

Surf sur représentations mathématiques

Copyrights divers

01/10/2020

Pour se “rafraichir” les idées : un petit surf d'échauffement !

Attention : Ceci n'est pas un support de cours (un cours rédigé est par ailleurs à votre disposition), c'est une collection de sollicitations visuelles pour solliciter votre attention aux mathématiques du machine learning.

De nombreuses illustrations de ce document sont extraites de l'excellent ouvrage :

Mathematics for machine learning

Marc Peter Deisenroth

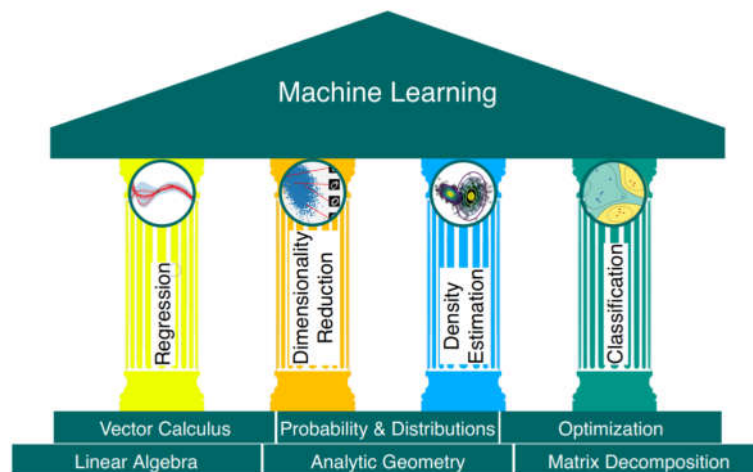
A. Aldo Faisal

Chen Soon Ong

Quelques rares illustrations sont extraites de l'excellent cours : Christophe Bertault - Mathématiques en MPSI

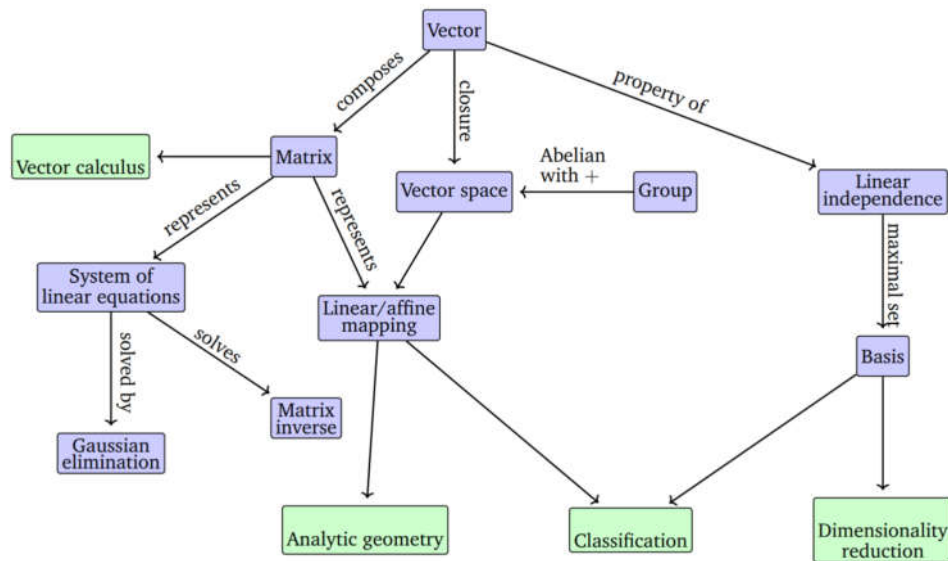
Introduction

Introduction - WHY ?



