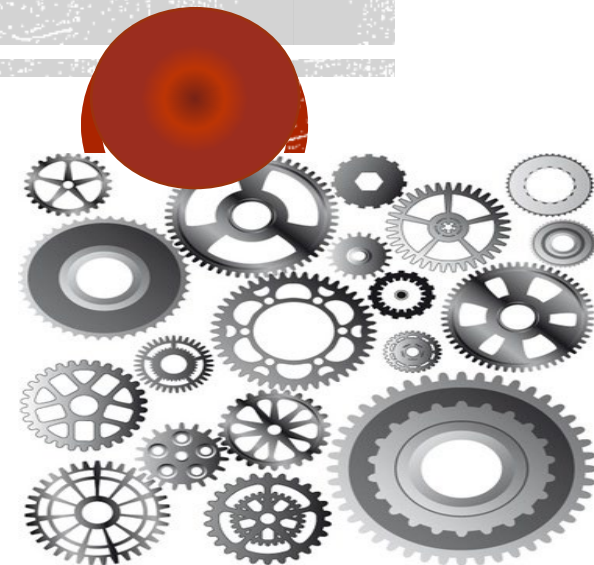
MACHINE LEARNING : CONNECTIONS WITH OUR CURRICULUM!

Brève introduction pour non-spécialistes

Brief introduction for non-specialists

Julie Plante - Vanier College

Colloque de l'AQPC Juin 2022



EN PREMIER LIEU : UN GROS MERCI...

- Je suis financée par un fonds de recherche NovaScience Volet 2b pour disséminer les possibilités de l'IA dans l'éducation collégiale, en collaboration avec Prof. T.H. Kwok du département de Génie Mécanique de l'Université Concordia.
- Je fus accompagnée par Ron Friedman, étudiant en Informatique à Vanier, grâce à une bourse MITACS FRQNT 20-21 dans la mise en place du site web vaniercollegeai.ca .

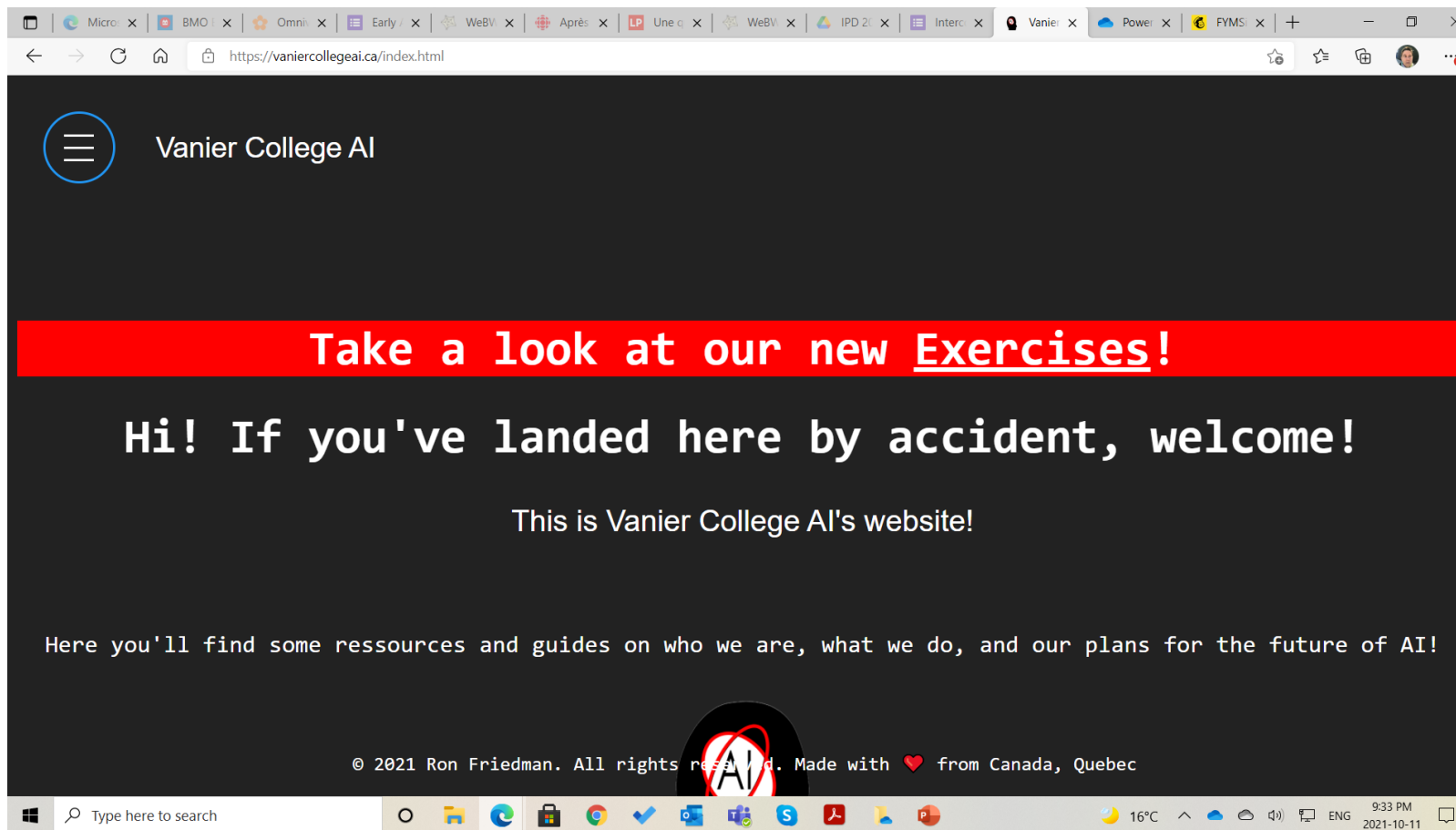
Québec 

VANIER
CÉGEP / COLLEGE

Mitacs



BEAUCOUP DE RESSOURCES SUR WWW.VANIERCOLLEGEAI.CA



The screenshot shows a web browser window with the URL <https://vaniercollegeai.ca/index.html>. The page has a dark background with a red banner at the top that reads "Take a look at our new Exercises!". Below the banner, the text says "Hi! If you've landed here by accident, welcome!" and "This is Vanier College AI's website!". At the bottom of the page, it says "Here you'll find some ressources and guides on who we are, what we do, and our plans for the future of AI!". The footer contains the copyright notice "© 2021 Ron Friedman. All rights reserved. Made with ❤️ from Canada, Quebec". The browser's taskbar is visible at the bottom, showing various application icons and the system tray with the date and time "9:33 PM 2021-10-11".



UNE VISITE AU VANIERCOLLEGEAI.CA

The screenshot shows a web browser window displaying the navigation menu of the Vanier College AI website. The menu items are listed on the left side of a dark background. Three orange arrows point from text annotations on the right to specific menu items:

- An arrow points from the text "Intro à la programmation et Python." to the "Documentation" menu item.
- An arrow points from the text "Ressources en IA" to the "ML Course — PPT Slides" menu item.
- An arrow points from the text "Exercices en Python." to the "Exercices Folder" menu item.

The browser's address bar shows the URL <https://vaniercollegeai.ca/index.html>. The Windows taskbar at the bottom shows the time as 11:25 AM on 2021-10-18.



VOYEZ LES EXERCICES ÉTAPE-PAR-ÉTAPE.

