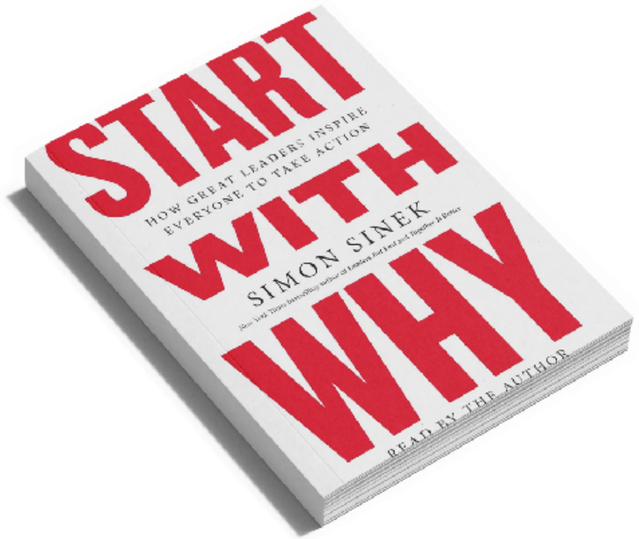
Techniques de réduction de dimension

Hicham Janati

hjanati@insea.ac.ma

INSEA

Techniques de réduction de dimension

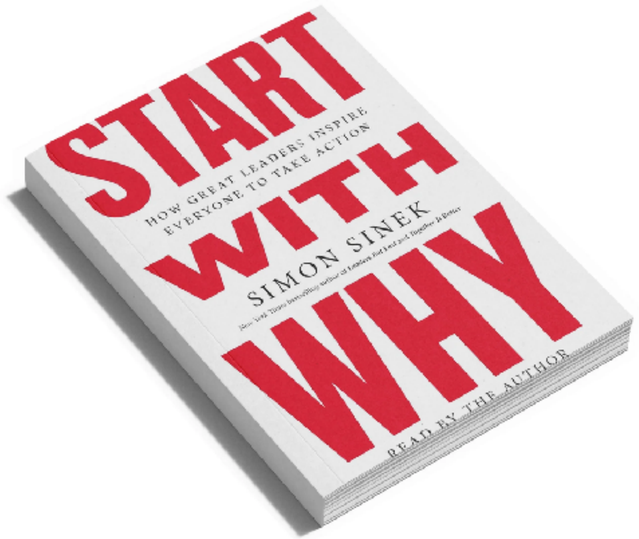


Hicham Janati

hjanati@insea.ac.ma

INSEA

Pourquoi “des techniques de réduction de dimension” ?

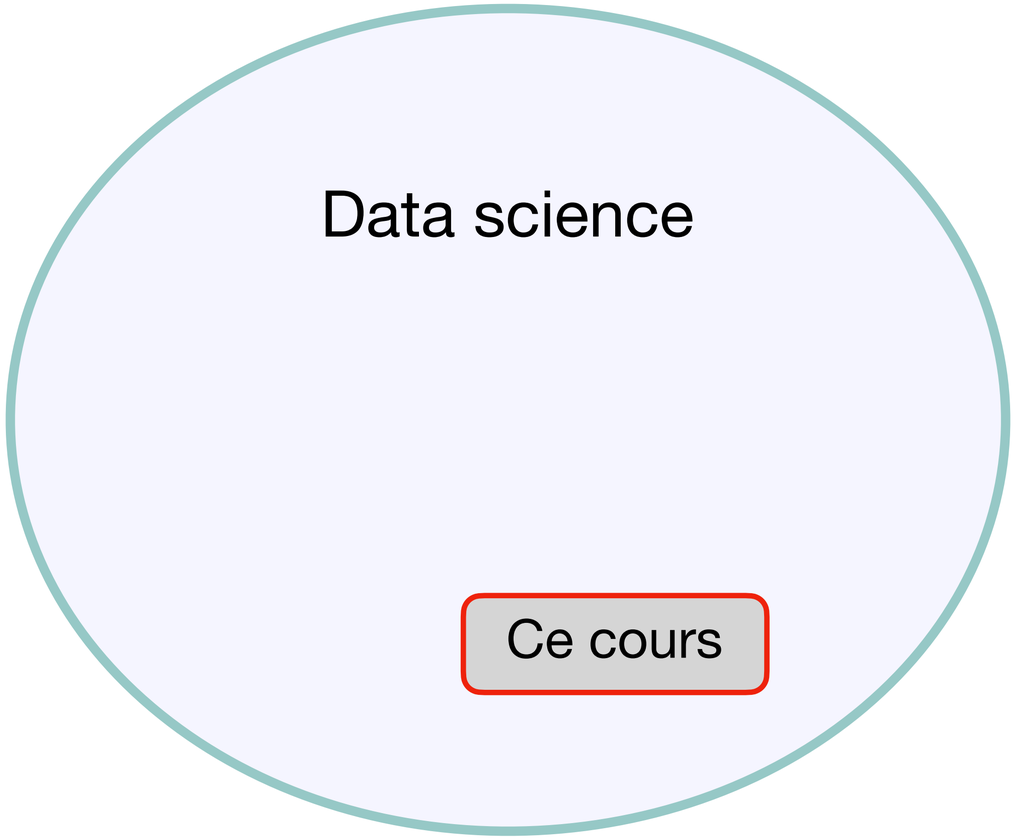


Hicham Janati

hjanati@insea.ac.ma

INSEA

Avant de répondre à cette question: il faut prendre du recul



INSEA

Data science: Introduction

INSEA

Data science: Extraire de la “valeur” à partir des données

INSEA

Data science: Extraire de la “valeur” à partir des données

1. Quantifier un effet ou une différence:

“Un Labo pharmaceutique invente un médicament pour l’hypertension. Est-il efficace ?”

Statistiques

“Vous créez deux publicités de votre produit. Laquelle des deux est plus rentable ?”

“Vous gérez la plateforme Glovo. Offrir un code promo aux nouveaux clients fait-il augmenter votre profit à long terme ?”

2. Prédire

Churn prediction: Prédire si un client va bientôt annuler son abonnement (téléphonique, Netflix, ChatGPT…)

Times series forecasting: Prédire le chiffre d’affaire d’une entreprise, la demande ou le prix du Bitcoin..

Maintenance: Prédire si une machine (usine) ou un train risque de tomber en panne

Reconnaissance d’objets dans une image / vidéo

Machine Learning

3. Générer

Systèmes de recommendation: Youtube, Netflix, Feed réseaux sociaux

Traduction, speech-to-text, text-to-speech

Large Language Models (LLMs): ChatGPT, Claude, Gemini
Diffusion models (DALL-E, midjourney…)

Modèles et algorithmes basés sur:

Les probabilités / stats: Quantifier l'incertitude, simuler / générer des nombres

L'algèbre linéaire: Compresser les données et faciliter les calculs algébriques

L'analyse / Calcul différentiel: Algorithmes d'optimisation

+ Théorie de l'information, théorie des graphes

Nécessaire pour comprendre mais pas pour "faire"

INSEA

Ce que vous allez faire:

Statisticien / Data analyst

1. Data cleaning / merging
2. Visualization / Dashboards
3. A/B tests, linear models, time series
4. Little to no code
5. Very close to business incentives

Data scientist “classique”

1. Data cleaning / merging
2. Visualization / Dashboards
3. Training ML models / benchmarks
4. Medium to high code
5. Close to business incentives

Data engineer

1. Data architecture
2. SQL / noSQL
3. Using / Serving APIs
4. Familiar with cloud providers

MLOps

1. Deploying models
2. Managing and optimizing servers
3. CI / CD

Software dev

1. Web / mobile dev
2. Front / backend

AI engineer

1. Hosting pre-trained models (Computer vision, audio)
2. Using LLMs through APIs (ChatGPT, Claude..)
3. Agents (chatbots, RAG ..)

Les probabilités: Quantifier l’incertitude, simuler / générer des nombres

L’algèbre linéaire: Compresser les données et faciliter les calculs algébriques

L’analyse / Calcul différentiel: Algorithmes d’optimisation

Programmation

Ce que vous allez faire:

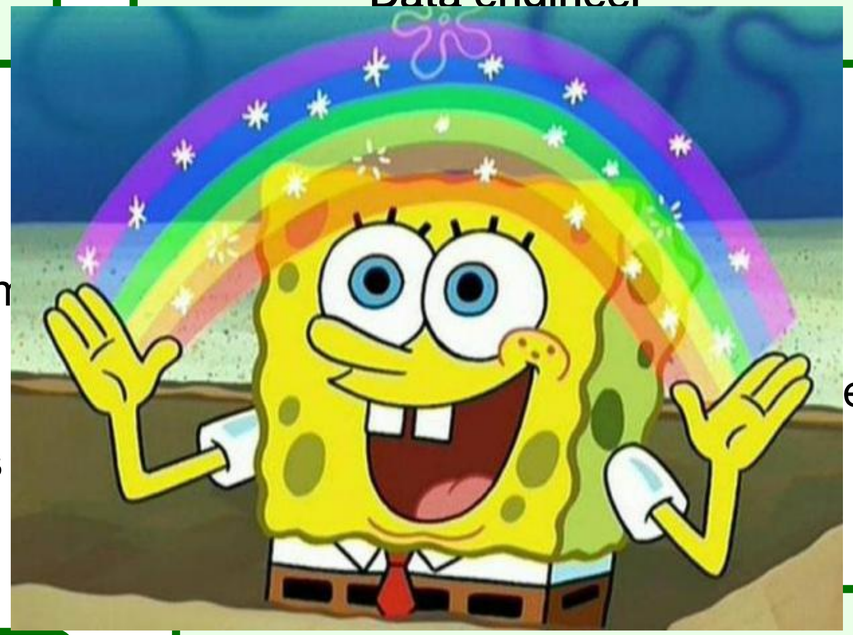
Statisticien / Data analyst

1. Data cleaning / merging
2. Visualization / Dashboards
3. A/B tests, linear models, time series
4. Little to no code
5. Very close to business incentives

Data scientist “classique”

1. Data cleaning / merging
2. Visualization / Dashboards
3. Training ML models / benchm
4. Medium to high code
5. Close to business incentives

Data engineer



ers

MLOps

1. Deploying models
2. Managing and op
3. CI / CD

ev

dev

nd

AI engineer

1. Hosting pre-trained models (Computer vision, audio)
2. Using LLMs through APIs (ChatGPT, Claude..)
3. Agents (chatbots, RAG…)

Les probabilités: Quantifier l’incertitude, simuler / générer des nombres

L’algèbre linéaire: Compresser les données et faciliter les calculs algébriques

L’analyse / Calcul différentiel: Algorithmes d’optimisation

Programmation

Adaptation, auto-formation, apprentissage continu…

A-t-on vraiment besoin des maths ?

- Pour faire des preuves au quotidien: Non
- Pour comprendre le fonctionnement des modèles: Oui

- Etes-vous un technicien ou un ingénieur ?
- Si l'IA programme mieux que vous, quelle est votre valeur ajoutée ?

Quelques conseils

La “motivation” et la “discipline” sont des illusions

1. La motivation n’est vue que par autrui
2. La discipline disparaît avec la passion
3. Il faut explorer
4. Il faut apprendre à apprendre

Comment on apprend ?

1. Active learning by doing VS passive listening

Profitez de votre présence en amphi en étant actif: réfléchir aux questions / programmation

2. Trial and error avec feedback VS peur de se tromper

Gamify learning (le prendre comme un jeu): adapter la difficulté et progresser avec des petits pas

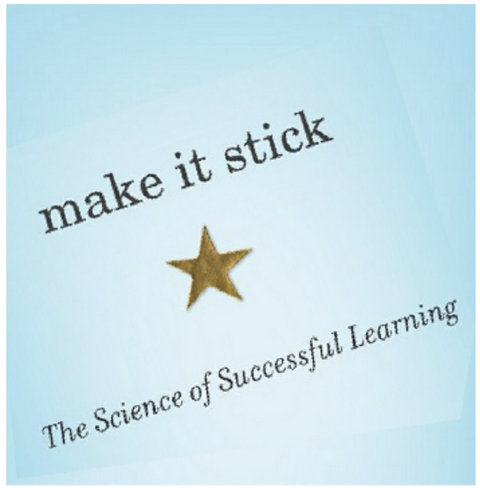
En info, la programmation vous donne un feedback immédiat

3. Self-testing en continu VS se tester à l'examen

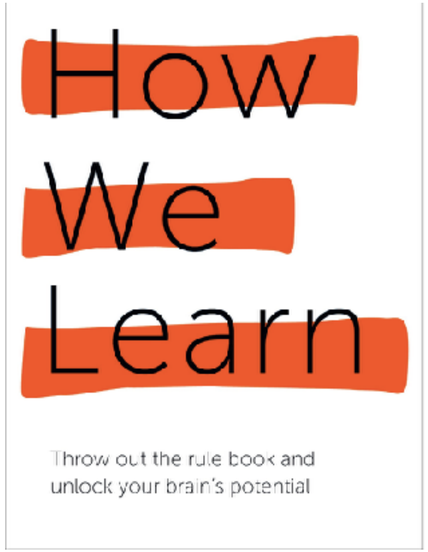
Soyez votre propre professeur: vous avez compris un concept uniquement si vous savez l'expliquer

En 2025, j'ai donné un QCM blanc anonyme avec corrigé avant l'examen. À votre avis combien l'ont fait ?