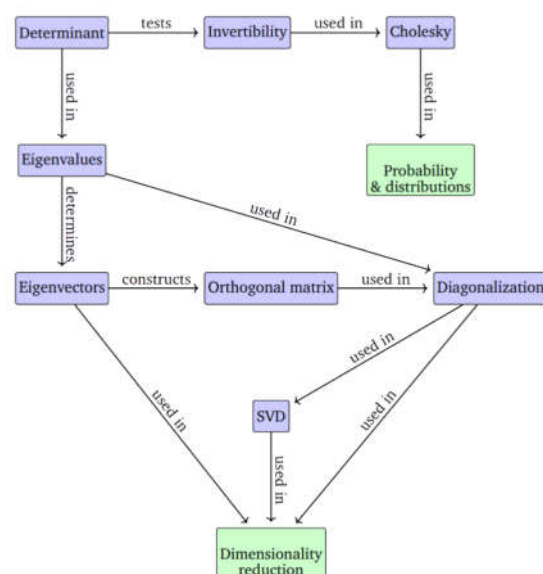
Contributions mathématiques au machine learning

Quelques thèmes naturellement reliés aux vecteurs et au ML



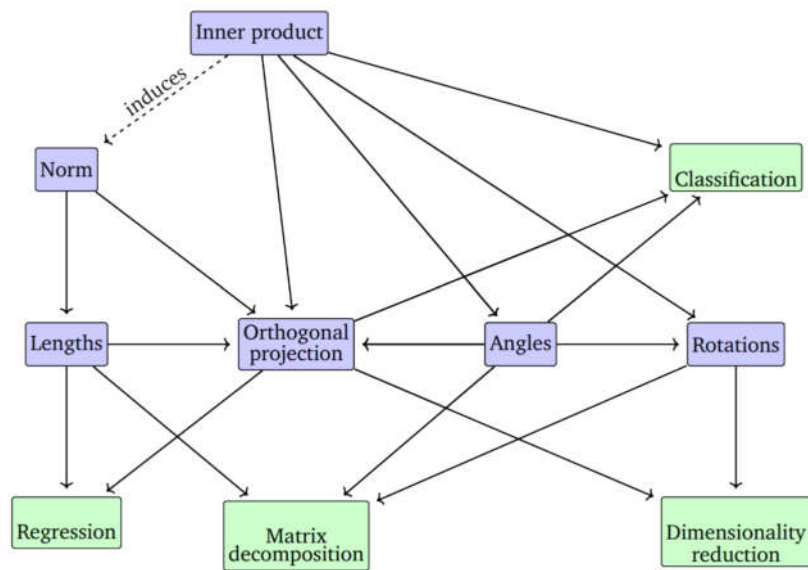
Autours des vecteurs

Quelques thèmes naturellement reliés aux déterminant/matrices et au ML



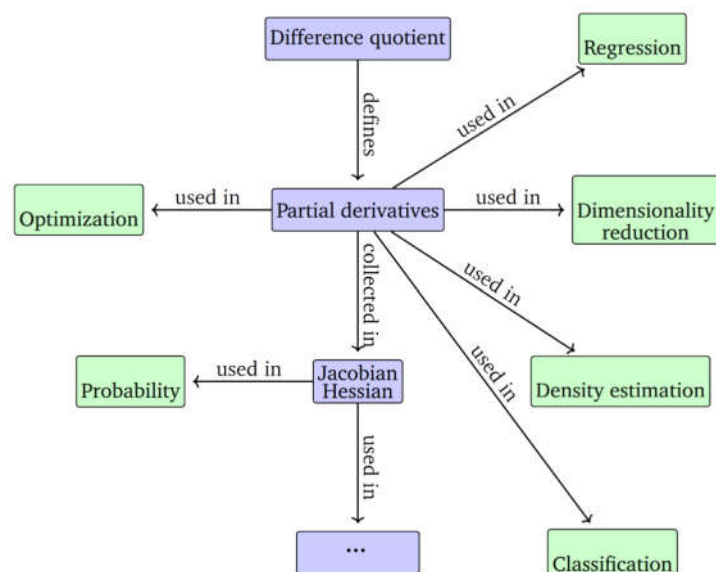
Autours des déterminants/matrices

Quelques thèmes naturellement reliés aux produits scalaires et au ML



Autours du produit scalaire

Quelques thèmes naturellement reliés aux dérivée partielles et au ML

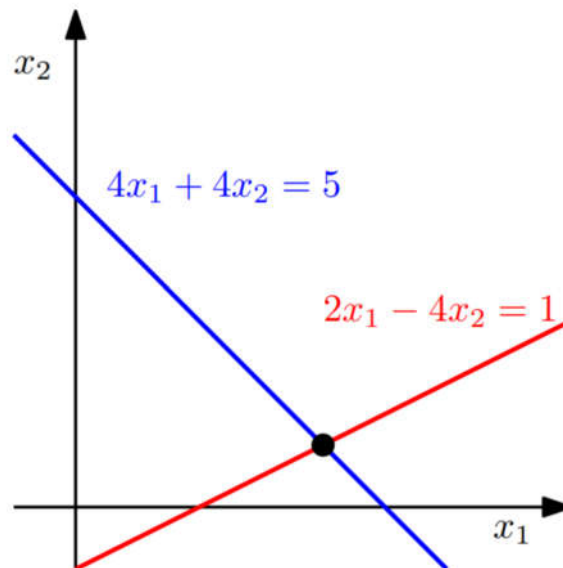


Autours des dérivées partielles

Linéarité - bases

Equations linéaires

Il n'existe pas toujours une ou plusieurs solutions



Système d'équations linéaires

Formulation sous forme de matrices