The screenshot shows a web browser window displaying a SharePoint OneDrive page. The browser's address bar shows the URL: https://eduvaniercollegeqc-my.sharepoint.com/personal/plantej_vaniercollege_qc_ca/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fplantej_vaniercollege_qc_ca%2F.... The page header includes the Vanier Cégep/College logo and 'Office 365'. Below the header, there is a 'Download' button and a 'Sort' dropdown menu. The main content area shows a breadcrumb path: 'Julie Plante > AI research OD > Ron's work > Folder of Exercises for Students'. A table lists the contents of the folder:

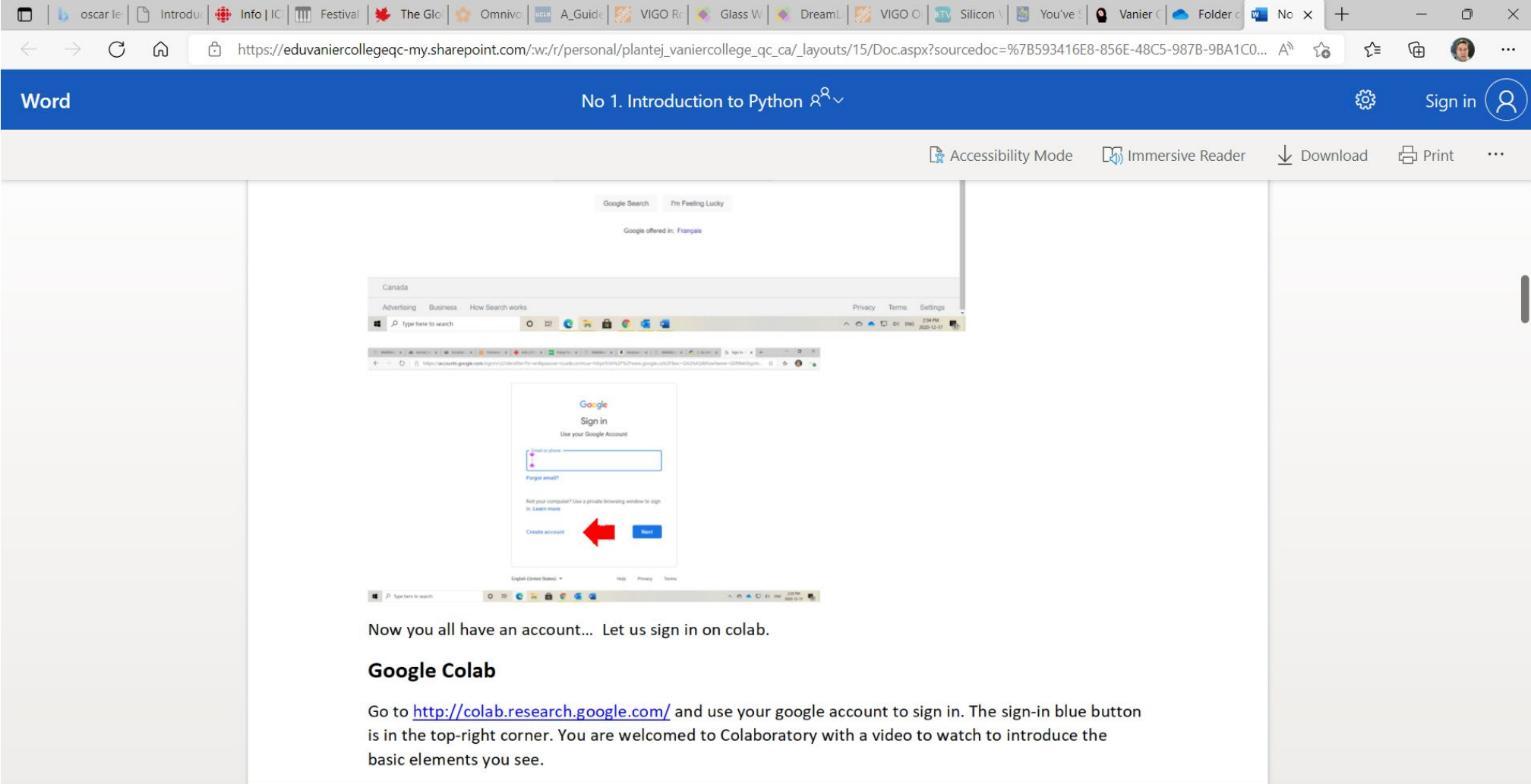
Name	Modified	Modified By	File size	Sharing
Exercices for Colleges	May 14, 2021	Julie Plante	13 items	Shared
Exercices for High Schools	May 14, 2021	Julie Plante	6 items	Shared
No 1. Introduction to Python.docx	March 12, 2021	Julie Plante	689 KB	Shared
No 2. Exercise 1.docx	March 12, 2021	Julie Plante	15.0 KB	Shared
Useful Links.docx	May 25, 2021	Julie Plante	13.0 KB	Shared

A red arrow points to the file 'No 1. Introduction to Python.docx', which is highlighted in the table. A tooltip for this file is visible, showing the name 'No 1. Introduction to Python.docx'.

The Windows taskbar at the bottom shows various application icons, including the Start menu, search, weather, and several office applications. The system tray on the right indicates the language is 'ENG CMS', the time is '11:32 AM', and the date is '2022-06-01'.



DANS LE DOSSIER *EXERCISES FOLDER*, ÉTAPE PAR ÉTAPE



Word No 1. Introduction to Python

Accessibility Mode Immersive Reader Download Print

Google Search I'm Feeling Lucky

Google offered in: Français

Canada

Advertising Business How Search works Privacy Terms Settings

Type here to search

Sign in

Use your Google Account

Forgot email?

Not your computer? Use a private browsing window to sign in. [Learn more](#)

Create account Sign in

Now you all have an account... Let us sign in on colab.

Google Colab

Go to <http://colab.research.google.com/> and use your google account to sign in. The sign-in blue button is in the top-right corner. You are welcomed to Colaboratory with a video to watch to introduce the basic elements you see.

Colloque de l'AQPC Juin 2022

100% Give Feedback to Microsoft

ENG CMS 10:20 PM 2022-05-31 24



JUSQU'ÀUX PREMIÈRES LIGNES DE CODE AVEC COLAB.

is in the top-right corner. you are welcomed to Colaboratory with a video to watch to introduce the basic elements you see.

62%
See 66 Editor suggestions

Waiting for colabresearch.google.com...

What is Colaboratory?

Colaboratory, or "Colab" for short, allows you to write and execute Python in your browser, with

- Zero configuration required
- Free access to GPUs
- Easy sharing

Whether you're a **student**, a **data scientist** or an **AI researcher**, Colab can make your work easier. Watch [Introduction to Colab](#) to learn more, or just get started below!

Getting started

The document you are reading is not a static web page, but an interactive environment called a **Colab notebook** that lets you write and execute code.

For example, here is a **code cell** with a short Python script that computes a value, stores it in a variable, and prints the result:

You should read the following overview of Colab please : [Colab basic features overview](#)

Page 1 of 8 1,422 words English (U.S.) Text Predictions: On Editor Suggestions: Showing 100% Fit Give Feedback to Microsoft





MAIS C'EST QUOI L' I. A. ?



VOCABULAIRE : INTELLIGENCE ARTIFICIELLE

- I. A.
- Un système quelconque qui répond à un signal.
- Une simple horloge-réveil ou votre cafetière programmable.
- Jusqu'à un robot autonome d'assemblage sur une chaîne de montage.
- Ou bien une belle montre "smartwatch" ...
- Et bien sûr... votre téléphone cellulaire!



ET PARFOIS MÊME DANS LE FROMAGE...!

Fat est
LE PREMIER ROBOT D'AFFINAGE
INSTALLÉ EN AMÉRIQUE
et provient de
LA VILLE DE ROMONT
EN SUISSE

Merci à la Fromagerie du Presbytère de Sainte-Élisabeth de Warwick!



VOCABULAIRE : INTELLIGENCE ARTIFICIELLE

De Wikipedia en Français on trouve:

L'intelligence artificielle (IA) est « l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine ». Encyclopédie Larousse.

Elle englobe donc un ensemble de concepts et de technologies, plus qu'une discipline autonome constituée. Des instances, telle la CNIL, notant le peu de précision de la définition de l'IA, l'ont présentée comme « le grand mythe de notre temps ».