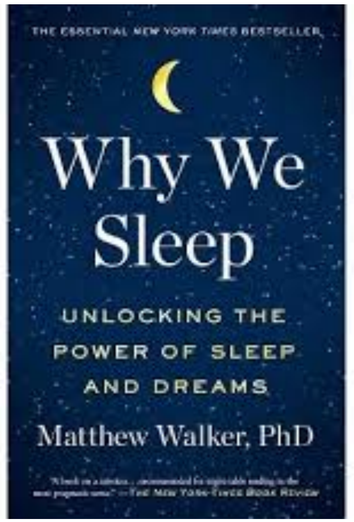
Pour aller plus loin:



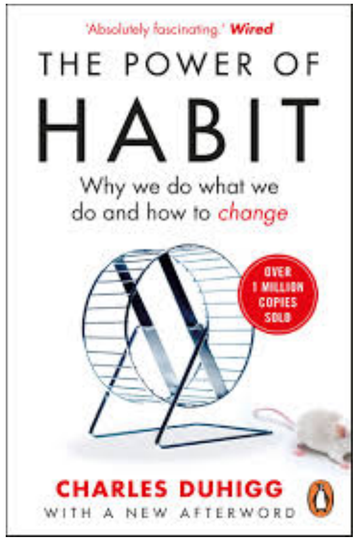
Peter C. Brown



Benedict Carey



Matthew Walker



Charles Duhigg

Techniques de réduction de dimension

Organisation

- On va alterner entre cours / TD / TP (Python): **toujours venir avec votre ordinateur**
- Un examen final + QCM / TP à rendre au cours du semestre
- -2.5% par absence sauf si parmi le top 10
- Ponctualité: porte fermée = interdit d'entrer

INSEA

Introduction: Machine Learning “à la main”

Exemple en faible dimension: e-commerce

Les données visiteurs d'un site e-commerce

Visitor ID	Name	Country	Pages Scrolled	Facebook Ad	Time on Site (min)	Purchase
1	John Smith	France	6	No	12.5	Yes
2	Maria Garcia	Spain	8	Yes	9.2	No
3	Elena Petrova	France	12	Yes	18.0	Yes
4	Ahmed Hassan	Egypt	5	No	8.4	No
5	Fatima Benali	Morocco	7	No	11.1	No
6	Anna Svensson	Spain	10	Yes	15.3	Yes
7	Carlos Mendoza	Portugal	9	No	14.6	No
			...			
2434	Emil Ivanov	Romania	4	Yes	7.4	No
2435	Amina Khaled	Egypt	11	No	16.2	Yes
2436	Layla El-Masry	Morocco	13	Yes	19.5	No
2440	Hiro Tanaka	France	6	No	10.3	Yes

Une observation: une ligne de la base données

Une variable (*feature*): une colonne de la base de données

Taille des données: Nombre d'observations = 2440

Dimension des données: Nombre de **variables = 6**

INSEA

Les données visiteurs d'un site e-commerce

Visitor ID	Name	Country	Pages Scrolled	Facebook Ad	Time on Site (min)	Purchase
1	John Smith	France	6	No	12.5	Yes
2	Maria Garcia	Spain	8	Yes	9.2	No
3	Elena Petrova	France	12	Yes	18.0	Yes
4	Ahmed Hassan	Egypt	5	No	8.4	No
5	Fatima Benali	Morocco	7	No	11.1	No
6	Anna Svensson	Spain	10	Yes	15.3	Yes
7	Carlos Mendoza	Portugal	9	No	14.6	No
...						

Une variable peut être numérique (continue) i.e prendre des valeurs dans $\mathbb{R}$

Une variable peut être catégorielle (discrète) i.e prendre des valeurs dans un ensemble fini

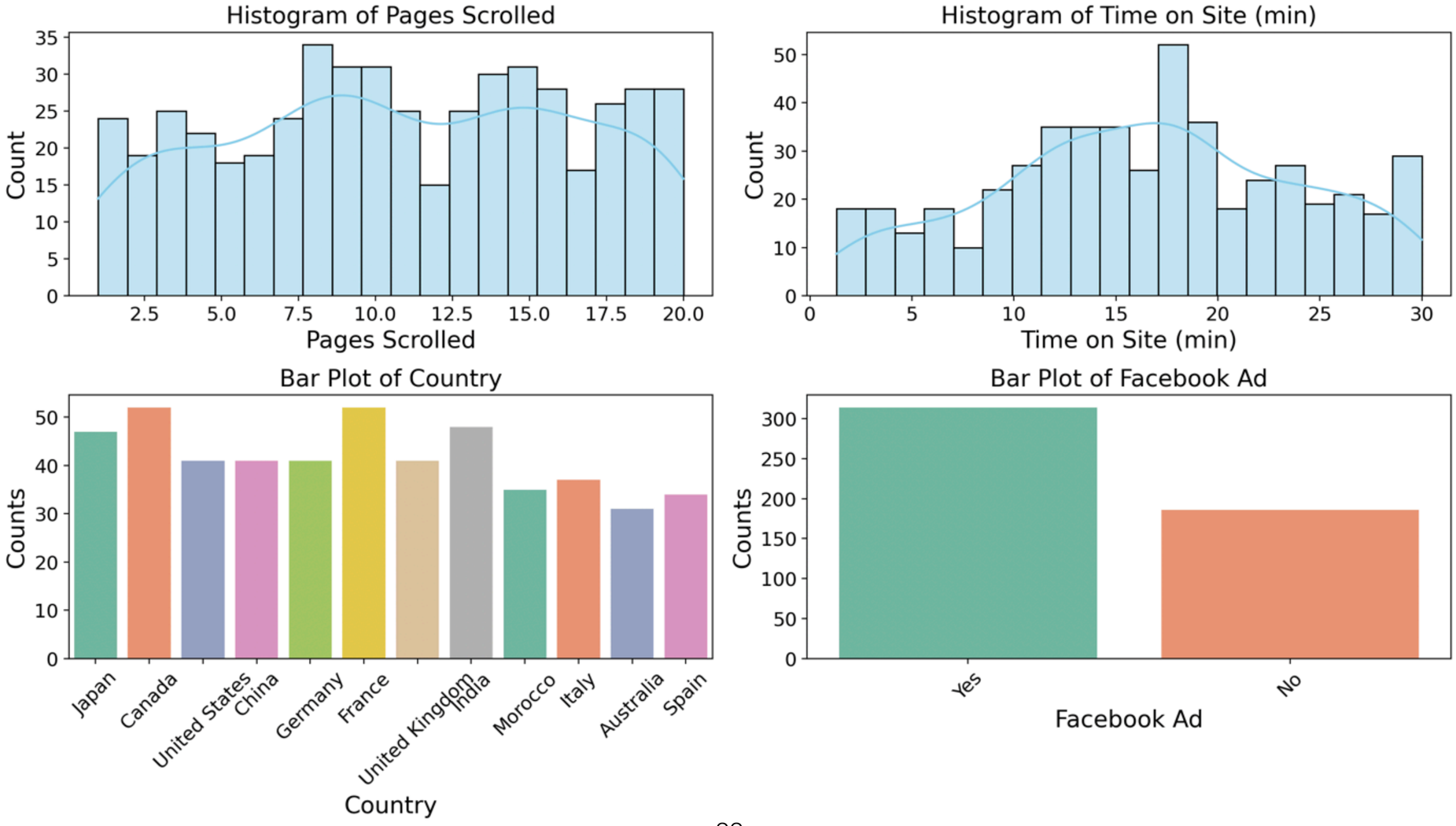
Formellement:

On a plusieurs observations $\mathbf{x}_1, \ldots, \mathbf{x}_{2440} \in \mathbb{R}^6$ du vecteur aléatoire $\mathbf{X} \in \mathbb{R}^6$

INSEA

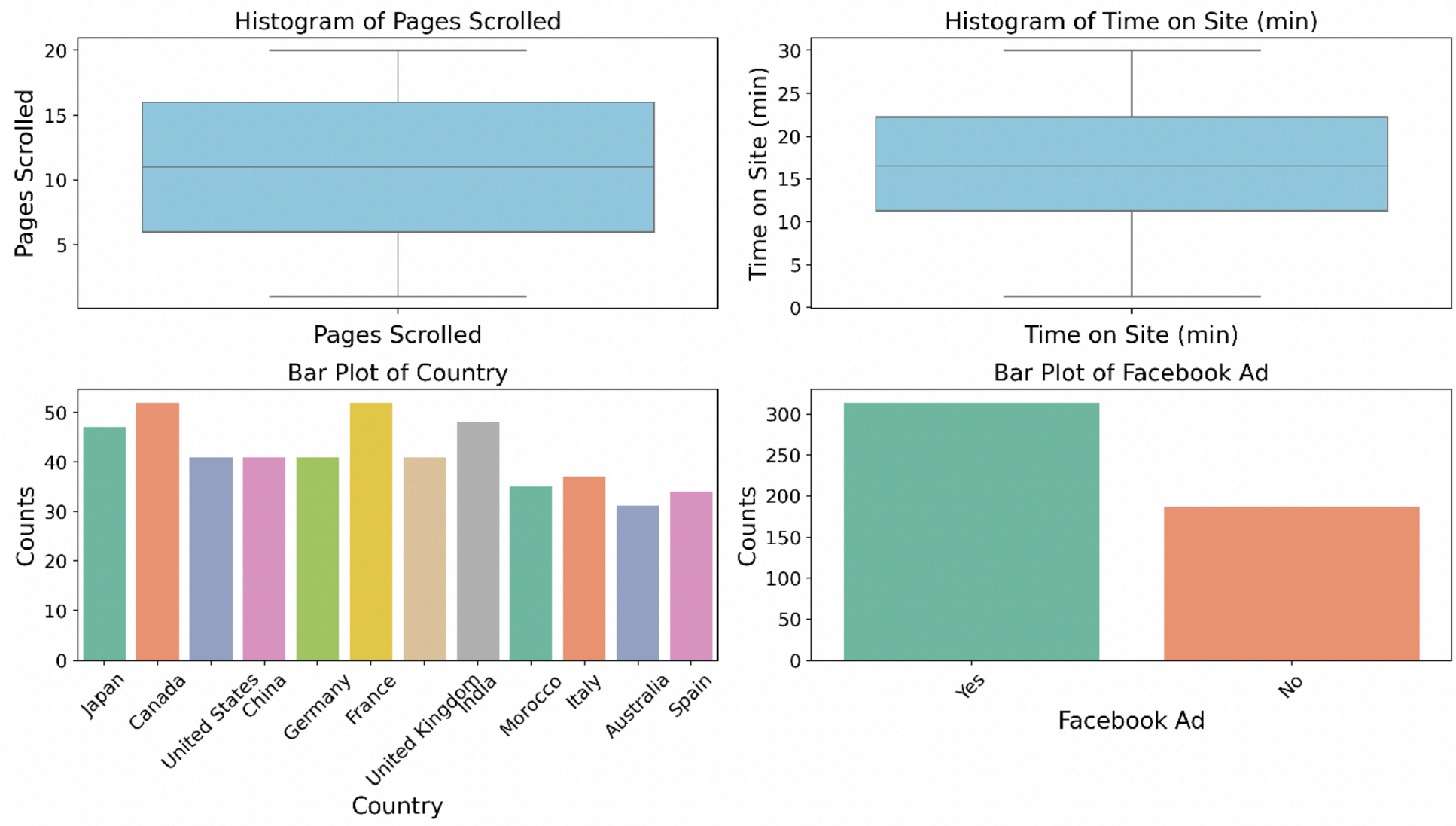
Exemple en faible dimension: e-commerce

En faible dimension on peut visualiser la distribution de chaque variable



Exemple en faible dimension: e-commerce

En faible dimension on peut visualiser la distribution de chaque variable



On souhaite développer un algorithme de prédiction: lorsqu'un visiteur est sur le site, l'algorithme devrait prédire si le visiteur va acheter ou non.

Visitor ID	Name	Country	Pages Scrolled	Facebook Ad	Time on Site (min)	Purchase
1	John Smith	France	6	No	12.5	Yes
2	M	S	8	Y	9.2	No
3	E					Yes
4	A					No
5	E					No
6	A					Yes
7	C					No
			...			
2434	Emil Ivanov	Romania	4	Yes	7.4	?
2435	Amina Khaled	Egypt	11	No	16.2	?
2436	Layla El-Masry	Morocco	13	Yes	19.5	?
2440	Hiro Tanaka	France	6	No	10.3	?