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Système d'équations linéaires généralisé

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

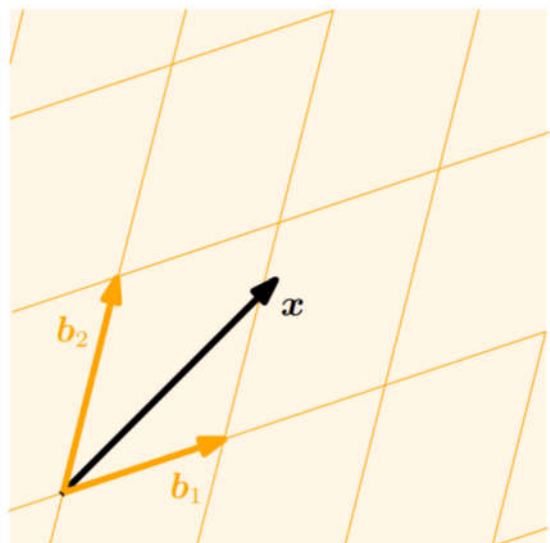
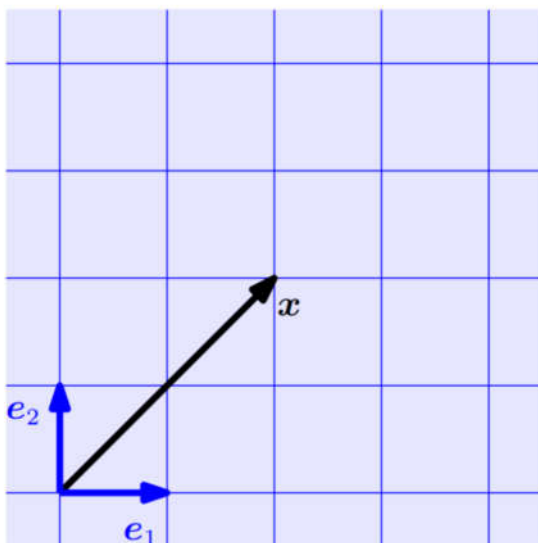
Forme matricielle

$$Ax = b$$

Matrices

Remarques basiques

Bien identifier les bases dans lesquelles on travaille



Un même vecteur dans 2 bases

Faire attention aux dimensions

$$\text{avec } \mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \in \mathbb{R}^{2 \times 3}, \mathbf{B} = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{3 \times 2}$$

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} \in \mathbb{R}^{2 \times 2}$$

$$\mathbf{BA} = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 4 & 2 \\ -2 & 0 & 2 \\ 3 & 2 & 1 \end{bmatrix} \in \mathbb{R}^{3 \times 3}$$