VOCABULAIRE : INTELLIGENCE ARTIFICIELLE

De Wikipedia en Anglais on trouve:

- It is intelligence demonstrated by machines, unlike the **natural intelligence** displayed by humans and animals. Leading AI textbooks define the field as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals. From Poole, Mackworth & Goebel 1998,



VOCABULAIRE : INTELLIGENCE ARTIFICIELLE

- Certains la définissent comme “N’importe quoi qui n’a pas encore été fait...”
- Beaucoup de différents groupes sous le thème IA.
- Pour nous, on regarde l’Apprentissage Machine ou bien “Machine Learning”.

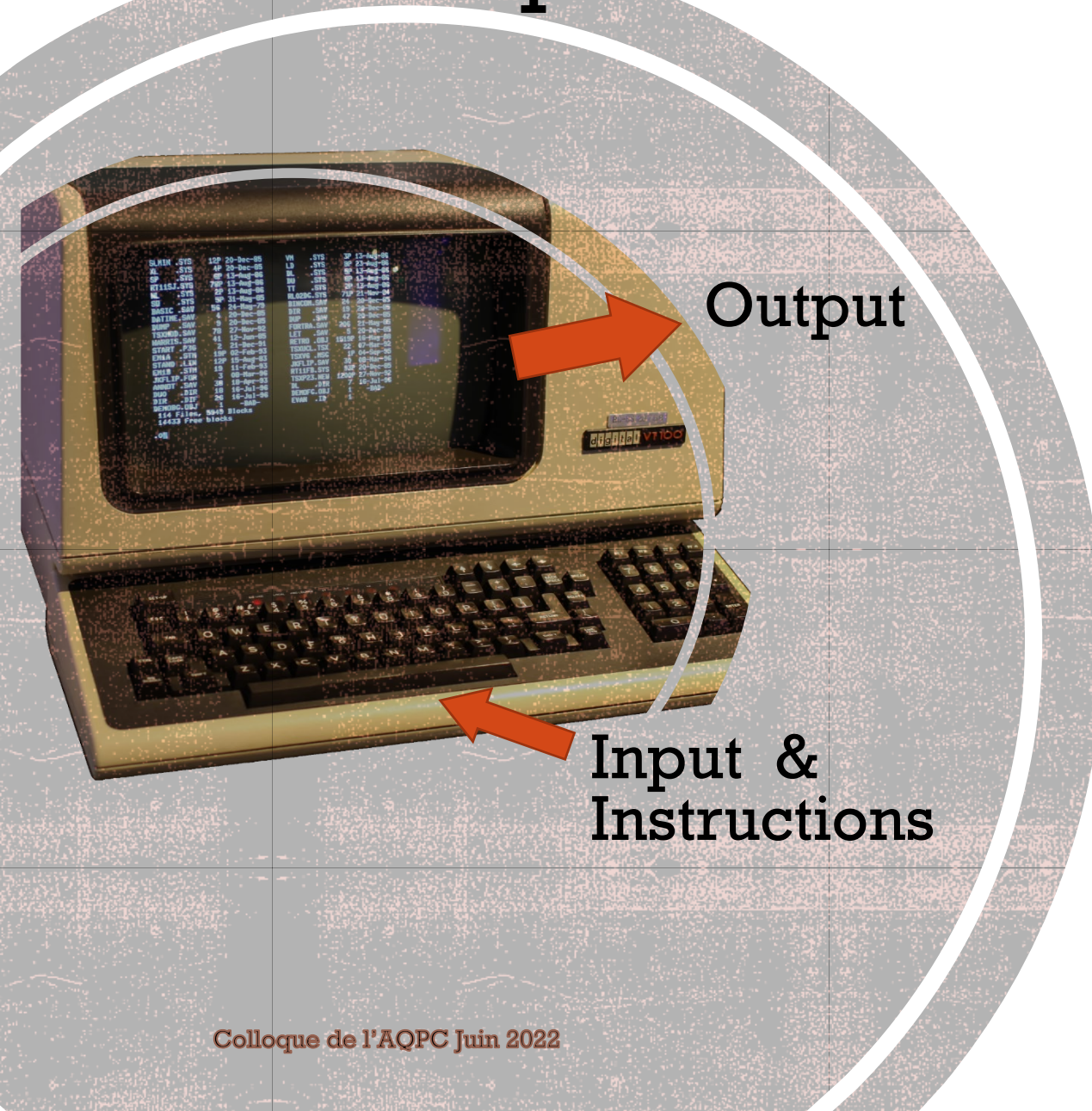


VOCABULAIRE : MACHINE LEARNING

- **Machine learning (ML)** est l'étude d'algorithmes qui s'améliorent en intégrant des données basées sur le passé, sur des expériences, de l'entraînement.
- Nous regarderons le ML supervisé pour cette présentation.
 - Ensemble de paires input (données) – output (action désirée).
- Notons ici la différence entre un algorithme « classique » et le ML .



Le Classique...



Le ML Supervisé

Paramètres
du
modèle ou
Instructions



POINT CLÉ À RETENIR!

- **Sujet : L'Apprentissage Automatisée**
- **Informatique: Programmation / Résolution de problème**
- **L'approche AA : Input + Output \rightarrow Modèle vs
Le Classique : Input + Instructions \rightarrow Output**



MACHINE LEARNING EN BREF...

- On a un **modèle**. Très souvent (ici, pour aujourd'hui) linéaire.
- On a les mesures des caractéristiques, les $x_0, x_1 \dots x_k$ en colonne.
- On a les **paramètres**, $w_0, w_1 \dots w_k$ aussi en colonne.

- La prédiction sera :

$$\text{prédiction} = W^T X$$

- Mais la **magie** est dans l'obtention des paramètres (weigths)...!



PARENTHÈSE : LE ML NON-SUPERVISÉ

- Ceci ne fait pas partie de notre discussion mais... il est agréable d'en savoir un peu plus dans la vie...
- Le ML non-supervisé prend des « input data » et trouve des caractéristiques, « features », par lui-même. Ces algorithmes groupent les « input data » en sous-groupes, « clusters ».
- Exemple en analyse d'images : après plusieurs images, l'algorithme peut décider d'identifier une ou deux caractéristiques importantes telles que les côtés droits, les coins... Les pixels sombres, les pixels clairs...



LE ML NON-SUPERVISÉ – SUITE ET FIN

- Comment fait-on pour identifier les caractéristiques d'une image?
- La convolution par un noyau, « kernel »
- C'est une multiplication matricielle...ou presque! Le noyau est choisi pour trouver ... quelque chose!

La réponse à
« Pourquoi on apprend
les matrices » ?

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

- Ce noyau, multiplié à chaque pixel, et les 8 autres pixels autour, de votre image, trouvera les contours.



POINT CLÉ À RETENIR!

- **Sujet : Algèbre Linéaire, opérations matricielles.**
- **La Mathématique : Convolution par un noyau pour trouver des caractéristiques dans une image.**
- **Un lien très pertinent : [Convolution PDF](#)**



POINT CLÉ À RETENIR!

- **Sujet : Opérations matricielles, traitement de l'image.**
- **Informatique : Convolution par un noyau pour trouver des caractéristiques dans une image. Appel à des fonctions de librairies.**
- **Un lien très pertinent : [Convolution PDF](#)**



VOIR LE CODE JUPYTER CONVOLUTION EX.IPYNB



Au : [vaniercollegeai.ca /](http://vaniercollegeai.ca/)
ML Course - PPT slides



L'OBJECTIF DU ML

- On utilise des “training data” pour optimiser les paramètres (coefficients) d’un modèle proposé.
- Les paramètres et le modèle sont alors utilisés pour prédire des output en utilisant des données de test.
- Les « testing data » sont un sous-groupe de nos données.
- Si satisfaisant, le “modèle” sera utilisé avec de nouvelles données « en service »...!



DEUX PROBLÈMES RÉSOLUS PAR LA M.L.