On commence par visualiser les données

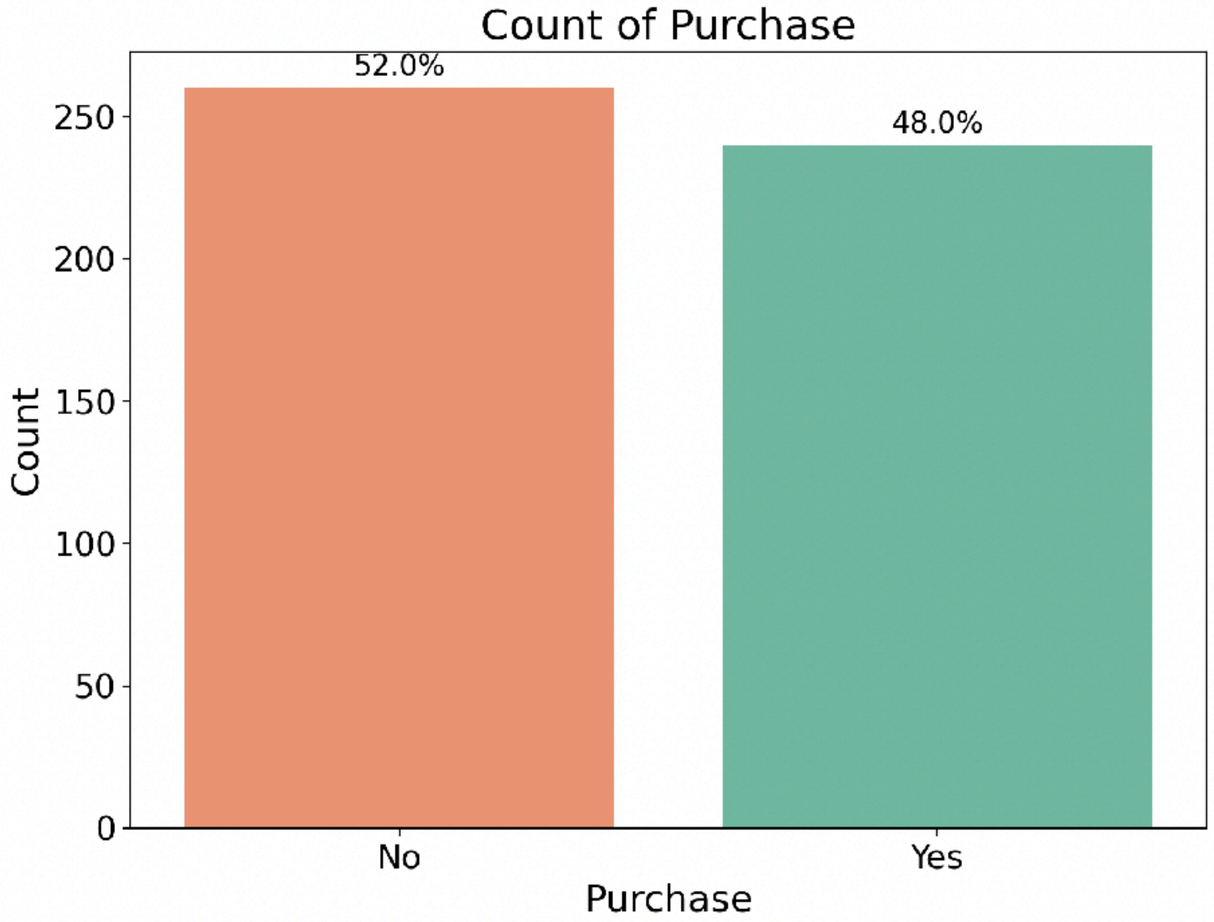
On cherche donc une fonction de la forme:

$$f(\text{Name, Pages Scrolled, Country, Time Spent, Facebook Ad}) \rightarrow \{\text{Yes, No}\}$$

f est appelée **classifieur**

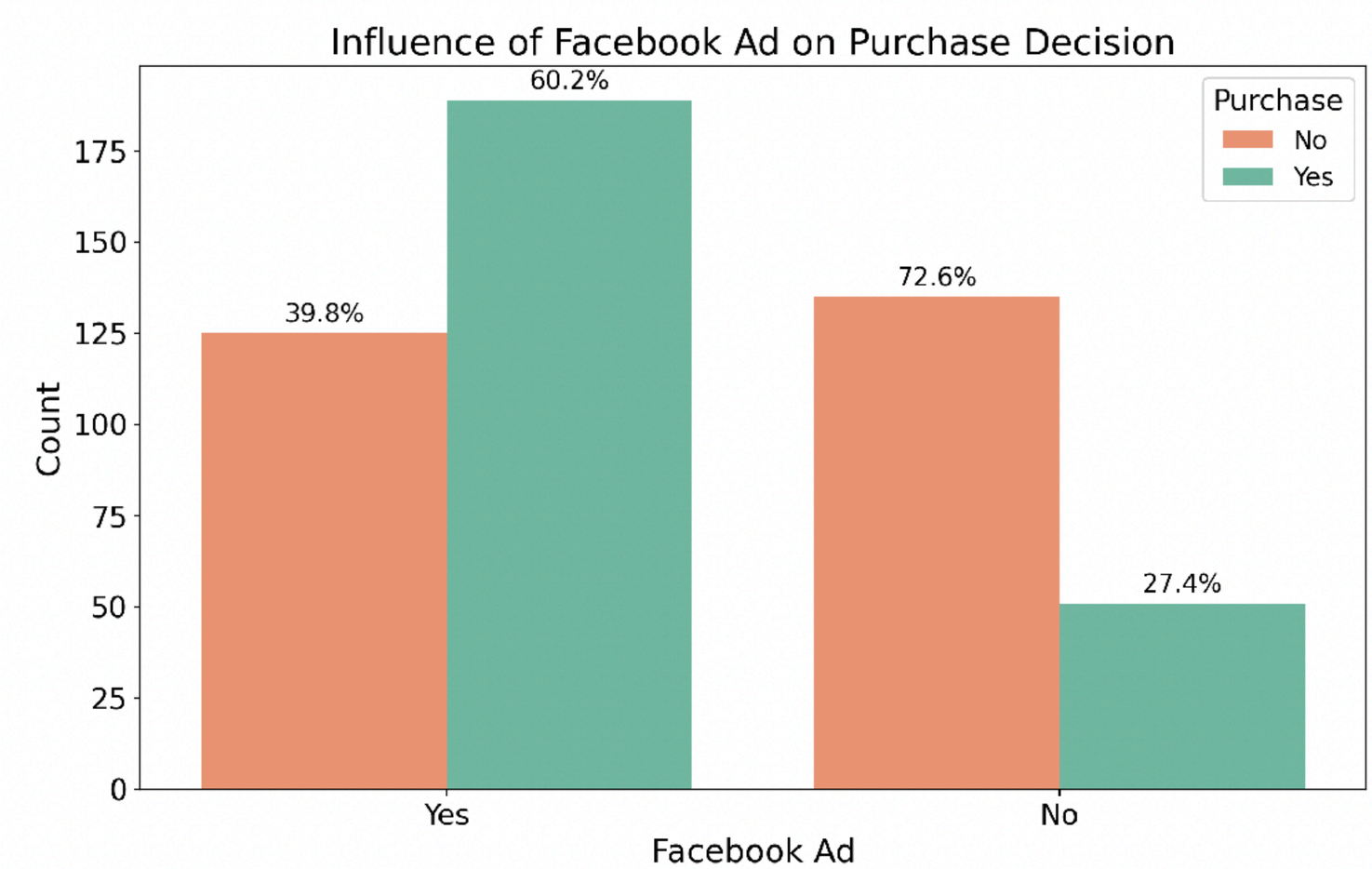
INSEA

Distribution de la variable d'intérêt:



Quel classifieur naïf peut-on déjà implémenter avec en moins 50% de précision ?

En utilisant une variable prédictive discrète



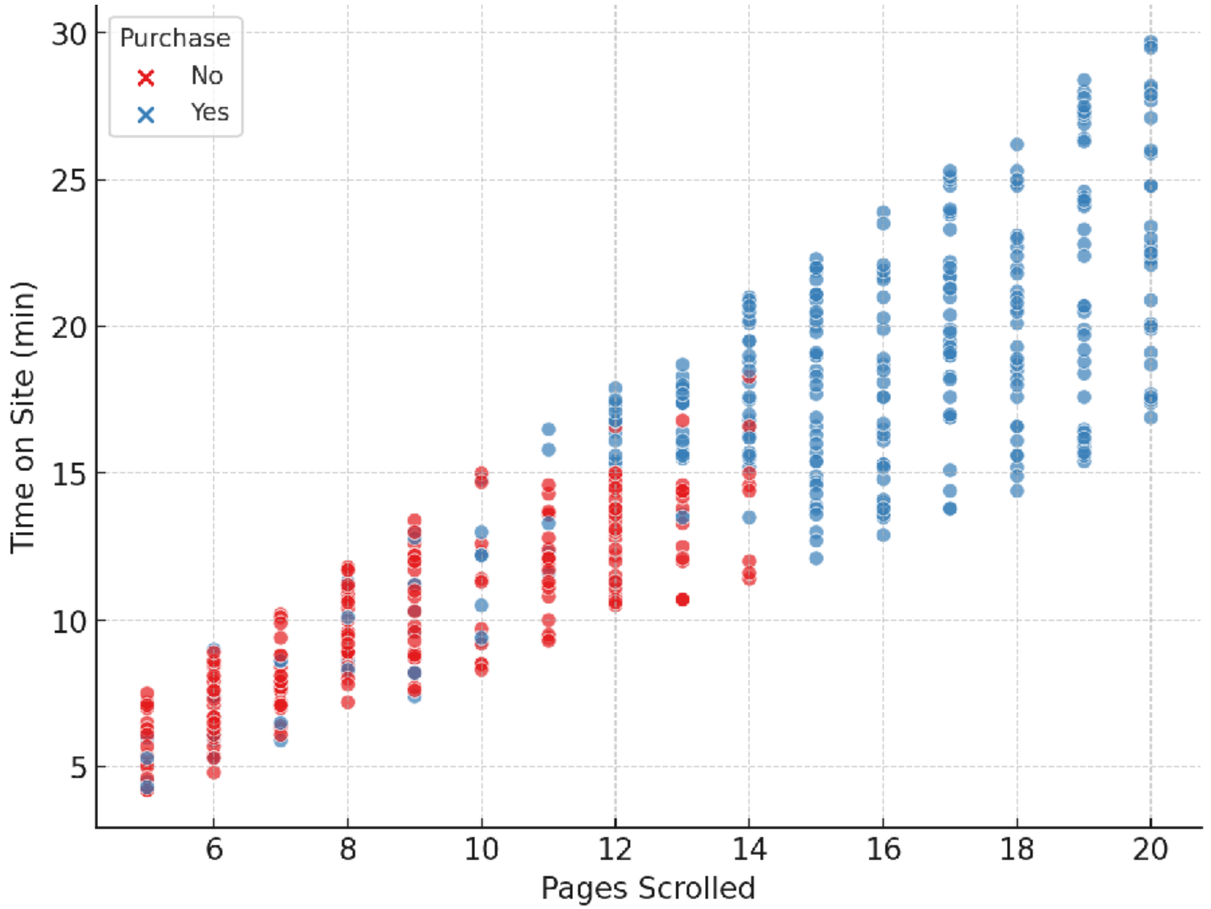
Comment améliorer le classifieur précédent ? Quelle est sa précision ?

En utilisant une variable prédictive continue

Définir une famille de classifieurs basés sur “Time on Site”.