Multiplication de matrices

Une matrice identité est une matrice carrée

$$\mathbf{I}_n := \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Identité

On a :

$$I_1 = (1) , \quad I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I_n = (\delta_{i,j})_{1 \leq i \leq n, 1 \leq j \leq n}$$

avec :

$$\delta_{i,j}$$

qui désigne le *symbole de Kronecker*

Associativité, distributivité, pas commutativité

$$\forall \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times p}, \mathbf{C} \in \mathbb{R}^{p \times q} : (\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

$$\begin{aligned} \forall \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}, \mathbf{C}, \mathbf{D} \in \mathbb{R}^{n \times p} : (\mathbf{A} + \mathbf{B})\mathbf{C} &= \mathbf{AC} + \mathbf{BC} \\ \mathbf{A}(\mathbf{C} + \mathbf{D}) &= \mathbf{AC} + \mathbf{AD} \end{aligned}$$

$$\forall \mathbf{A} \in \mathbb{R}^{m \times n} : \mathbf{I}_m \mathbf{A} = \mathbf{A} \mathbf{I}_n = \mathbf{A}$$

Propriétés de base

L'inverse d'une matrice n'existe pas toujours

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 4 & 4 & 5 \\ 6 & 7 & 7 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -7 & -7 & 6 \\ 2 & 1 & -1 \\ 4 & 5 & -4 \end{bmatrix}$$
$$\mathbf{AB} = \mathbf{I} = \mathbf{BA}.$$

les matrices carrées A et B sont inverses l'une de l'autre

Produit scalaire

Forme bilinéaire

De façon un peu “approximative”

Notre ami Wikipédia nous dit :

En mathématiques, plus précisément en algèbre linéaire, une **forme bilinéaire** est un type particulier d'application qui, à deux vecteurs d'un espace vectoriel (sur un certain corps commutatif), associe un scalaire (c'est-à-dire un élément de ce corps).

Certaines formes bilinéaires sont de plus des **produits scalaires**.

Les produits scalaires (sur les espaces vectoriels de dimension finie ou infinie) sont très utilisés, dans toutes les branches mathématiques, pour définir une **distance**.

... Puis :

Une forme désigne en mathématiques une application d'un espace vectoriel dans son corps des scalaires.

Matrice d'une forme bilinéaire

Avec e_i les vecteurs de la base, f fonction de deux vecteurs et à valeur dans son corps (typiquement, les réels pour les cas simples).

$$\text{mat } f = F = (f(e_i, e_j))_{1 \leq i \leq n, 1 \leq j \leq n} = \begin{pmatrix} f(e_1, e_1) & f(e_1, e_2) & \dots & f(e_1, e_n) \\ f(e_2, e_1) & f(e_2, e_2) & \dots & f(e_2, e_n) \\ \vdots & \vdots & \ddots & \vdots \\ f(e_n, e_1) & f(e_n, e_2) & \dots & f(e_n, e_n) \end{pmatrix}$$
$$f(x, y) = {}^t\mathbf{x} F \mathbf{y}$$

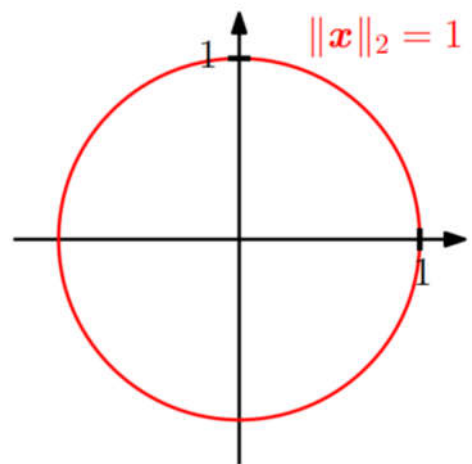
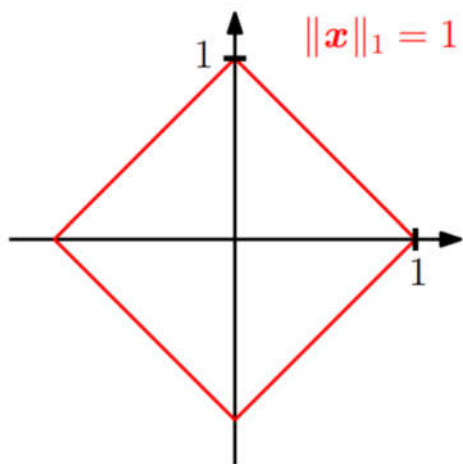
Si toutes les expressions sont dans les réels et que f est l'identité on se trouve dans le cas courant du produit scalaire euclidien (les anglais le nomme dot).