1. La régression (linéaire – suffit pour aujourd'hui) :

- Trouver les meilleurs paramètres basés sur des données existantes (x_i, y_i) pour le modèle : $y = w^T x + b$
- Où en fait W, X sont des vecteurs-colonnes. On note plutôt:
- $y = W^T X$, le vecteur X aura en dernier élément 1, le b sera le dernier élément de W .
- Attention : la régression est un outil très puissant quand on a des centaines de facteurs ou mesures x_i^j à chaque point (x_i, y_i), ainsi donc des centaines de paramètres dans le vecteur W .
- Nous ferons un exemple : prédire un temps de victoire au 100 mètres olympique basé sur les précédents temps de victoire.



DEUXIÈME PROBLÈME RÉSOLU PAR LA ML

2. La catégorisation (oui ou non):

- Trouver les paramètres basés sur des données existantes (x_i, y_i) pour répondre à la question « est-ce que les données x_i appartiennent à un élément de la catégorie y oui ou non? »
- Exemples typiques : filtres de nouvelles (Sports? Internationales? Économiques?), filtres de photos (De chats? De mon ami Robert? De moi!?)
- Nous ferons un exemple de base « théorique » : un point à trois facteurs est-il oui ou non représentatif d'un étudiant en difficulté?



MAIS LE PLUS IMPORTANT : COMMENT LE ML “APPREND” ???

- Voici la partie probabiliste.
- Prêts à faire un plongeon en statistiques?
Bayésiennes?

- On part!



VOICI UN EXEMPLE SIMPLE POUR NOUS GUIDER.

- Nous pourrions introduire quelques mots de vocabulaire.



ON VEUT DES PREDICTIONS BASÉES SUR DES DONNÉES HISTORIQUES.

- Un modèle de prédiction =
Fonction (X , Paramètres, Données du passé)
- On établit des paramètres A Priori pour commencer.
- On modifie les paramètres pour maximiser la chance de prédire les résultats passés à partir des données du passé.
- Les paramètres maximaux sont les paramètres A Posteriori



TRACER LA MEILLEURE LIGNE DROITE

- Tracer la meilleure ligne droite parmi des points expérimentaux (x,y) .
- N points, on a donc $x_1 \dots x_n$, $y_1 \dots y_n$.
- On a une seule mesure (caractéristique) à chaque point: x_i .
- Le modèle est... linéaire par notre choix: $y = w_0 + w_1 x$
-
- w_i pour... “Weights”. Les coefficients forment le modèle.
- Question: Un “bon modèle”, ca veut dire quoi?
Comment mesure-t-on ça?



ON PARLE DE L'ÉCART, "LOSS FUNCTION"

- La fonction d'écart, de perte.
- La différence entre les output désirés et les prédictions.
- Dans un modèle linéaire, on utilise :
 - $\sum (y_n - \text{prédiction}_n)^2$
 - $\sum (y_n - [w_0 + w_1 x_n])^2$

Les étudiants ont fait cela au secondaire!



LES PARAMÈTRES (A.K.A. WEIGHTS)

- Avec l'aide du Calcul différentiel, on trouve le minimum de la perte, loss function, nous donnant la bonne vieille recette:

- $\hat{x} = \frac{1}{N} \sum_{n=1}^N x_n$ la Moyenne de tous les xs,

- $\hat{y} = \frac{1}{N} \sum_{n=1}^N y_n$ la Moyenne de tous les ys.

- Paramètre $w_1 = \frac{\frac{1}{N}(\sum_{n=1}^N x_n y_n) - \hat{x} \hat{y}}{\frac{1}{N}(\sum_{n=1}^N x_n^2) - \hat{x} \hat{x}}$

- Paramètre $w_0 = \hat{y} - w_1 \hat{x}$



Voilà Pourquoi la dérivée!



POINT CLÉ À RETENIR!

- **Sujet : Calcul Différentiel**
- **La Mathématique : La dérivée pour redécouvrir les formules de statistique pour la meilleure droite alignée sur des points donnés.**
- **Un lien très pertinent : tout bon texte couvrant la régression linéaire.**



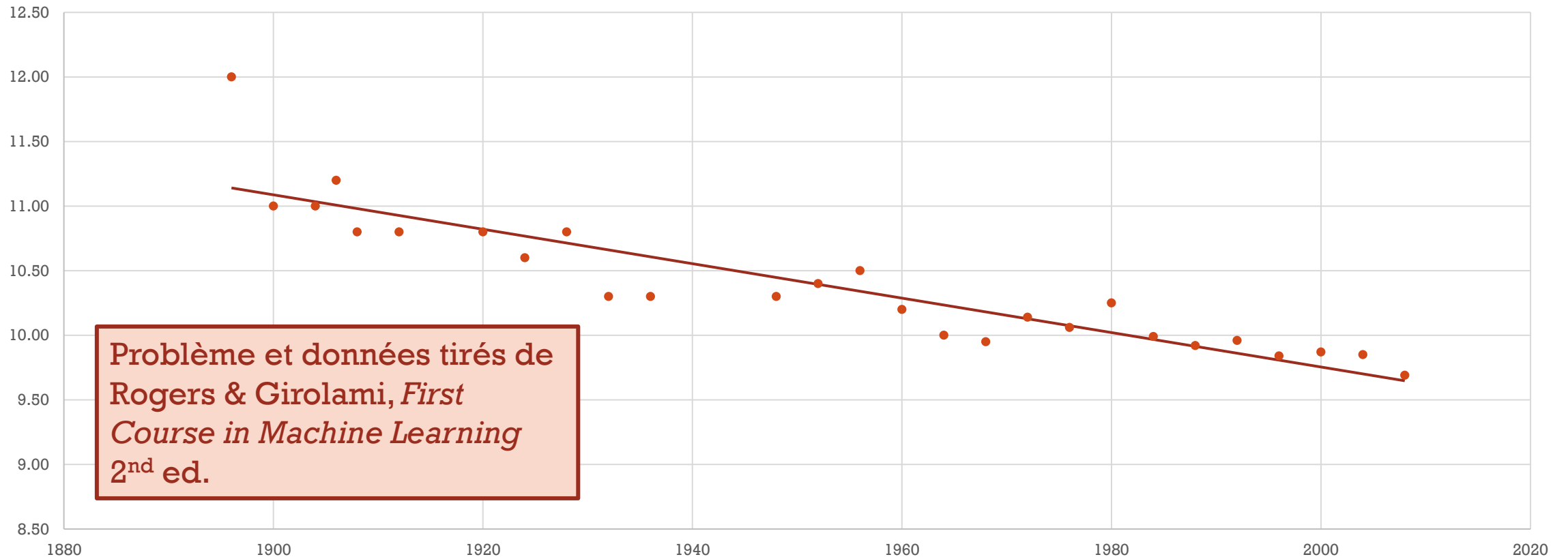
POINT CLÉ À RETENIR!

- **Sujet : Lecture de fichiers.**
- **Informatique : lecture des données dans un fichier (format .json, .csv, etc...).**
- **Un lien pertinent : L'API de votre langage préféré.**



DONNÉES VERSUS PRÉDICTIONS DU TEMPS GAGNANT AU 100 MÈTRES VIA LES STATS.