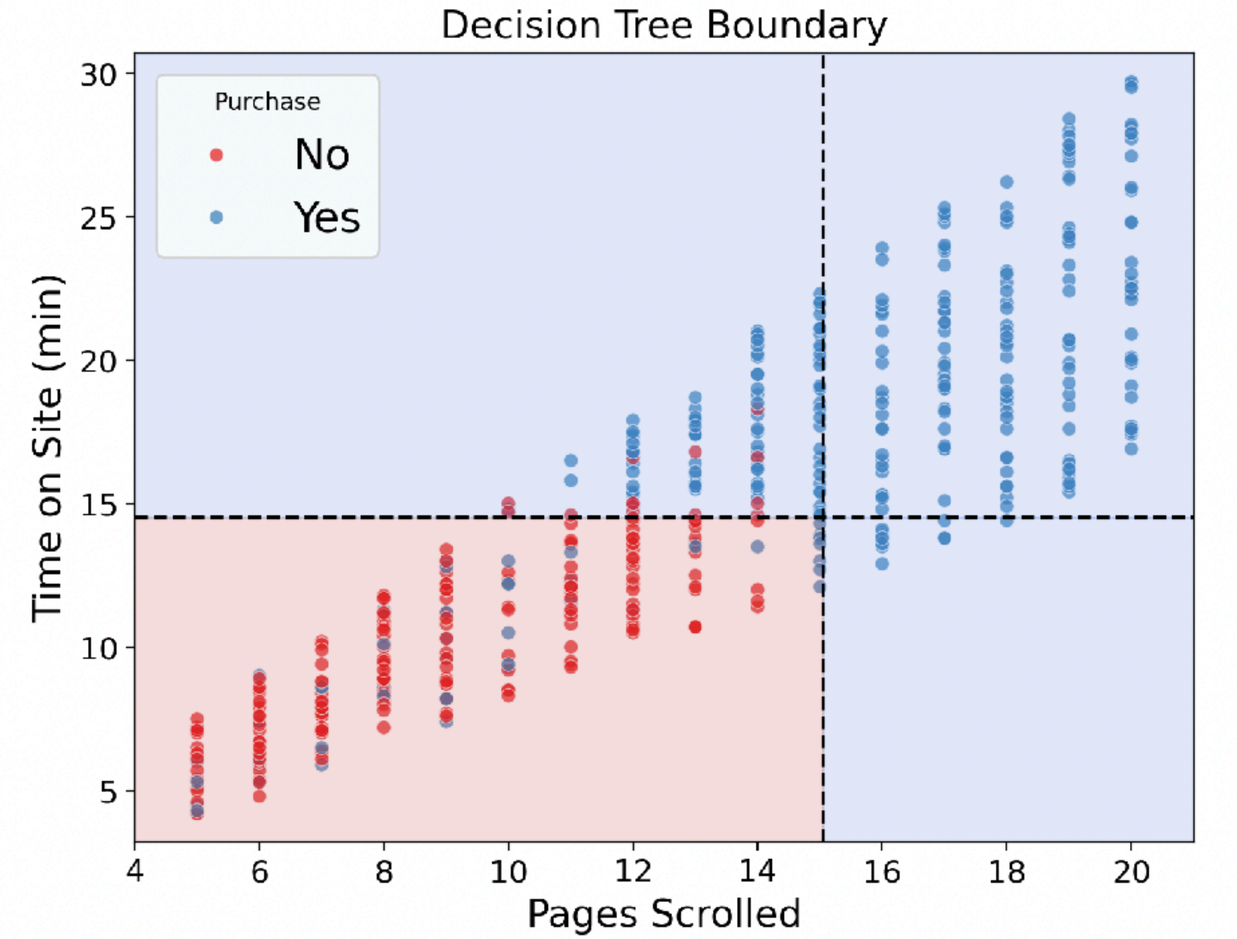
En utilisant deux variables prédictives

En utilisant deux variables prédictives



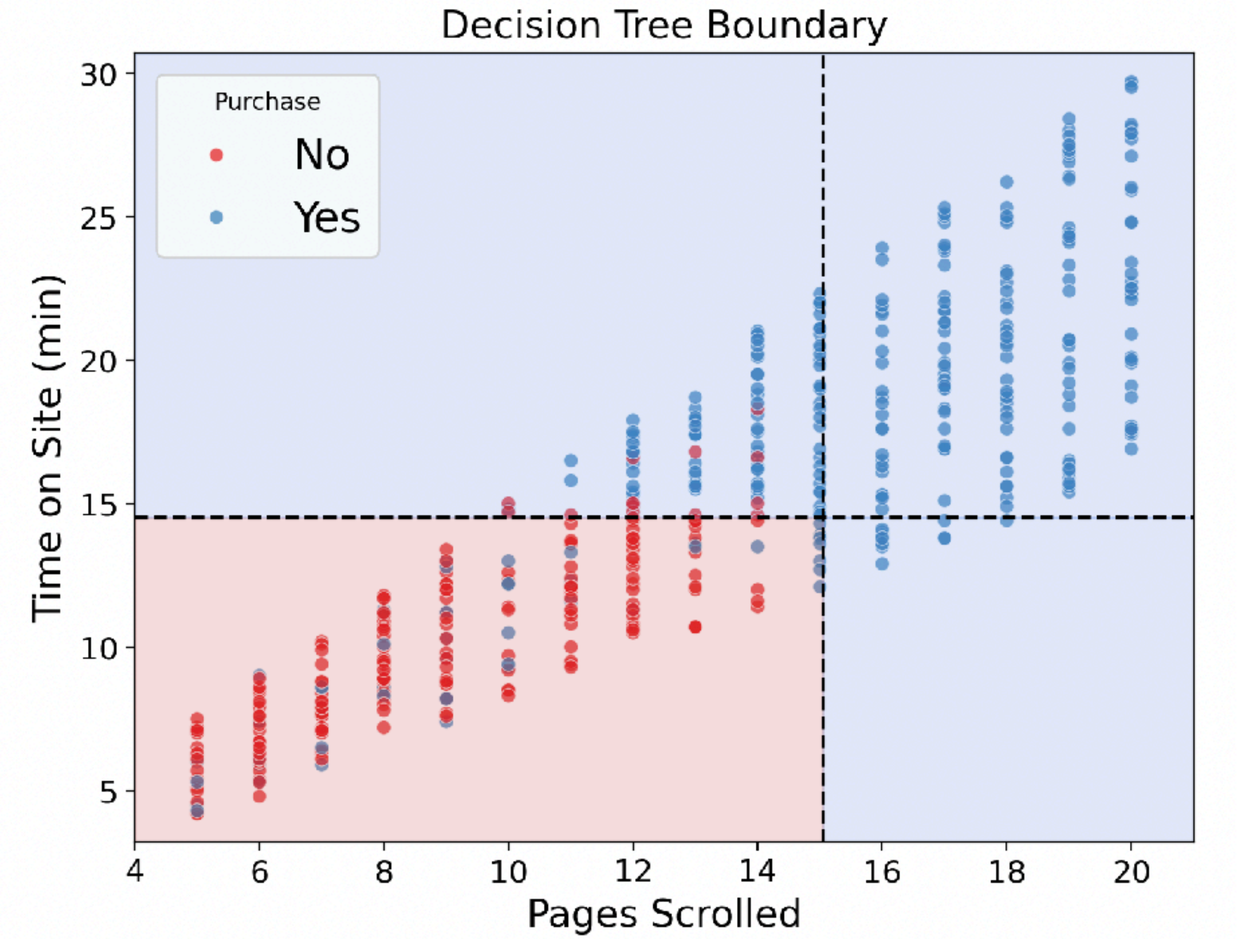
Définir un type de classifieurs basés sur “Time” et “Pages Scrolled”.

En utilisant deux variables prédictives



Définir un type de classifieurs basés sur “Time” et “Pages Scrolled”.

En utilisant deux variables prédictives



Comment peut-on rajouter la variable “Facebook Ad” dans la prise de décision ?

Dans cet exemple nous avons pu:

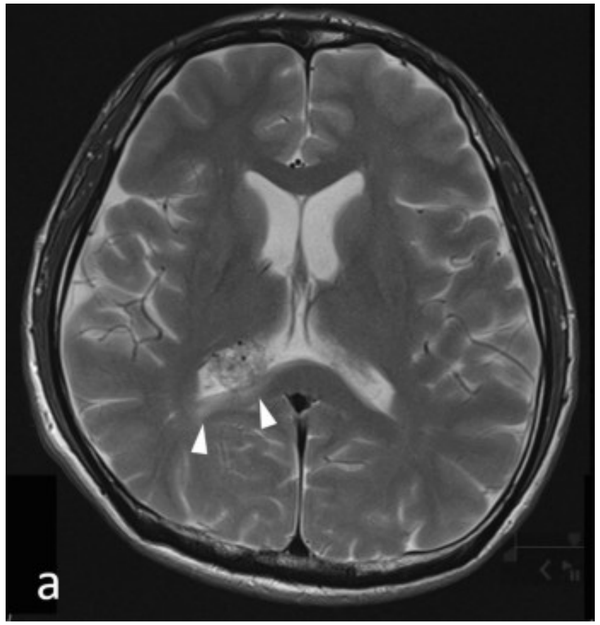
1. Voir quelques observations entières des données
2. Visualiser les distributions des variables une à une
3. Visualiser les interactions entre les variables: distribution jointe d'un vecteur aléatoire
4. Proposer des classifieurs simples en utilisant (presque) toutes les informations dans les données

Avec un grand nombre de variables, ceci devient impossible …

INSEA

Imagerie médicale

Scanner IRM du cerveau, résolution d'un mm^3. Nombre de pixels 200 x 200 = 40000 pixels. Chaque image est une matrice contenant 40000 nombres [0-256].

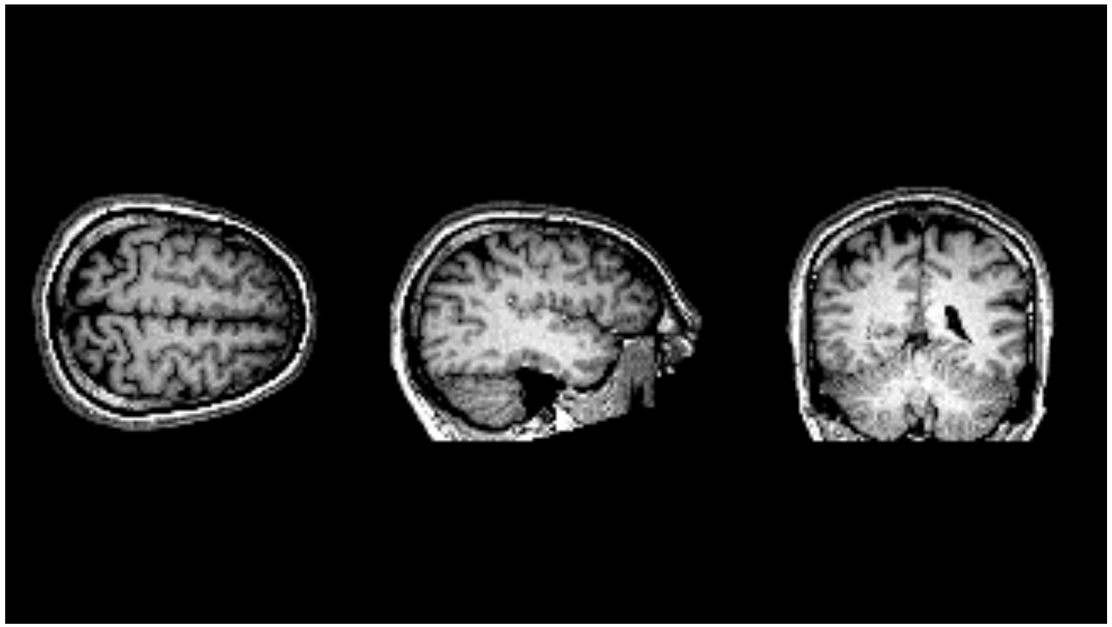


$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 120 & 150 & 200 & \cdots & 180 & 130 & 0 \\ 0 & 110 & 160 & 210 & \cdots & 190 & 140 & 0 \\ 0 & 100 & 140 & 230 & \cdots & 170 & 120 & 0 \\ 0 & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & 0 \\ 0 & 130 & 200 & 220 & \cdots & 160 & 150 & 0 \\ 0 & 140 & 180 & 240 & \cdots & 200 & 170 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Imagerie médicale

Mais on n’a pas qu’une seule image …

En moins une centaine de coupes différentes … on atteint facilement les millions.



Objectif: utiliser les données pour prédire et localiser une éventuelle tumeur

INSEA

Données texte

Les données sont des SMS avec un label: spam / non-spam

Texte SMS	**Spam (1) / Non-Spam (0)**
Salut, tu viens ce soir ?	0
Votre rendez-vous est confirmé pour demain.	0
Vous avez un colis en attente, cliquez ici pour récupérer.	**1**
Bonjour, comment vas-tu aujourd'hui ?	0
...	**...**
Inscrivez-vous maintenant pour gagner un iPhone !	**1**
J'ai bien reçu ton email, merci !	0
On se retrouve à 14h devant la gare.	0
N'oublie pas notre réunion demain matin.	0

Données texte

Les données sont des SMS avec un label: spam / non-spam

La transformation *CountVectorizer*

ID SMS	Type	Salut	tu	viens	...	colis	gagner	email	réunion
1	Ok	1	2	0	...	0	0	0	0
3	Ok	0	0	0	...	0	0	0	0
3	Spam	0	0	0	...	1	0	0	0
4	Ok	0	4	0	...	0	0	0	0
5	Spam	0	0	0	...	0	2	0	0
6	Ok	0	0	0	...	0	0	1	0
7	Ok	0	0	0	...	0	0	0	2
8	Ok	0	0	0	...	0	0	0	1
...	...	...	...	...	...	...	...	...	...

Quelle est la dimension de ces données ?

Dimension = taille du vocabulaire : nombre de mots distincts

INSEA

Pourquoi réduire la dimension ?

1. Pour visualiser les données
2. Pour compresser les données et les préparer aux modèles ML

Chapitre I: Algèbre linéaire revisitée

INSEA

Soit $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$

Le produit scalaire entre $\mathbf{a}$ et $\mathbf{b}$ s'écrit : $\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^{d} \mathbf{a}_i \mathbf{b}_i$

et peut aussi s'écrire:

$$(\mathbf{a}_1, \ldots, \mathbf{a}_d) \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_d \end{pmatrix} = \mathbf{a}^\top \mathbf{b} = \mathbf{b}^\top \mathbf{a} = \mathbf{a}^\top \mathbf{b} = \sum_{i=1}^{d} \mathbf{a}_i \mathbf{b}_i$$

Les vecteurs dans $\mathbb{R}^d$ sont toujours considérés des vecteurs colonnes.

Ainsi $\mathbb{R}^d$ et l'espace des matrices $\mathbb{R}^{d \times 1}$ sont identiques.

INSEA

Soit $\mathbf{a} \in \mathbb{R}^d$

$\text{diag}(\mathbf{a})$ correspond à la matrice diagonale $\begin{pmatrix} \mathbf{a}_1 & 0 & \dots & 0 \\ 0 & \mathbf{a}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 0 & \dots & 0 & \mathbf{a}_d \end{pmatrix}$

$\mathbb{1}_d$ correspond au vecteur de dimension d avec des 1 partout $\begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}^\top$

La matric Identité est donc $\mathrm{I}_d = \text{diag}(\mathbb{1}_d)$

Le déterminant d'une matrice carrée $\mathbf{A} \in \mathbb{R}^{d \times d}$ est noté par $\det(\mathbf{A})$ ou $|\mathbf{A}|$

INSEA

$$\text{Soit } A = \begin{bmatrix} 1 & 2 \\ 3 & 7 \\ 5 & 4 \end{bmatrix} \text{ et } x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad A = \begin{bmatrix} a_1 & a_2 \end{bmatrix}$$

Comment calculer Ax ?