$$\mathbf{x} \cdot \mathbf{y} = {}^t\mathbf{x} I_n \mathbf{y} = {}^t\mathbf{x} \mathbf{y}$$

Pour écrire “transposée”, s'il n'y a pas de confusion avec une puissance, il est plus commode d'écrire “à l'américaine” :

$$\mathbf{x}^T$$

Normes et distance



Influence de la norme sur la topologie

Norme issue d'un produit scalaire

$$\|x\| = \sqrt{\langle x|x \rangle}$$

Distance issue d'une norme

$$d(x, y) = \|y - x\|$$

Zoologie de matrices

Inversion et transposition ne fonctionnent pas de la même façon

$$\begin{aligned}
AA^{-1} &= I = A^{-1}A \\
(AB)^{-1} &= B^{-1}A^{-1} \\
(A+B)^{-1} &\neq A^{-1} + B^{-1} \\
(A^{\top})^{\top} &= A \\
(A+B)^{\top} &= A^{\top} + B^{\top} \\
(AB)^{\top} &= B^{\top}A^{\top}
\end{aligned}$$

Manipulations basiques

Rq: On n'inverse que les matrices carrées, mais on peut transposer toutes les matrices.

Trace d'une matrice

$$\text{tr}(A) := \sum_{i=1}^n a_{ii},$$

- $\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$ for $A, B \in \mathbb{R}^{n \times n}$
- $\text{tr}(\alpha A) = \alpha \text{tr}(A)$, $\alpha \in \mathbb{R}$ for $A \in \mathbb{R}^{n \times n}$
- $\text{tr}(I_n) = n$
- $\text{tr}(AB) = \text{tr}(BA)$ for $A \in \mathbb{R}^{n \times k}$, $B \in \mathbb{R}^{k \times n}$

$$\text{▪ } A \in \mathbb{R}^{a \times k}, K \in \mathbb{R}^{k \times l}, L \in \mathbb{R}^{l \times a} \quad \text{tr}(AKL) = \text{tr}(KLA)$$

$$\text{▪ } x, y \in \mathbb{R}^n \quad \text{tr}(xy^{\top}) = \text{tr}(y^{\top}x) = y^{\top}x \in \mathbb{R}$$

$$\begin{aligned}
&\Phi : V \rightarrow V \quad B = S^{-1}AS \\
&\text{▪ } \text{tr}(B) = \text{tr}(S^{-1}AS) = \text{tr}(ASS^{-1}) = \text{tr}(A)
\end{aligned}$$

Autour de la trace d'une matrice

Matrices de transformations linéaires

linéarité

Linéarité d'une transformation Φ (on aurait pu l'appeler f !!).

En anglais on parle de *linear map*.

$$\Phi(x + y) = \Phi(x) + \Phi(y)$$

Multiplication par un scalaire λ :

$$\Phi(\lambda x) = \lambda \Phi(x)$$

Remarque IMPORTANTE : Les dimensions de l'espace vectoriel de départ et d'arrivée sont ici quelconques.

Si l'espace d'arrivée est le corps des scalaires, sa dimension est "1" et on dit que l'on a une forme linéaire.