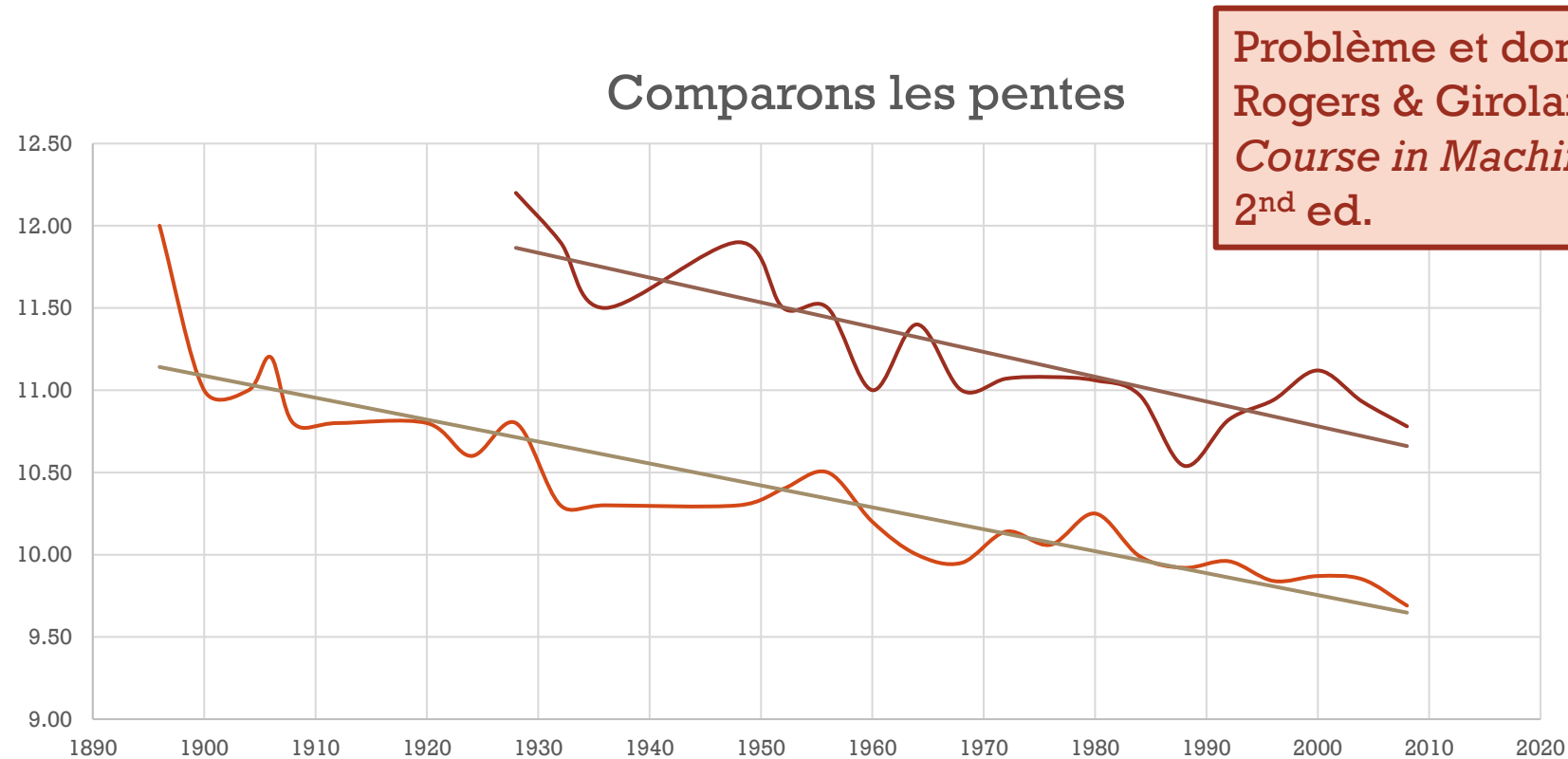
Men's winning time



Problème et données tirés de
Rogers & Girolami, *First
Course in Machine Learning*
2nd ed.



PRÉDICTIONS: LES FEMMES BATTRONT-ELLES LES HOMMES?



DES LIMITATIONS AUX STATISTIQUES

- Nous avons UNE prédiction, un paramètre sans précision.
- Nous avons UNE valeur, sans probabilité d'être proche ou éloigné de la bonne valeur.
- On aimerait un intervalle avec une probabilité de couvrir la véritable valeur dans cet intervalle.

Mais
comment?



INCLUONS DES PROBABILITÉS DANS LE MODÈLE...

- On ajoute un bruit aléatoire:

-

Noise = $\mathcal{N}(0, 0.5)$, une distribution Gaussienne centrée à 0, et variance = 0.5 (au choix...).

- Temps = Prédiction (year) = $w1 * year + w0 + \mathcal{N}(0, 0.5)$



IL FAUT COMMENCER : LES A PRIORIS...

- On recentre les coordonnées x (l'année) vers 0, 1, 2... (on nomme cela une Standardisation)
- On estime le b à 12.0 (le premier temps...! year = 0)
- On estime le m à -1.0, la pente entre les deux premiers points.
- On utilise une Gaussienne 2D centrée à (12, -1) :
- $\text{Apriori} = [12.0, -1.0]$ Ceci sera nos premiers paramètres.
- On estime une plus grande variance autour de b , m sera un petit nombre négatif.
- On choisit $\text{AprioriVar} = \begin{bmatrix} 100 & 0 \\ 0 & 5 \end{bmatrix}$; sans covariance entre w_0 et w_1 .



QUE VOULONS-NOUS ? QU'AVONS-NOUS ?

- On veut une **distribution** de probabilité des **paramètres**, sachant les points donnés et des paramètres initiaux.
- $\text{Prob}(W \mid \text{Temps, années})$
- On nomme ceci l'A Posteriori.

- Nous avons nos A Priori, $\text{Prob}(W)$, gaussienne.

- On peut calculer $\text{Prob}(\text{Temps} \mid W, \text{années})$.
- La barre le dit : « sachant W » ...!



PROB (T | W) : SON CALCUL

- Si on connaît le W, c'est facile...! C'est une gaussienne:
 - Temps = $w_1 * \text{year} + w_0 + \mathcal{N}(0, 0.5)$
 - C'est donc une Gaussienne $\mathcal{N}(w_1 * y + w_0, 0.5)$
 - On ne fait qu'ajouter une constante $w_1 * y + w_0$ à une gaussienne.
-
- On nomme ceci la vraisemblance.
 - En Anglais : Likelihood.



ARRIVE BAYES ET LA VISION BAYÉSIENNE...

Théorème de Bayes :

$$\text{Probabilité (A | B)} = \text{Prob (B | A)} * \text{Prob(A)} / \text{Prob(B)}$$

Difficile d'expliquer rapidement... Voici un exemple:

Il y a un test qui dit positif ou négatif, mais il y a la vérité: avoir ou pas le cancer.

Il y a une différence entre la tumeur et le test.

A test screening 80% of cancer patients!
This cancer affects 1% of the population.

	Cancer (1%)	No Cancer (99%)
Test Pos	80%	0.8%
Test Neg	20%	0.4%



DEUX DONNÉES NÉCESSAIRES...

- La sensibilité d'un test: nombre de cancéreux positifs / nombre de cancéreux.
- Exemple : 80%

- La spécificité d'un test: nombre de gens sans cancers et négatif / nombre de gens sans cancer.
- Exemple : 90.4%



EXEMPLE INTUITIF DE BAYES

A: avoir le cancer.

B: test positif.

	Cancer (1%)	No Cancer (99%)
Test Pos	80%	9.6%
Test Neg	20%	90.4%

La question : Prob (Avoir le cancer | test positif) ???

$P(A | B) = \text{Prob} (\text{positif} | \text{ayant le cancer}) * \text{Prob}(\text{cancer}) / \text{Prob}(\text{test positif})$

1000 personnes	Ont le cancer (10)	N'ont pas le cancer (990)
Test positif	8	990-895 = 95
Test négatif	10-8 = 2	990 * 90.4% = 895



ARRIVE BAYES ET LA VISION BAYÉSIENNE...

Théorème de Bayes :

$$\text{Probabilité (A | B)} = \text{Prob (B | A)} * \text{Prob(A)} / \text{Prob(B)}$$

A: avoir le cancer.

B: test positif.

1000 personnes	Ont le cancer (10)	N'ont pas le cancer (990)
Test positif	8	95
Test négatif	2	895

$$P(A | B) = 80\% * 1\% / (80\%*1\%+9.6\%*99\%)$$

$$P(A | B) = 0.8 * 0.01 / (0.10304) = 7.76\% \quad \text{Donc on se calme!}$$

	Cancer (1%)	No Cancer (99%)
Test Pos	80%	9.6%
Test Neg	20%	90.4%



POINT CLÉ À RETENIR!