1

Ligne - colonne $\begin{bmatrix} 1 & 2 \\ 3 & 7 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1x_1 + 2x_2 \\ 3x_1 + 7x_2 \\ 5x_1 + 4x_2 \end{bmatrix}$

produits scalaires des lignes avec $x = (x_1, x_2)$

Vue bas niveau (Calcul)

2

Colonne - ligne $\begin{bmatrix} 1 & 2 \\ 3 & 7 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 7 \\ 4 \end{bmatrix}$

$$Ax = x_1 a_1 + x_2 a_2$$

combinaison linéaire des colonnes a_1 et a_2

Vue haut niveau (Compréhension)

INSEA

Soit $A \in \mathbb{R}^{n \times d}$. En notation colonnes: $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_d \end{bmatrix}$

Soit $\mathbf{x} \in \mathbb{R}^d$.

L'ensemble des $A\mathbf{x}$ pour $\mathbf{x} \in \mathbb{R}^d$ correspond donc à toutes les combinaisons linéaires des $\mathbf{a}_i$: $\sum_{i=1}^{d} x_i \mathbf{a}_i$

Cet ensemble est l'image de A (*Column space of A*)

C'est l'ensemble vectoriel engendré par les colonnes de A.

On appelle sa dimension le rang de A, qui est toujours inférieure ou égale à min(n, d).

Le rang de A est égal au rang de sa transposée.

Le rang de A est la dimension de l'espace engendré par ses lignes (*row space of A*)

Règle d'or: pour bien comprendre un concept donné par un nombre, il faut étudier ses extrêmes pour le cerner

Soit $A \in \mathbb{R}^{n \times d}$. En notation colonnes: $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_d \end{bmatrix}$

1 Extrême maximal

On suppose que $n > d$ et que le rang est maximal: $rank(A) = d$.

Alors aucune colonne ne peut être exprimée linéairement en fonction des autres.

On ne peut pas “compresser” l’information des colonnes de A avec une transformation linéaire.

2 Extrême minimal

Le rang 0 est trivial: il correspond à la matrice nulle. Étudions le cas rang = 1.

Un exemple trivial serait par exemple de prendre une seule colonne **a** et de la multiplier par des scalaires

$$A = \begin{bmatrix} b_1\mathbf{a} & b_2\mathbf{a} & \dots & b_d\mathbf{a} \end{bmatrix}$$

? **Quel est le terme général de A càd A_{ij} ? Comment peut-on écrire A à l’aide d’un produit matriciel simple ?**

$$A_{ij} = b_j a_i = a_i b_j$$

$$A = \begin{bmatrix} b_1\mathbf{a} & b_2\mathbf{a} & \dots & b_d\mathbf{a} \end{bmatrix} = \begin{bmatrix} \\ a \\ \\ \end{bmatrix} \begin{bmatrix} b_1 & \dots & b_d \end{bmatrix} = \mathbf{a}\mathbf{b}^\top$$

Outer product (produit extérieur)

$$\forall \mathbf{a} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^d \quad \text{rank}(\mathbf{a}\mathbf{b}^\top) = 1$$

à ne pas confondre avec:

$$\begin{bmatrix} a \end{bmatrix} \begin{bmatrix} \\ \mathbf{b} \\ \\ \end{bmatrix} = \mathbf{a}^\top \mathbf{b} = \langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^{d} a_i b_j$$

Inner / dot product (produit scalaire)

Soit $A \in \mathbb{R}^{n\times d}$ et $B \in \mathbb{R}^{d\times m}$. On note la i-ème ligne de A par $A_{i\bullet}$ et sa j-ème colonne par $A_{\bullet j}$.

1 Vue lignes-colonnes (Méthode usuelle)

$$AB = \begin{bmatrix} A_{1\bullet}^\top \\ \vdots \\ A_{n\bullet}^\top \end{bmatrix} \begin{bmatrix} B_{\bullet 1} & \dots & B_{\bullet m} \end{bmatrix} = \begin{bmatrix} A_{1\bullet}^\top B_{\bullet 1} & \dots & A_{1\bullet}^\top B_{\bullet m} \\ \vdots & A_{i\bullet}^\top B_{\bullet j} & \vdots \\ A_{n\bullet}^\top B_{\bullet 1} & \dots & A_{n\bullet}^\top B_{\bullet m} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{k=1}^d A_{1k}B_{k1} & \dots & \sum_{k=1}^d A_{1k}B_{km} \\ \vdots & \sum_{k=1}^d A_{ik}B_{kj} & \vdots \\ \sum_{k=1}^d A_{nk}B_{k1} & \dots & \sum_{k=1}^d A_{nk}B_{km} \end{bmatrix}$$

? **Trouvez la vue colonnes-lignes du produit matriciel ci-dessus.**

2 Vue colonnes-lignes

$$= \sum_{k=1}^d \begin{bmatrix} A_{1k}B_{k1} & \dots & A_{1k}B_{km} \\ \vdots & A_{ik}B_{kj} & \vdots \\ A_{nk}B_{k1} & \dots & A_{nk}B_{km} \end{bmatrix} = \sum_{k=1}^d A_{\bullet k}B_{k\bullet}^\top = \begin{bmatrix} A_{\bullet 1} & \dots & A_{\bullet d} \end{bmatrix} \begin{bmatrix} B_{1\bullet}^\top \\ \vdots \\ B_{d\bullet}^\top \end{bmatrix}$$

INSEA

Soit $A \in \mathbb{R}^{n \times d}$ et $B \in \mathbb{R}^{d \times m}$. On note la i-ème ligne de A par $A_{i\bullet}$ et sa j-ème colonne par $A_{\bullet j}$.

$$AB = \begin{bmatrix} A_{\bullet 1} & \dots & A_{\bullet d} \end{bmatrix} \begin{bmatrix} B_{1\bullet}^\top \\ \vdots \\ B_{d\bullet}^\top \end{bmatrix} = \sum_{k=1}^{d} A_{\bullet k} B_{k\bullet}^\top$$

Somme de d matrices de rang 1.

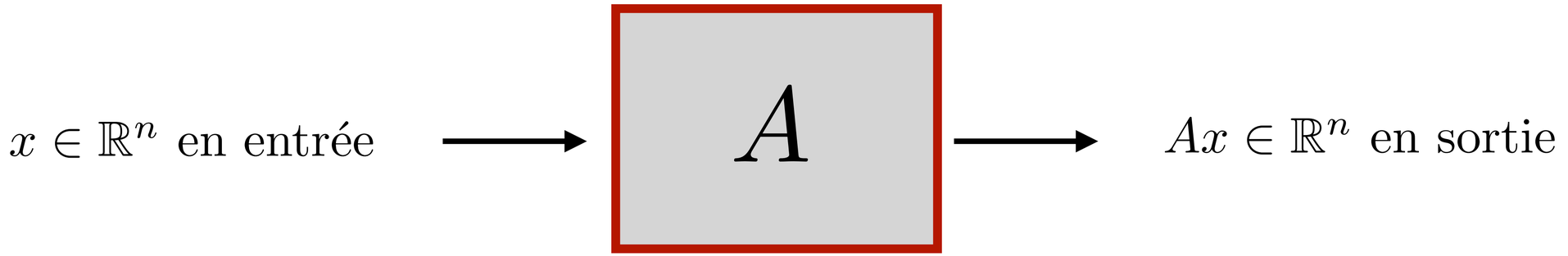
? **Comment peut-on retrouver la vue colonnes-lignes du produit matriciel ci-dessus directement à partir de AB ?**

La vue colonnes-lignes peut être également obtenue avec un produit matriciel par blocs:

$$\left[\begin{array}{c|c|c} A_{\bullet 1} & \dots & A_{\bullet d} \end{array}\right] \left[\begin{array}{c} B_{1\bullet}^\top \\ \hline \vdots \\ \hline B_{d\bullet}^\top \end{array}\right] = \sum_{k=1}^{d} A_{\bullet k} B_{k\bullet}^\top$$

INSEA

Pour simplifier et avoir le même espace d'arrivée et de départ, on suppose que A est carrée dans $\mathbb{R}^{n\times n}$.



Règle d'or: pour bien comprendre un concept donné par une transformation, il faut étudier sa stabilité / ses invariances, c-à-d, les entrées sur lesquelles peu de changement a lieu.

Existent-t-il des directions (vecteurs) inchangées par A ?

INSEA

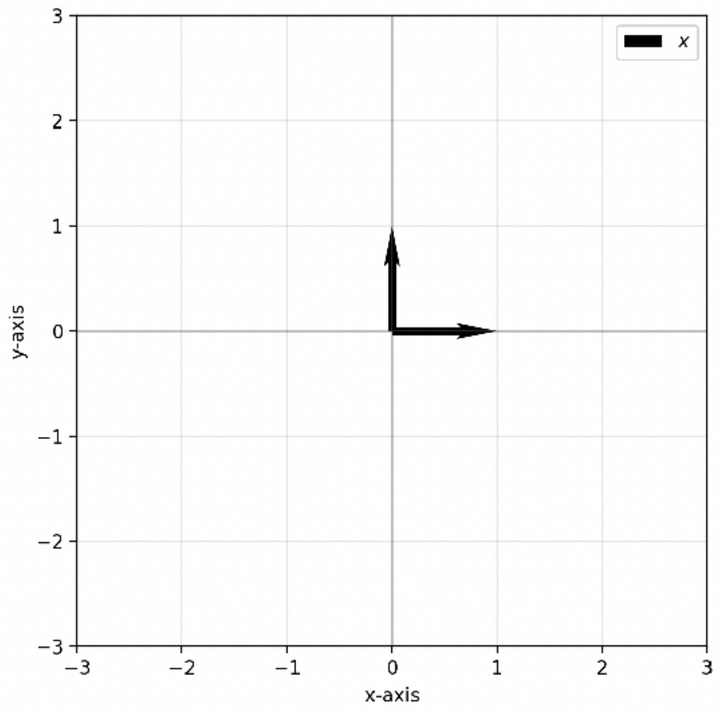
Pour comprendre l'action de la matrice, faisons quelques visualisations en 2D

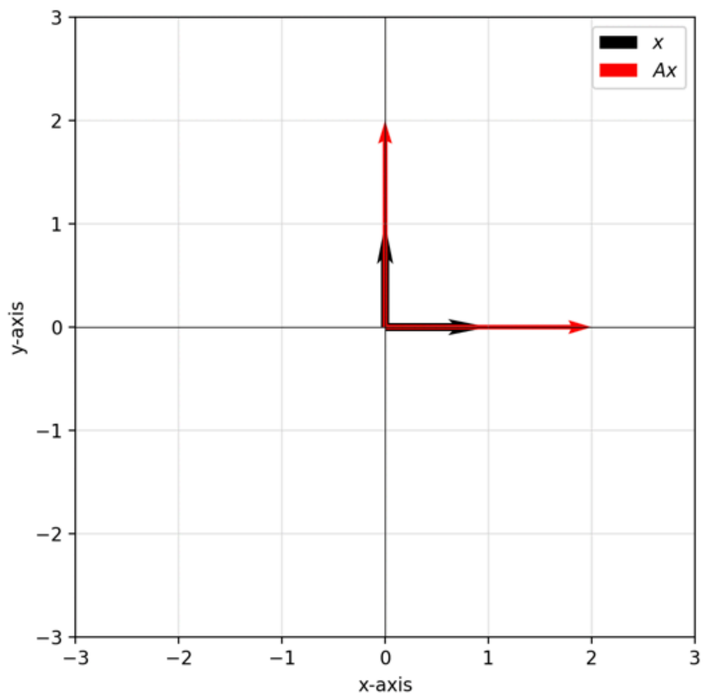
Commençons par des matrices simples avec beaucoup de 0

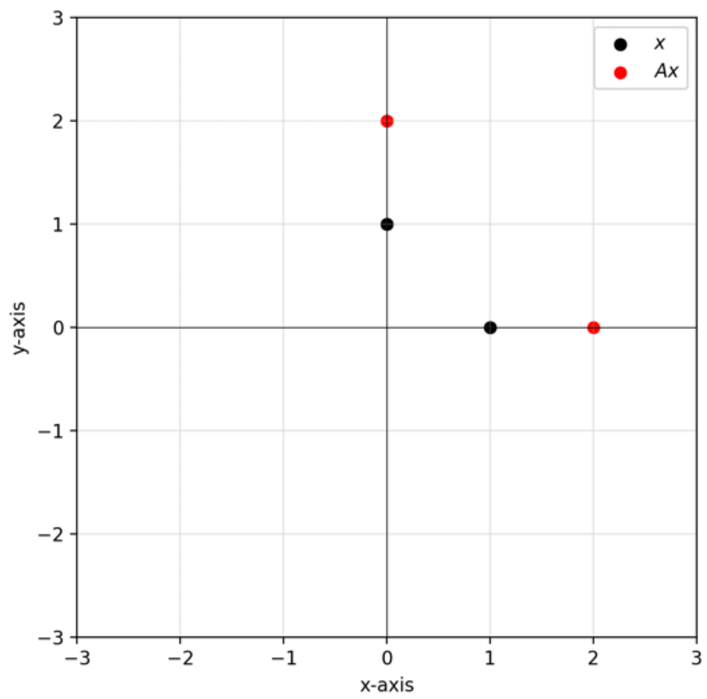
1. Matrices diagonales

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

On prend deux vecteurs x de la base Euclidienne:





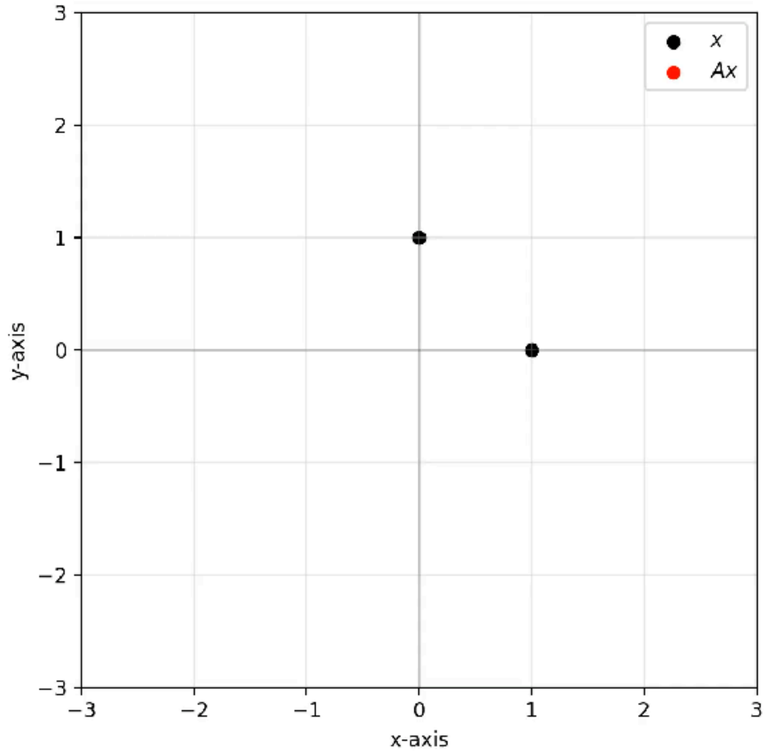


Pour comprendre l'action de la matrice, faisons quelques visualisations en 2D

Commençons par des matrices simples avec beaucoup de 0

1. Matrices diagonales

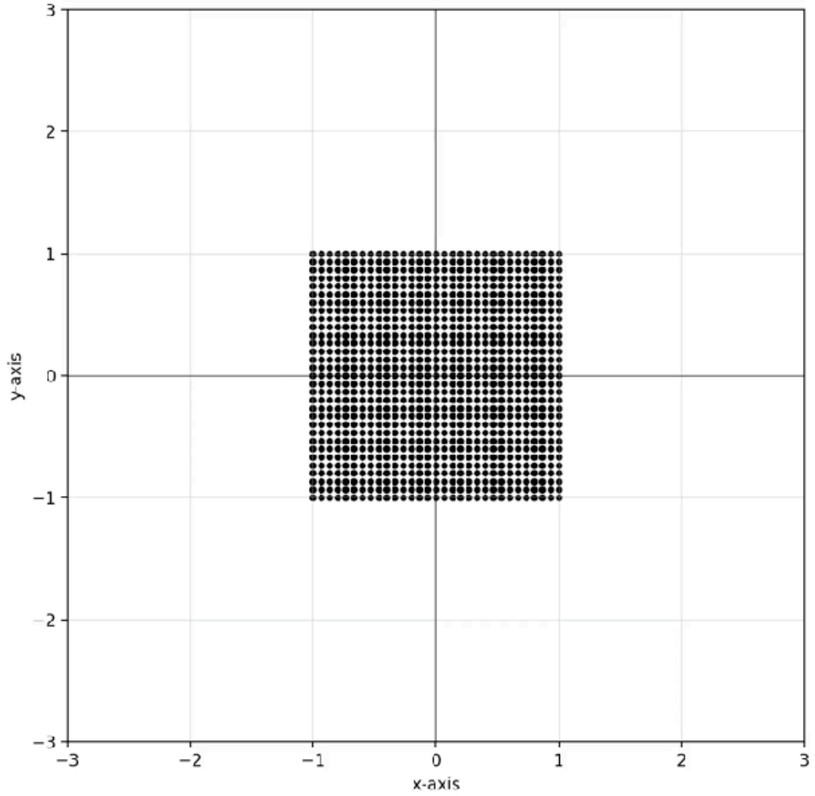
$$A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

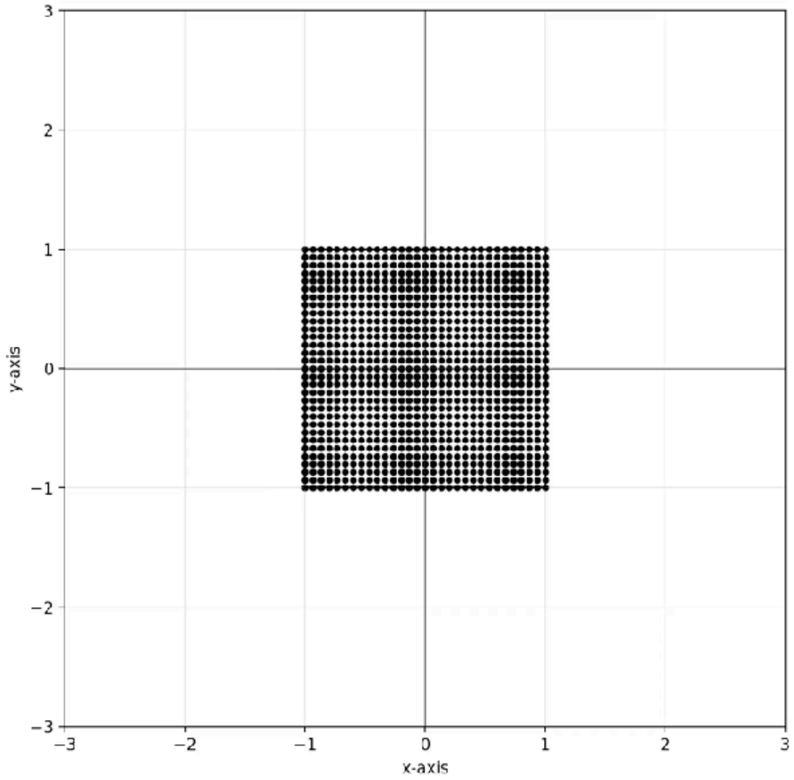


Commençons par des matrices simples avec beaucoup de 0

1. Matrices diagonales

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$





Scaling
(mise à l'échelle)

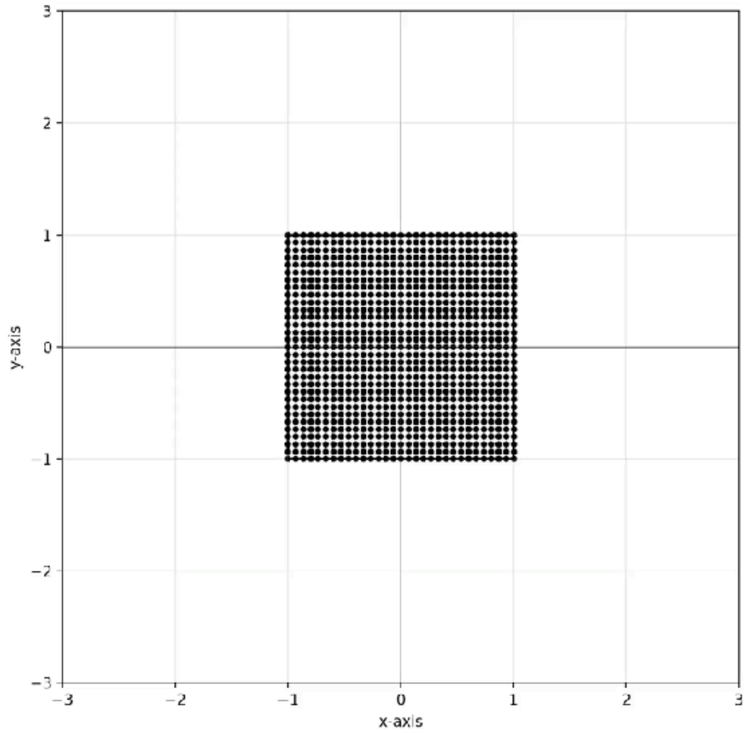
Existent-t-il des directions (vecteurs) inchangées par A ?

Oui, toutes les directions !

Commençons par des matrices simples avec beaucoup de 0

1. Matrices diagonales + élément négatif

$$A = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$



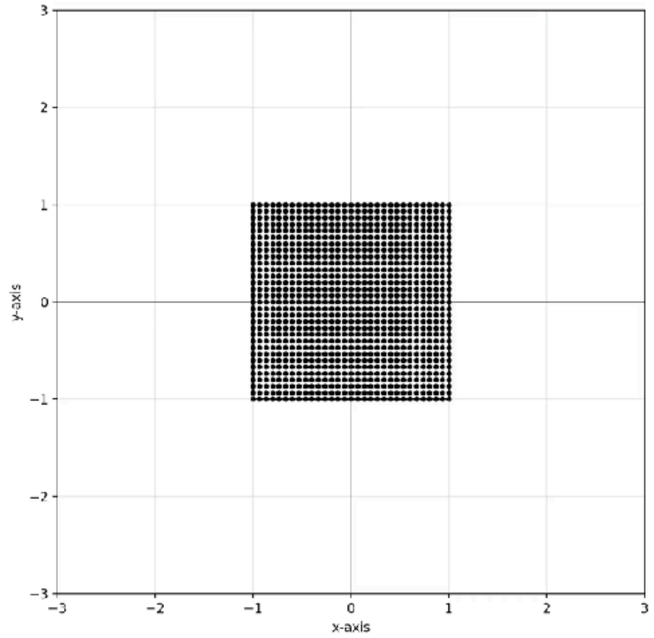
Effet de miroir par rapport à l'axe x:
On parle de réflection

Existent-t-il des directions (vecteurs) inchangées par A ?

Oui, toutes les directions !

Commençons par des matrices simples avec beaucoup de 0

2. Matrices à diagonales nulles $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$



réflection par rapport à la droite y = x

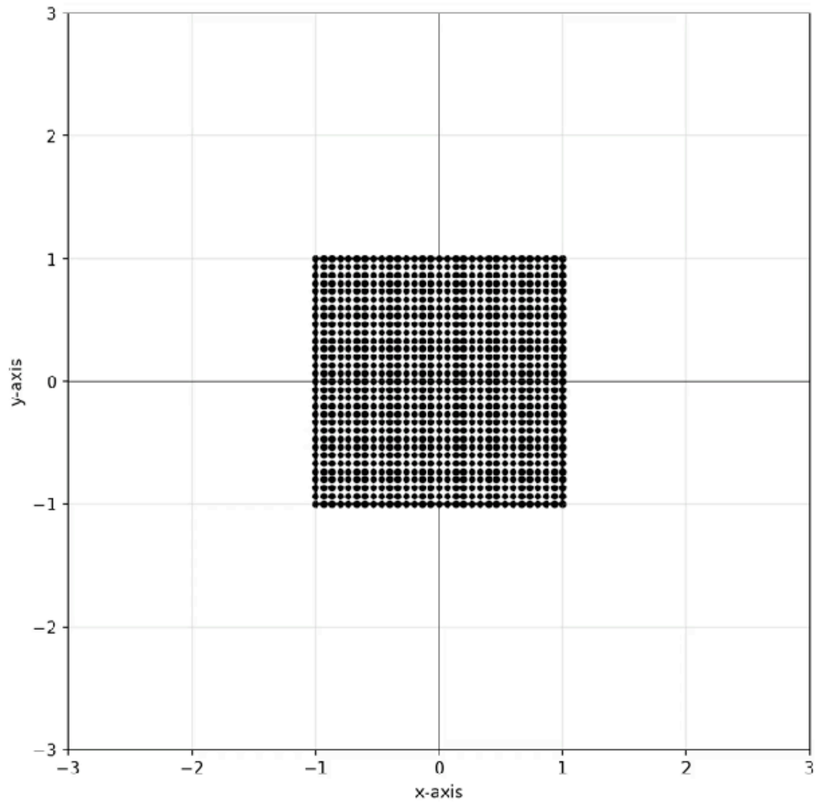
Existent-t-il des directions (vecteurs) inchangées par A ?

Oui, l'axe de réflection: y = x

Commençons par des matrices simples avec beaucoup de 0

3. Matrices à diagonales nulles + élément négatif

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$



rotation avec un angle -90°

Existent-t-il des directions (vecteurs) inchangées par A ?

Non !

Existent-t-il des directions (vecteurs) inchangées par A ?

1. Matrices diagonales $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ $\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$ Oui

2. Matrices à diagonales nulles $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ Oui

3. Matrices à diagonales nulles + élément négatif $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ Non

Matrices symétriques

$$A = A^\top$$

Matrices anti-symétriques

$$A = -A^\top$$

Toute matrice A peut être décomposée: $A = \overbrace{\frac{A + A^\top}{2}}^{\text{Partie symmétrique}} + \overbrace{\frac{A - A^\top}{2}}^{\text{Partie anti-symmétrique}}$

“stable” “instable”

INSEA

Existent-t-il des directions (vecteurs) inchangées par A ?

Formellement: $\exists x \in \mathbb{R}^n, \lambda \in \mathbb{R}$ tels que $Ax = \lambda x$?

vecteur propre valeur propre

Si $A = -A^\top$ Non !

Si $A = A^\top$ alors:

Le théorème spectral garantit l'existence de n valeurs propres λ_i associées à n vecteurs propres **orthogonaux** e_i de norme 1 qui vérifient pour tout $i = 1..n$:

$$Ae_i = \lambda_i e_i$$

?