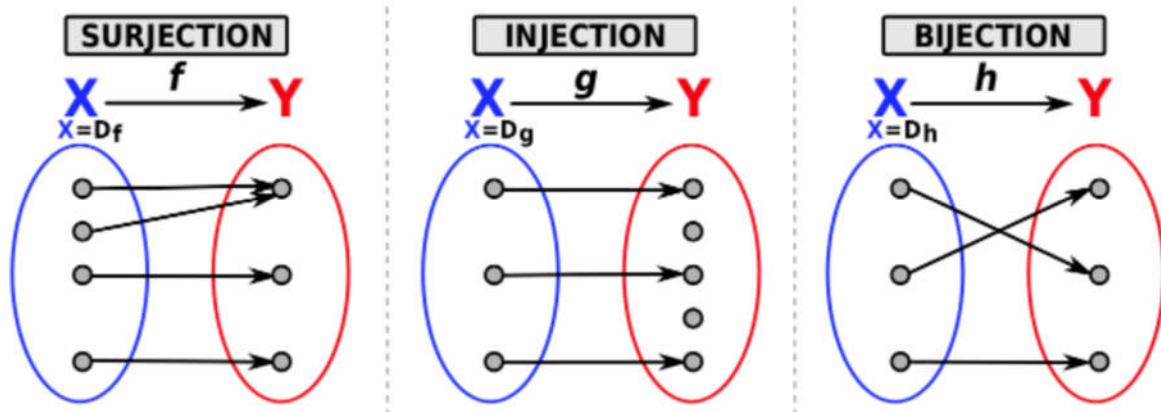
Matrice d'une transformation linéaire

$$\Phi : \mathbb{R}^4 \rightarrow \mathbb{R}^2, \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 - x_3 \\ x_1 + x_4 \end{bmatrix}$$

Une transformation linéaire

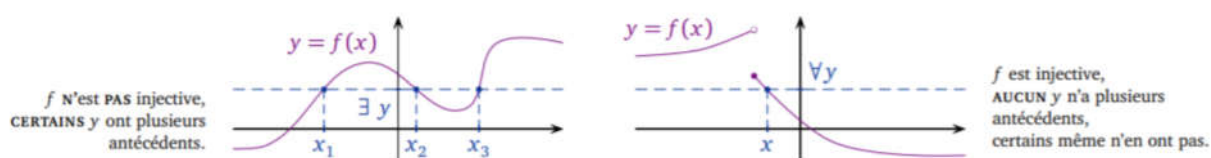
injection, surjection, bijections : rappel

Les bases



Autour de la bijection

traduction pour une fonction (pas forcément linéaire !)

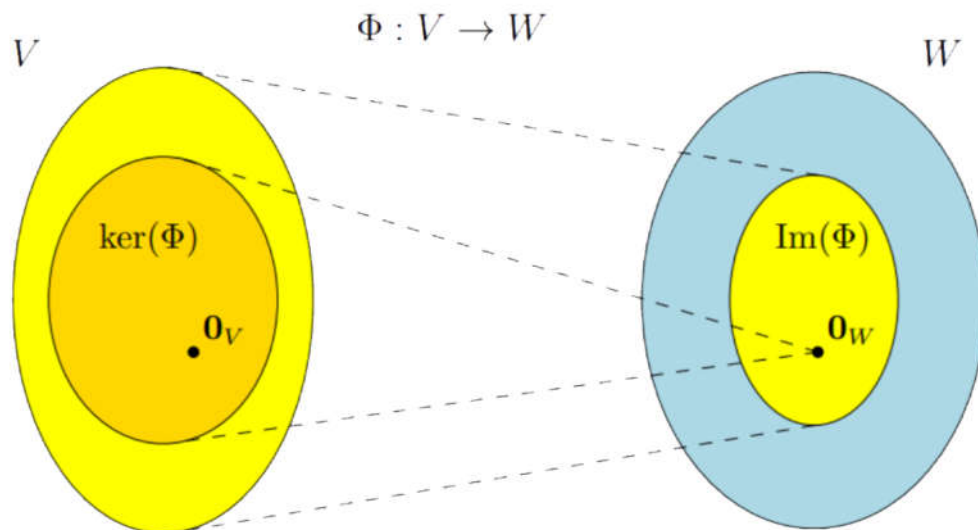


Autour de la bijection

Noyau (kernel) et image

Quand la tranformation n'est pas bijective il est possible que :