- **Sujet : Statistiques et Probabilités.**
- **La Mathématique : Le théorème de Bayes appliqué à l'AA pour « apprendre ».**
- **Un lien très pertinent : Un bon texte de probabilités couvrant le Théorème de Bayes.**



JOUONS AVEC BAYES ET LES PROBABILITÉS...

- Par le théorème de Bayes :

Prob (Paramètres | Temps, données passées) =

$$\frac{\text{Prob (Temps | Param., données)} \cdot \text{Prob (Param. | hypothèse initiales)}}{\text{Prob (Temps sur l'entièreté des params)}}$$

ou

- $$\text{Prob (W | T, Y)} = \frac{\text{Prob (T | W, Y)} \cdot \text{Prob (W | hypothèse initiales)}}{\text{Prob (T) sur l'entièreté des W}}$$



UN MIRACLE APPELÉ LES DISTRIBUTIONS CONJUGUÉES (RÉF WIKIPEDIA CI-BAS)

- Deux distributions qui sont dans la même famille de distributions conjuguées générerons une distribution de la même famille.

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters ^[note 1]	Interpretation of hyperparameters	Posterior predictive ^[note 5]
Normal with known variance σ^2	μ (mean)	Normal	μ_0, σ_0^2	$\frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n x_i}{\sigma^2} \right), \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1}$	mean was estimated from observations with total precision (sum of all individual precisions) $1/\sigma_0^2$ and with sample mean μ_0	$\mathcal{N}(\tilde{x} \mu'_0, \sigma_0'^2 + \sigma^2)^{[4]}$

- https://en.wikipedia.org/wiki/Conjugate_prior#Table_of_conjugate_distributions



UN MIRACLE APPELÉ DES DISTRIBUTIONS CONJUGUÉES... ET BEAUCOUP DE CALCUL DE PROB!

- On trouve l'expression exacte de la distribution de nos deux paramètres. Un vecteur de dimension 1 X 2.
- Une distribution gaussienne avec :
- $Moyenne = PosterioriVar_w * \left(\frac{1}{\sigma^2} X^T t + APrioriVar^{-1} \cdot APrioriMean \right)$
- $PosterioriVar_w = \left(\frac{1}{\sigma^2} X^T X + APrioriVar^{-1} \right)^{-1}$
- Voir le code en exemple pour le 100 mètres.
- Bon exemple de travail matriciel!



LES RÉSULTATS

- La manière traditionnelle, vive les statistiques:
- $m = -0.06008547008547001$ et $b = 11.17074074074074$

-
- Avec l'Intelligence Artificielle, les a priori se corrigent.
 - Valeurs de nos a priori: [12. -1.]
 - Et la variance... $\begin{bmatrix} 100. & 0. \\ 0. & 5. \end{bmatrix}$
 - Résultats avec AI:
 - Le $m = -0.06071254360244203$ et $b : 11.181266535991748$
 - **Variance** = $\begin{bmatrix} 0.07005356 & -0.00396523 \\ -0.00396523 & 0.00030507 \end{bmatrix}$

C'est quoi la difference ?
Pourquoi tout le travail
de probabilité?



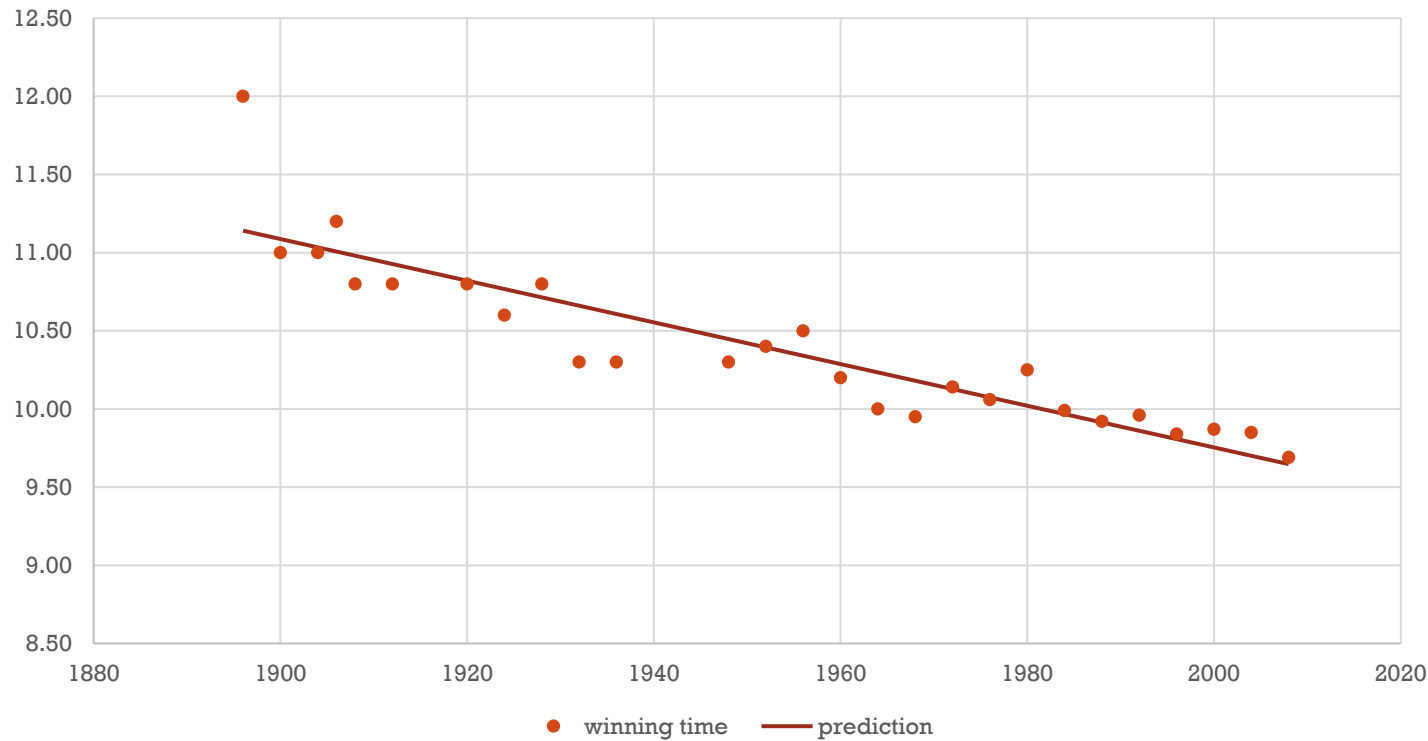
CE N'EST PAS JUSTE DEUX NOMBRES, M ET B.

- C'est une distribution de valeurs!
- Et on peut prédire plus qu'un nombre, c'est une probabilité autour de ce nombre.
- On peut dire qu'en 2012, on trouvera un temps gagnant **moyen** de : $27 * -0.06071254360244203 + 11.181266535991748 = 9.54s$
- Ceci est une gaussienne.
- Moyenne de 9.54 avec une **variance de 0.57**.
- **Covariance** de nos paramètres: pas complètement indépendants!



LES COVARIANCES...

Men's winning time



$$\text{Var} = \begin{bmatrix} 0.070054 & -0.003965 \\ -0.003965 & 0.000305 \end{bmatrix}$$

On voit que si b augmente, la pente diminue dans le négatif.

Et vice-versa!



APRÈS LA RÉGRESSION, LA CATÉGORISATION.

- L'autre grand problème en IA est la catégorisation entre plusieurs catégories.
- Est-ce un chat ou un chien sur cette image?
- Est-ce que cet étudiant est à risque ou non?
- La question demande une réponse binaire, oui ou non, 0 ou 1.
- Ici encore, l'algèbre linéaire et le truc des paramètres fonctionnent.
- Mais on a besoin de la fonction sigmoïde.