On pose $\mathbf{P} = \begin{bmatrix} e_1 & \dots & e_n \end{bmatrix}$

1. Calculer $\mathbf{AP}$ et $\mathbf{P}^\top\mathbf{P}$.
2. En déduire $\mathbf{A}$ en fonction de $\mathbf{P}$ et des λ_i.

INSEA

$$\mathbf{AP} = \begin{bmatrix} \mathbf{A}\boldsymbol{e}_1 & \dots & \mathbf{A}\boldsymbol{e}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{e}_1 & \dots & \lambda_n\boldsymbol{e}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1 & \dots & \boldsymbol{e}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \lambda_2 & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix} = \mathbf{P}\Lambda$$

$$\text{Avec } \Lambda \overset{\text{def}}{=} \text{diag}(\lambda_1, \dots, \lambda_n)$$

Comme les $\boldsymbol{e_i}$ sont orthonormaux, on a $\boldsymbol{e_i}^\top \boldsymbol{e_j} = 1$ si $i = j$ et 0 sinon.

$$\mathbf{P}^\top\mathbf{P} = \begin{bmatrix} \boldsymbol{e_1}^\top \\ \vdots \\ \boldsymbol{e_n}^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{e_1} & \dots & \boldsymbol{e_n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{e_1}^\top\boldsymbol{e_1} & \dots & \boldsymbol{e_1}^\top\boldsymbol{e_n} \\ \vdots & \boldsymbol{e_i}^\top\boldsymbol{e_i} & \vdots \\ \boldsymbol{e_n}^\top\boldsymbol{e_1} & \dots & \boldsymbol{e_n}^\top\boldsymbol{e_n} \end{bmatrix} = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & 1 & \vdots \\ 0 & \dots & 1 \end{bmatrix} = \text{I}_n$$

Ainsi, $\mathbf{P}^\top = \mathbf{P}^{-1}$ et donc $\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}^\top$

P est dite: matrice orthogonale

? **Utilité du théorème spectral**

Écrire $\mathbf{A}$ comme une somme de matrices simples en utilisant le produit colonnes-lignes vu précédemment.

$$\mathbf{P}\Lambda\mathbf{P}^\top = \begin{bmatrix} \lambda_1 \boldsymbol{e}_1 & \dots & \lambda_n \boldsymbol{e}_n \end{bmatrix} \mathbf{P}^\top = \begin{bmatrix} \lambda_1 \boldsymbol{e}_1 & \dots & \lambda_n \boldsymbol{e}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_1{}^\top \\ \dots \\ \boldsymbol{e}_1{}^\top \end{bmatrix} = \sum_{i=1}^{n} \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i{}^\top$$

On choisit les indices i des $\boldsymbol{e}_i$ tels que $\lambda_1 \geq \lambda_2 .. \geq \lambda_n$.

Ainsi, la décomposition spectrale de $\mathbf{A}$ donne une somme pondérée de matrices de rang 1:

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i{}^\top$$

Théorème spectral: énoncé

Soit $\mathbf{A} \in \mathbb{S}_d$. Alors il existe une matrice orthogonale $\mathbf{P} \in \mathbb{R}^{d \times d}$ et une matrice diagonale Λ telles que:

$$\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}^\top$$

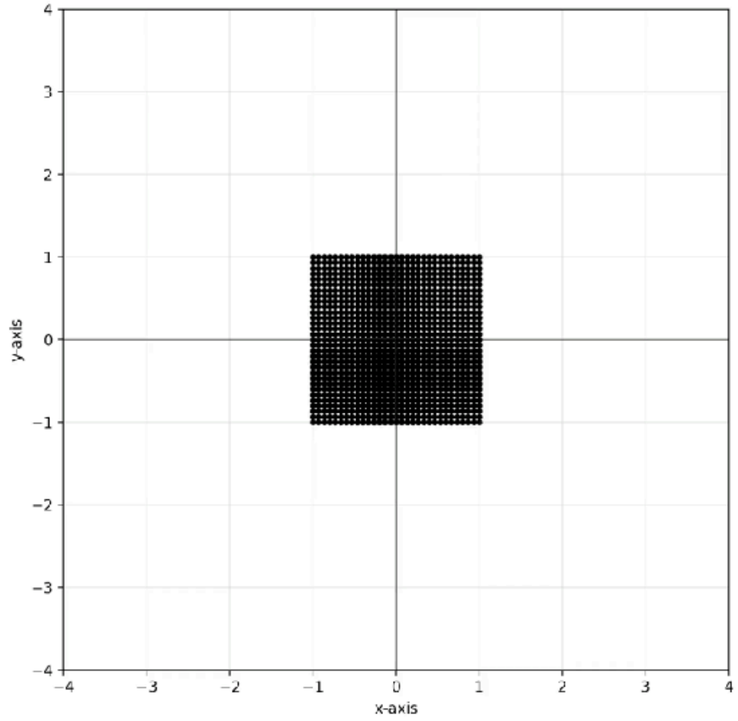
INSEA

Quelle est l'action géométrique d'une matrice **orthogonale** en 2D ?

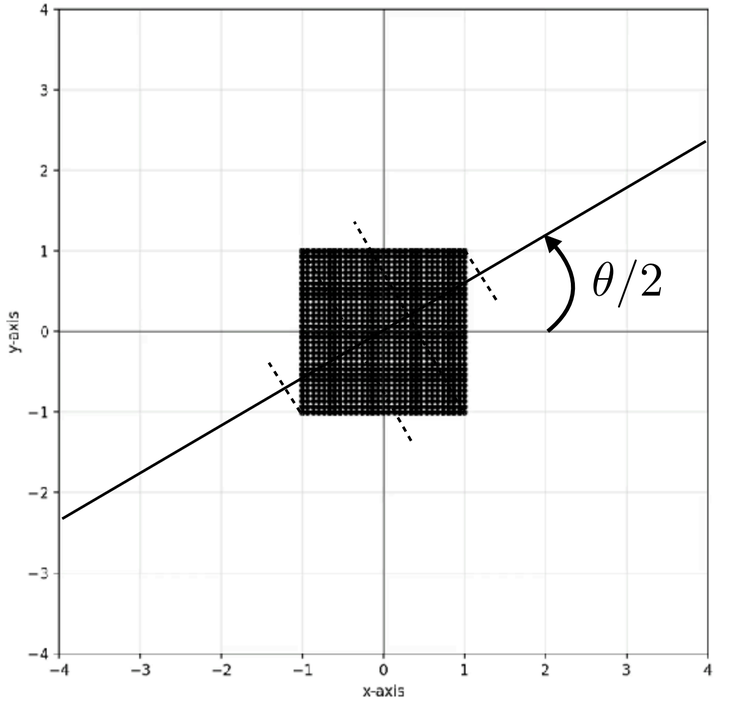
Avec les conditions d'orthogonalité et norme = 1, seuls deux formats sont possibles:

$$\mathbf{P} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \mathbf{P} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

Prenons $\theta = \pi/3 = 60$ degrés



Rotation d'angle θ



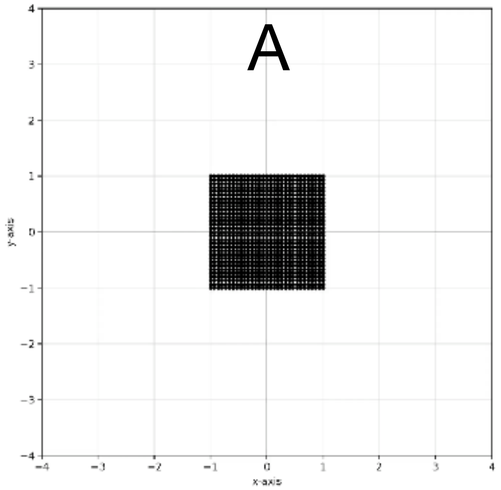
Réflection par rapport à l'axe $\theta/2$

Action d'une matrice symétrique via le théorème spectral:

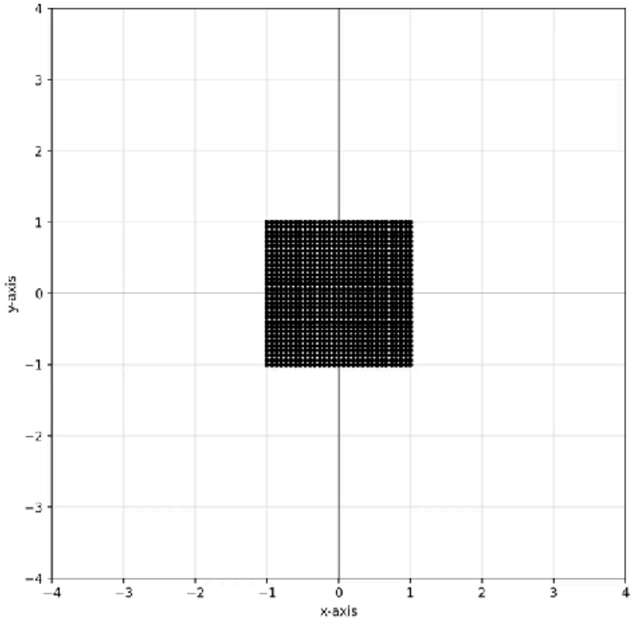
$$\mathbf{A} = \begin{bmatrix} 2.5 & -0.5 \\ -0.5 & 2.5 \end{bmatrix} = \mathbf{P}\Lambda\mathbf{P}^\top \quad \text{avec}$$

$$\mathbf{P} = \begin{bmatrix} 0.7 & -0.7 \\ 0.7 & 0.7 \end{bmatrix} \quad \text{et} \quad \Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

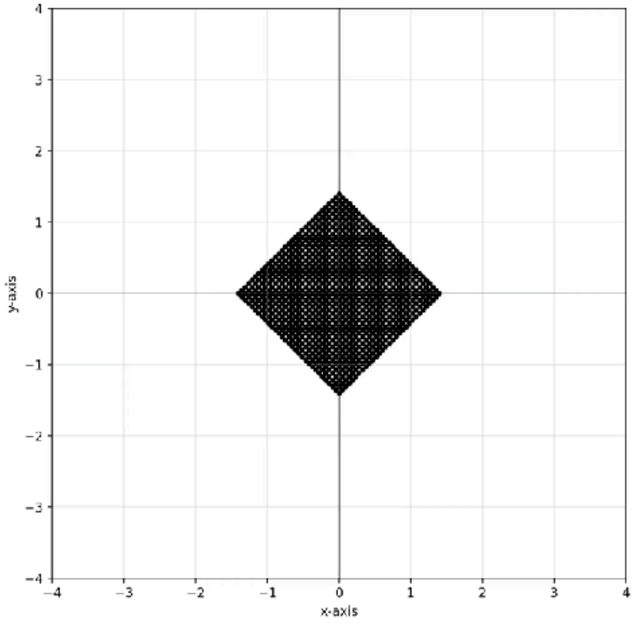
A



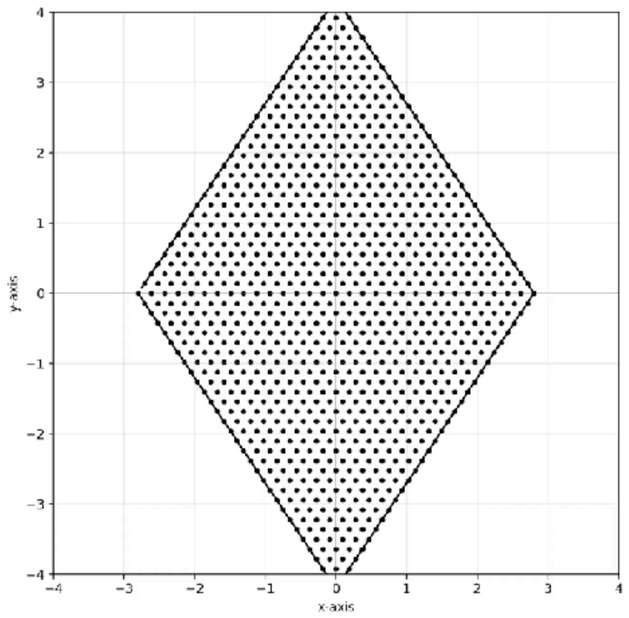
$\mathbf{P}^\top$



Λ



$\mathbf{P}$



INSEA

? **Comment peut-on la calculer en pratique ?**

Exercice: Fonction quadratique et quotient de Rayleigh

Soit $\mathbf{x} \in \mathbb{R}^n$. On pose $\varphi : \mathbf{x} \mapsto \mathbf{x}^\top \mathbf{A}\mathbf{x}$.

1. Calculez $\varphi(\boldsymbol{e_1})$ et $\varphi(\boldsymbol{e_n})$ en fonction des λ_i.
2. Évaluez $\varphi(\mathbf{x})$ et proposez un encadrement de $\frac{\mathbf{x}^\top \mathbf{A}\mathbf{x}}{\|\mathbf{x}\|^2}$ valable pour tout $\mathbf{x} \in \mathbb{R}^n$.
3. En déduire la solution de $\max_{\mathbf{x}} \frac{\mathbf{x}^\top \mathbf{A}\mathbf{x}}{\|\mathbf{x}\|^2}$ et $\min_{\mathbf{x}} \frac{\mathbf{x}^\top \mathbf{A}\mathbf{x}}{\|\mathbf{x}\|^2}$.