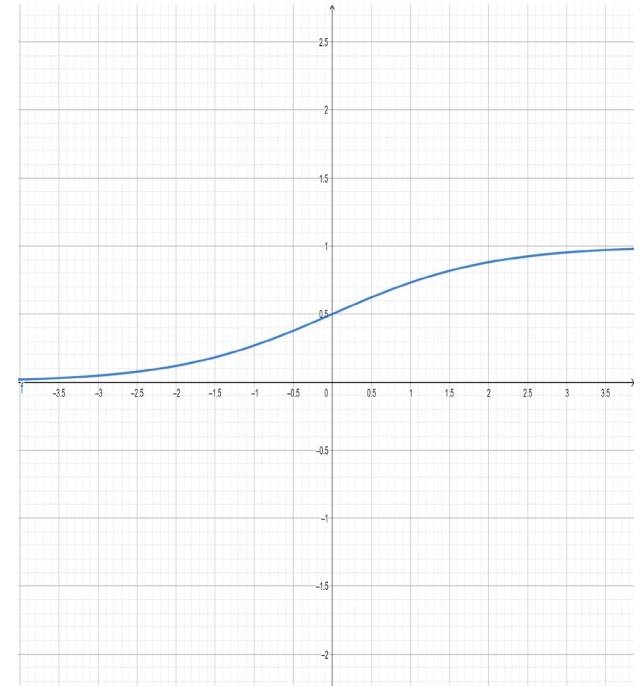
CATÉGORISATION: ENTRE 0 ET 1

- La prémisse : Beaucoup de données $(x_1, x_2, x_3 \dots, x_k, y)_j$ où les x_i s représentent k caractéristiques numériques; y est la catégorie établie.
- Le j est l'indice sur les N données.
- On utilise encore un modèle linéaire avec k paramètres, un paramètre par caractéristique.
- Voyons $k = 2$, et N données. $X_j = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$; Y_j est soit 0 ou 1 ;
 $W = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$; $X = [X_0 \quad \dots \quad X_N]$ l'ensemble des données, une matrice.
- Mais... $W^T X_j$ est un nombre réel... pas 0 ou 1... Et on cherche la probabilité d'être 0 ou 1...



LA FONCTION SIGMOÏDE POUR OBTENIR T.

- $\text{Prob} (T_j = 1 \mid X_j, W) = \frac{1}{1 + e^{-W^T X_j}}$
- La fonction sigmoïde prend le produit des paramètres et des caractéristiques, et donne un nombre entre 0 et 1.
- Parfait pour représenter une probabilité!
- On établit ensuite si $T = 1$ est « oui » ou « non », c'est purement une question d'étiquette!



JOUER AVEC LA FONCTION SIGMOÏDE

- $\text{Prob} (T_j = 0 \mid X_j, W) = 1 - \frac{1}{1+e^{-W^T X_j}} = \frac{e^{-W^T X_j}}{1+e^{-W^T X_j}}$
- On travaille avec une variable binaire, T.
- Et on dit « oui » si la probabilité ≥ 0.5 { ou « non » ... }



POINT CLÉ À RETENIR!

- **Sujet : Fonction exponentielle, Calcul Différentiel**
- **La Mathématique : La fonction sigmoïde qui représente la probabilité d'une catégorie et sa dérivée.**
- **Un lien très pertinent : [Wikipedia Sigmoid Function](#)**



ET SI LE MIRACLE DE LA CONJUGAISON N'EST PAS DISPONIBLE ??

- Dans ce cas, la vraisemblance est : $\text{Prob}(T = t \mid X_j, W) = \text{Prob}(T_j=1 \mid X_j, W)^t \times \text{Prob}(T_j=0 \mid X_j, W)^{1-t} = \left(\frac{1}{1+e^{-W^T X_j}} \right)^t \left(\frac{e^{-W^T X_j}}{1+e^{-W^T X_j}} \right)^{1-t}$
- On peut prendre notre A Priori encore dans les gaussiennes, centrée en 0 et avec une variance donnée. Ici, la forme 2-D, deux paramètres.

$$P(W) = \mathcal{N}(0, \text{Variance}) = \frac{1}{\sqrt{|\Sigma|} 2\pi} e^{-\frac{W^T \Sigma^{-1} W}{2}} \text{ où } \Sigma \text{ est la matrice de covariance } \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$

- Mais nous n'avons pas de famille conjuguée.
- On se tourne vers le « MAP » : Maximum A Posteriori



ON VEUT UN MAXIMUM : LA DÉRIVÉE = 0.

- Notre A Posteriori :

$$\text{Prob} (W | T, Y) = \frac{\text{Prob} (T | W, Y) \cdot \text{Prob} (W | \text{hypothèse initiales})}{\text{Prob} (T) \text{ sur l'entièreté des } W}$$

$$\text{■} = \frac{\left(\frac{1}{1+e^{-W^T X_j}} \right)^t \left(\frac{e^{-W^T X_j}}{1+e^{-W^T X_j}} \right)^{1-t} \cdot \mathcal{N}(0, \sigma^2)}{\text{Prob} (T) \text{ sur l'entièreté des } W}$$

Mais le dénominateur ne dépend pas de W, c'est un intégration sur les valeurs de W. Donc une constante... Si on ne regarde que le numérateur ?



LA DÉRIVÉE DE QUOI... PAR RAPPORT À QUOI?

- $\text{Prob} (W | T, Y) = \left(\frac{1}{1+e^{-W^T X_j}} \right)^t \left(\frac{e^{-W^T X_j}}{1+e^{-W^T X_j}} \right)^{1-t} \cdot \frac{1}{\sqrt{|\Sigma|} 2\pi} e^{-\frac{W^T \Sigma^{-1} W}{2}}$
- Ceci est horrible au premier regard... Mais un excellent exercice de **dérivation logarithmique!**
- **Log*** $\left(\left(\frac{1}{1+e^{-W^T X_j}} \right)^t \left(\frac{e^{-W^T X_j}}{1+e^{-W^T X_j}} \right)^{1-t} \cdot \frac{1}{\sqrt{|\Sigma|} 2\pi} e^{-\frac{W^T \Sigma^{-1} W}{2}} \right)$ n'est pas si terrible!
- Le Maximum du logarithme indique bien le maximum de notre \bar{A} Posteriori.
- * En IA, Log est la notation pour ln, logarithme naturel.



POINT CLÉ À RETENIR!

- **Sujet : Calcul Différentiel**
- **La Mathématique : La dérivation logarithmique pour trouver la première et seconde dérivées de la distribution A Posteriori.**
- **Un lien très pertinent : un bon livre de Calcul Différentiel.**