INSEA

$$\varphi(\boldsymbol{e_1}) = \boldsymbol{e_1}^\top \mathbf{A} \boldsymbol{e_1} = \sum_{i=1}^{n} \lambda_i \boldsymbol{e_1}^\top \boldsymbol{e_i} \boldsymbol{e_i}^\top \boldsymbol{e_1} = \sum_{i=1}^{n} \lambda_i (\boldsymbol{e_1}^\top \boldsymbol{e_i})(\boldsymbol{e_i}^\top \boldsymbol{e_1}) = \sum_{i=1}^{n} \lambda_i (\boldsymbol{e_1}^\top \boldsymbol{e_i})^2 = \lambda_1 (\boldsymbol{e_1}^\top \boldsymbol{e_1})^2 = \lambda_1 \|\boldsymbol{e_1}\|^2 = \lambda_1$$

$$\varphi(\boldsymbol{e_n}) = \boldsymbol{e_n}^\top \mathbf{A} \boldsymbol{e_n} = \sum_{i=1}^{n} \lambda_i \boldsymbol{e_n}^\top \boldsymbol{e_i} \boldsymbol{e_i}^\top \boldsymbol{e_n} = \sum_{i=1}^{n} \lambda_i (\boldsymbol{e_n}^\top \boldsymbol{e_i})(\boldsymbol{e_i}^\top \boldsymbol{e_n}) = \sum_{i=1}^{n} \lambda_i (\boldsymbol{e_n}^\top \boldsymbol{e_i})^2 = \lambda_n (\boldsymbol{e_n}^\top \boldsymbol{e_n})^2 = \lambda_n \|\boldsymbol{e_n}\|^2 = \lambda_n$$

$$\varphi(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} = \sum_{i=1}^{n} \lambda_i \mathbf{x}^\top \boldsymbol{e_i} \boldsymbol{e_i}^\top \mathbf{x} = \sum_{i=1}^{n} \lambda_i (\mathbf{x}^\top \boldsymbol{e_i})(\boldsymbol{e_i}^\top \mathbf{x}) = \sum_{i=1}^{n} \lambda_i (\mathbf{x}^\top \boldsymbol{e_i})^2$$

$$\min(\lambda_i) \sum_{i=1}^{n} (\mathbf{x}^\top \boldsymbol{e_i})^2 \leq \varphi(\mathbf{x}) \leq \max(\lambda_i) \sum_{i=1}^{n} (\mathbf{x}^\top \boldsymbol{e_i})^2$$

Or $$\sum_{i=1}^{n} (\mathbf{x}^\top \boldsymbol{e_i})^2 = \mathbf{x}^\top (\sum_{i=1}^{n} \boldsymbol{e_i} \boldsymbol{e_i}^\top) \mathbf{x} = \mathbf{x}^\top \mathbf{P} \mathbf{P}^\top \mathbf{x} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|^2$$

$$\lambda_n = \min(\lambda_i) \leq \frac{\varphi(\mathbf{x})}{\|\mathbf{x}\|^2} \leq \max(\lambda_i) = \lambda_1$$

Ces bornes sont atteintes avec $\mathbf{x} = \boldsymbol{e_1}$ et $\mathbf{x} = \boldsymbol{e_n}$.

INSEA

Pour $\mathbf{A} \in \mathbb{S}_n$ et $\mathbf{x} \in \mathbb{R}^n$:

$$\frac{\boldsymbol{e_n}^\top \mathbf{A} \boldsymbol{e_n}}{\|\boldsymbol{e_n}\|^2} = \lambda_n = \min(\lambda_i) \leq \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|^2} \leq \max(\lambda_i) = \lambda_1 = \frac{\boldsymbol{e_1}^\top \mathbf{A} \boldsymbol{e_1}}{\|\boldsymbol{e_1}\|^2}$$

On en déduit que:

1. On peut calculer $\boldsymbol{e_1}, \boldsymbol{e_n}, \lambda_1, \lambda_n$ en résolvant les problèmes d'optimisation $\min_{\mathbf{x}} \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|^2}$ et $\max_{\mathbf{x}} \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|^2}$.
2. $\forall \mathbf{x} \neq 0 \quad \mathbf{x}^\top \mathbf{A} \mathbf{x} > 0 \Leftrightarrow \min(\lambda_i) = \lambda_n > 0$. Dans ce cas, $\mathbf{A}$ est dite définie positive

?

Quels problèmes d'optimisation doit-on résoudre pour calculer $\boldsymbol{e_2}, \boldsymbol{e_{n-1}}, \lambda_2, \lambda_{n-1}$?

INSEA

Soit $\mathbf{A} \in \mathbb{S}_d$. Alors il existe une matrice orthogonale $\mathbf{P} \in \mathbb{R}^{d \times d}$ et une matrice diagonale Λ telles que:

$$\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}^\top$$

Puissance et racine matricielles

Et donc, comme $\mathbf{P}\mathbf{P}^\top = \mathrm{I}_d$, pour tout $q \in \mathbb{N}_*$, on a:

$$\mathbf{A}^q = \underbrace{\mathbf{P}\Lambda\mathbf{P}^\top\mathbf{P}\Lambda\mathbf{P}^\top \dots \mathbf{P}\Lambda\mathbf{P}^\top}_{q \text{ fois}} = \mathbf{P}\Lambda^q\mathbf{P}^\top$$

Si $\mathbf{A}$ est semi-définie positive, on définit la racine carrée matricielle:

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{P}\Lambda^{\frac{1}{2}}\mathbf{P}^\top$$

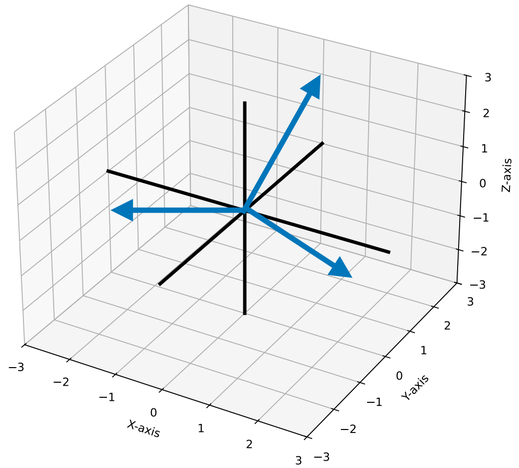
Où $\Lambda^{\frac{1}{2}} = \mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})$

INSEA

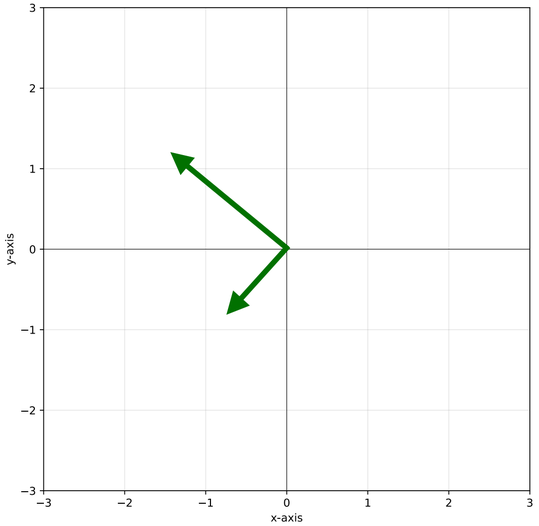
Toute $\mathbf{A} \in \mathbb{S}_d$ transforme une base orthogonale donnée par les colonnes de $\mathbf{P} \in \mathbb{R}^{d \times d}$ en elle-même avec un *rescaling* de ses amplitudes: $\mathbf{AP} = \mathbf{P}\Lambda$

Qu'en est-il d'une matrice quelconque ?

Soit $\mathbf{A} \in \mathbb{R}^{n \times d}$.



$\mathbf{A}$

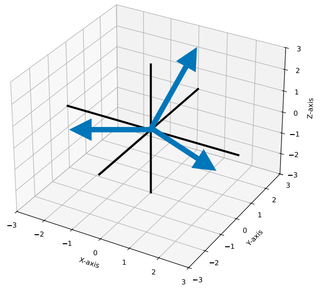


Oui !

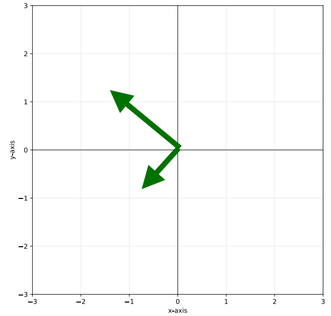
$\mathbb{R}^d$ $\mathbb{R}^n$

Existe-t-il une base orthonormée $\mathbf{V}$ dans $\mathbb{R}^d$ transformée par $\mathbf{A}$ en une base orthonormée $\mathbf{U}$ dans $\mathbb{R}^n$?

Soit $\mathbf{A} \in \mathbb{R}^{n \times d}$.



$\mathbf{A}$

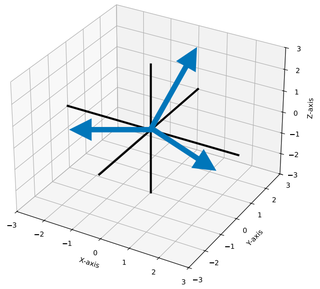


$\mathbb{R}^d$ $\mathbb{R}^n$

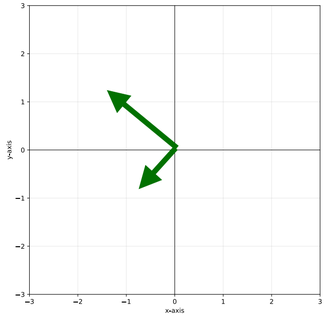
Il existe deux matrices orthogonales $\mathbf{V} \in \mathbb{R}^{d \times d}$ et $\mathbf{U} \in \mathbb{R}^{n \times n}$ et une matrice **rectangulaire** diagonale $\Sigma \in \mathbb{R}^{n \times d}$ telles que:

$$\mathbf{AV} = \mathbf{U}\Sigma$$

Soit $\mathbf{A} \in \mathbb{R}^{n \times d}$.



$\mathbf{A}$



$\mathbb{R}^d$ $\mathbb{R}^n$

Il existe deux matrices orthogonales $\mathbf{V} \in \mathbb{R}^{d \times d}$ et $\mathbf{U} \in \mathbb{R}^{n \times n}$ et une matrice **rectangulaire** diagonale $\Sigma \in \mathbb{R}^{n \times d}$ telles que:

$$\mathbf{AV} = \mathbf{U}\Sigma$$

si $d > n$: $$\mathbf{AV} = \begin{bmatrix} \mathbf{A}\boldsymbol{v}_1 & \dots & \mathbf{A}\boldsymbol{v}_d \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_n \end{bmatrix} \begin{bmatrix} \sigma_1 & \dots & 0 & \dots & 0 \\ \vdots & \sigma_2 & \vdots & 0 & \vdots \\ 0 & \dots & \sigma_n & \dots & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1\boldsymbol{u}_1 & \dots & \sigma_n\boldsymbol{u}_n & 0 & \dots & 0 \end{bmatrix}$$

si $n > d$: $$\mathbf{AV} = \begin{bmatrix} \mathbf{A}\boldsymbol{v}_1 & \dots & \mathbf{A}\boldsymbol{v}_d \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_n \end{bmatrix} \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \sigma_2 & \vdots \\ 0 & \dots & \sigma_d \\ \vdots & 0 & \vdots \end{bmatrix} = \begin{bmatrix} \sigma_1\boldsymbol{u}_1 & \dots & \sigma_d\boldsymbol{u}_d \end{bmatrix}$$

INSEA

SVD

Soit $\mathbf{A} \in \mathbb{R}^{n \times d}$.

Il existe deux matrices orthogonales $\mathbf{V} \in \mathbb{R}^{d \times d}$ et $\mathbf{U} \in \mathbb{R}^{n \times n}$ et une matrice **rectangulaire** diagonale $\Sigma \in \mathbb{R}^{n \times d}$ telles que:

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^\top$$

Si $\text{rang}(\mathbf{A}) = r$, il existe r triplets $(\sigma_i, \mathbf{v}_i, \mathbf{u}_i)$ tels que: $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i \qquad \sigma_i > 0$

σ_i: valeur singulière de $\mathbf{A}$ $\qquad$ $\mathbf{v}_i, \mathbf{u}_i$: vecteurs singuliers de $\mathbf{A}$

Alors on peut obtenir une forme réduite de la SVD avec $\mathbf{U} \in \mathbb{R}^{n \times r}$, $\mathbf{V} \in \mathbb{R}^{d \times r}$ et Σ diagonale et inversible de taille r:

$$\mathbf{A} = \mathbf{U}_r \Sigma_r \mathbf{V}_r^\top$$

$$\mathbf{U}_r = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_r \end{bmatrix} \qquad \Sigma_r = \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \sigma_2 & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix} \qquad \mathbf{V}_r = \begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{v}_r \end{bmatrix}$$

INSEA

$$\mathbf{A} = \mathbf{U}_r \Sigma_r \mathbf{V}_r^\top$$

$$\mathbf{U}_r = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_r \end{bmatrix} \qquad \Sigma_r = \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \sigma_2 & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix} \qquad \mathbf{V}_r = \begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{v}_r \end{bmatrix}$$

Ainsi, toute matrice $\mathbf{A}$ s'écrit comme une somme pondérée de matrices de rang 1:

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

?

On suppose que le rang est très petit par rapport à n et d. Pouvez-vous penser à des applications directes des résultats ci-dessus ?

1. Comparez la taille en mémoire occupée par A et par sa décomposition singulière.

2. Comparez la complexité du produit matriciel Ax avec ou sans SVD.

INSEA