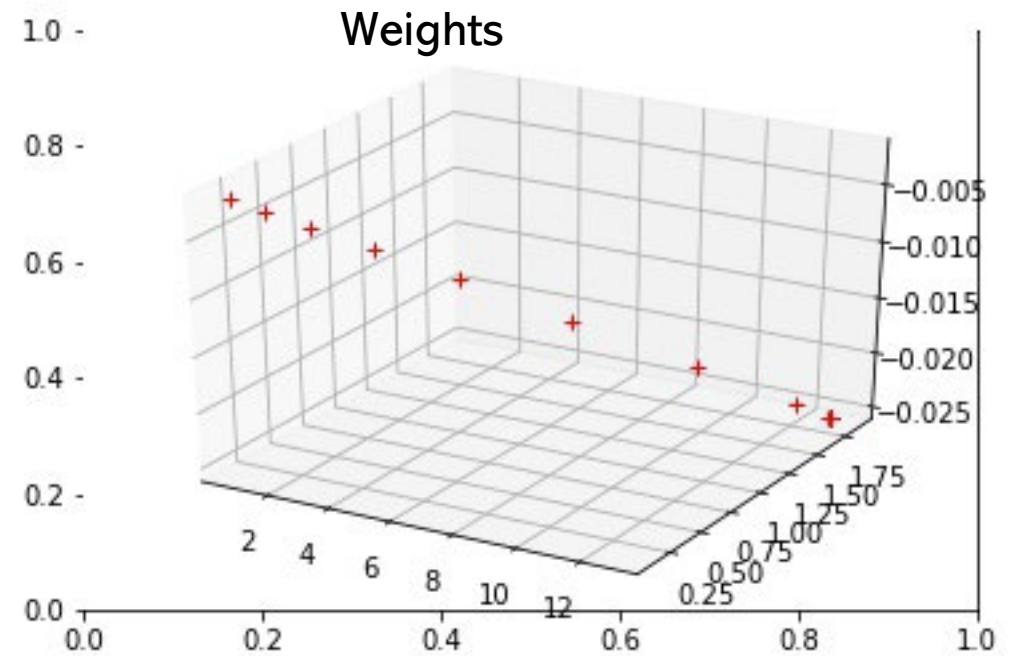
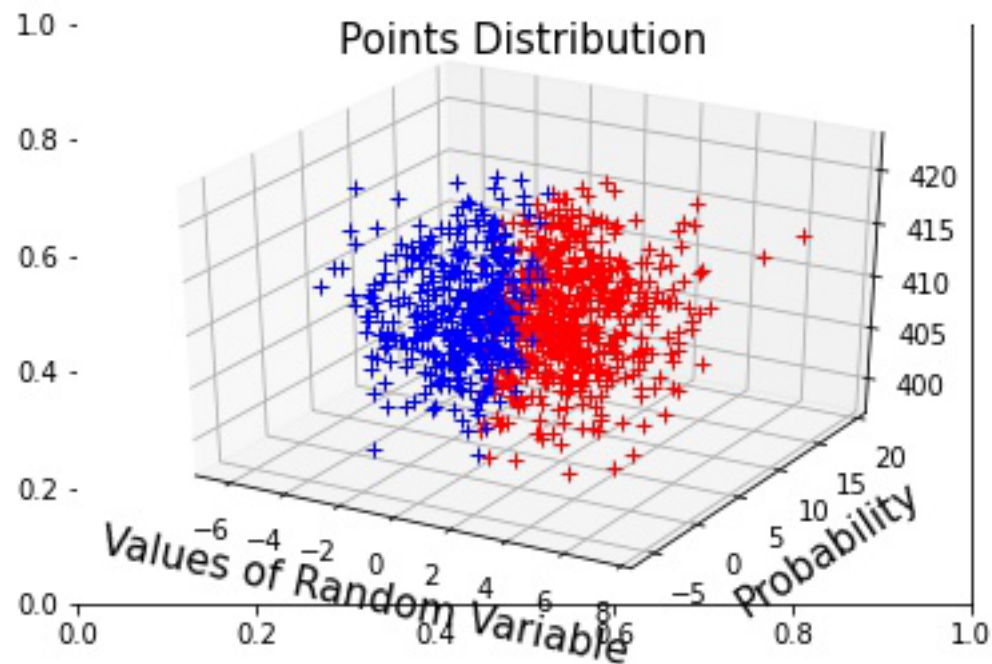


ON UTILISE LA MÉTHODE NEWTON-RAPHSON

- Encore une autre belle justification de notre cours de Calcul Différentiel!
- On veut le maximum du Log ... donc un zéro de la dérivée...
- Donc nous aurons besoin de la première et seconde dérivées.
- $W^1 = W^0 - f' / f'' = W^0 - f' \cdot (f'')^{-1}$ car f'' sera une matrice.
- 1^{re} dérivée : $\frac{1}{\sigma^2} w + \sum_{j=1}^N x_j (t_j - P_j)$ où il y a N points de données, $P_j = \left(\frac{1}{1 + e^{-w^T x_j}} \right)$
- 2^{de} dérivée : $-\frac{1}{\sigma^2} I - \sum_{j=1}^N x_j x_j^T P_j (1 - P_j)$



EXEMPLE COMPLETE AVEC 3 PARAMÈTRES.



Voir le code sous vaniercollegeai.ca / PowerPoint – Intro to ML / Categorisation



AVEC LE MAXIMUM DU LOG DE LA VRAISEMBLANCE...

- On a le meilleur vecteur de paramètres.
- On peut rebâtir la distribution de l'A Posteriori en simulant une Gaussienne autour du maximum trouvé. La variance sera trouvée par échantillonnage.
- Ouf... Voilà notre tour complété!



POINT CLÉ À RETENIR!

- **Sujet : Calcul Différentiel**
- **La Mathématique : La méthode de Newton-Raphson pour trouver les meilleurs paramètres.**
- **Un lien très pertinent : votre livre de calcul différentiel.**



POINT CLÉ À RETENIR!

- **Sujet : Méthodes itératives.**
- **Informatique : Une itération jusqu'à une solution acceptable, contrôle de boucles.**
- **Un lien pertinent : Les définitions des contrôles de boucles de votre langage préféré.**



RESSOURCES SUR LE SITE / ML COURSE –PPT SLIDES

The screenshot shows a web browser window displaying a SharePoint OneDrive page. The browser's address bar shows the URL: `eduvaniercollegeqc-my.sharepoint.com/personal/plantej_vaniercollege_qc_ca/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fplantej_vaniercollege_qc_ca%2FDocuments%2FAI%20r...`. The page header includes the Vanier College logo and 'Office 365'. Below the header, there is a 'Download' button and a 'Sort' dropdown menu. The main content area shows a breadcrumb trail: 'Julie Plante > AI research OD > PowerPoints - Intro to ML'. A table lists the following items:

Name	Modified	Modified By	File size	Sharing
100 meter AI example	October 13, 2021	Julie Plante	2 items	Shared
Categorisation	January 4	Julie Plante	1 item	Shared
Convolution Example	October 13, 2021	Julie Plante	2 items	Shared
Beta_Distribution.ipynb	September 28, 2021	Julie Plante	25.5 KB	Shared
Machine Learning chapter 1.pptx	February 2	Julie Plante	3.96 MB	Shared
Machine Learning chapter 2.pptx	June 30, 2021	Julie Plante	1.63 MB	Shared
Machine Learning chapter 3.pptx	April 8	Julie Plante	1.72 MB	Shared
Machine Learning chapter 4.pptx	April 8	Julie Plante	1.75 MB	Shared
Machine Learning chapter 5.pptx	April 12	Julie Plante	639 KB	Shared

The Windows taskbar at the bottom shows the system tray with the date '2022-06-06' and time '4:07 PM'. The system language is set to 'ENG CMS'.





OUPS! ET LA MÉCANIQUE NEWTONNIENNE?

- Pas le temps! Désolée.
- Voir la référence d'une ML qui cherche le meilleur ensemble de paramètres... et qui répond à une équation différentielle!

Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations, Maziar Raissi, Alireza Yazdani, George Em Karniadakis, dans <http://science.sciencemag.org/>, Juin 2021.



LECTURES INTÉRESSANTES . . .

- Sejnowski, T. (2018) *Deep Learning Revolution*, MIT Press.
- Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*, MIT Press, retrieved from <http://www.deeplearningbook.org>.
- Rogers, S., Girolami, M. (2017) *First Course in Machine Learning* 2nd Ed., CRC Press 
- Turing, A. M. (1950) *Computing Machinery and Intelligence*, Oxford University Press on behalf of Mind Association, retrieved from <https://www.jstor.org/stable/2251299> 
- Walsh, T. (2018), *Machines that think*, Prometheus Books.
- Fan, S. (2019), *Will AI replace us?*, Thames & Hudson.
- From Python Software Foundation, *Python documentation contents*, retrieved from <https://docs.python.org/3/contents.html>.
- From TensorFlow youtube channel, *intro to machine learning – part 1*, retrieved from <https://www.youtube.com/watch?v=KNAWp2S3w94>.



ET VOS QUESTIONS !??

- À vous la parole!
- Qui a la première question... ???



Merci bien!!

- Pour toute question et commentaire :

plantej@vanier.college

Site: vaniercollegeai.ca