- l'image de l'espace vectoriel de départ puisse ne pas être l'ensemble de l'espace vectoriel d'arrivé,
- Il puisse y avoir plusieurs antécédents à un vecteur de l'espace d'arrivé.



Kernel et image

A méditer sur les noyaux et les images

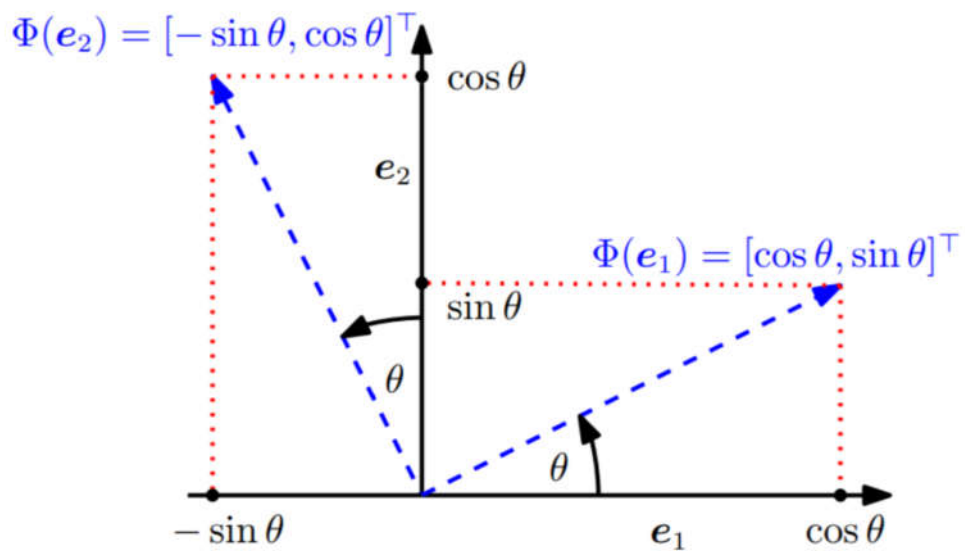
On appelle **rang** (*rank* en anglais) de cette transformation linéaire (et de sa matrice associée A) la dimension de son image : $\dim(\text{Im}(\Phi))$.

Le noyau est tout simplement la solution de $A\mathbf{x} = \mathbf{0}$.

Φ est **injective** ssi son noyau est réduit au vecteur nul de l'espace de départ $\ker(\Phi) = \{0_W\}$ Φ est **injective** ssi $\dim(\text{Im}(\Phi))$ est égal à la dimension de l'espace **de départ**.

Sous l'angle géométrique

Matrice de rotation et généralisation

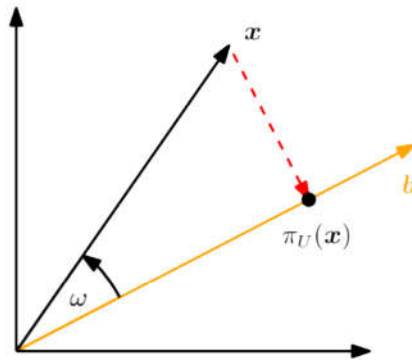


Une rotation dans le plan

$$R_{ij}(\theta) := \begin{bmatrix} I_{i-1} & 0 & \dots & \dots & 0 \\ 0 & \cos \theta & 0 & -\sin \theta & 0 \\ 0 & 0 & I_{j-i-1} & 0 & 0 \\ 0 & \sin \theta & 0 & \cos \theta & 0 \\ 0 & \dots & \dots & 0 & I_{n-j} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Given rotation

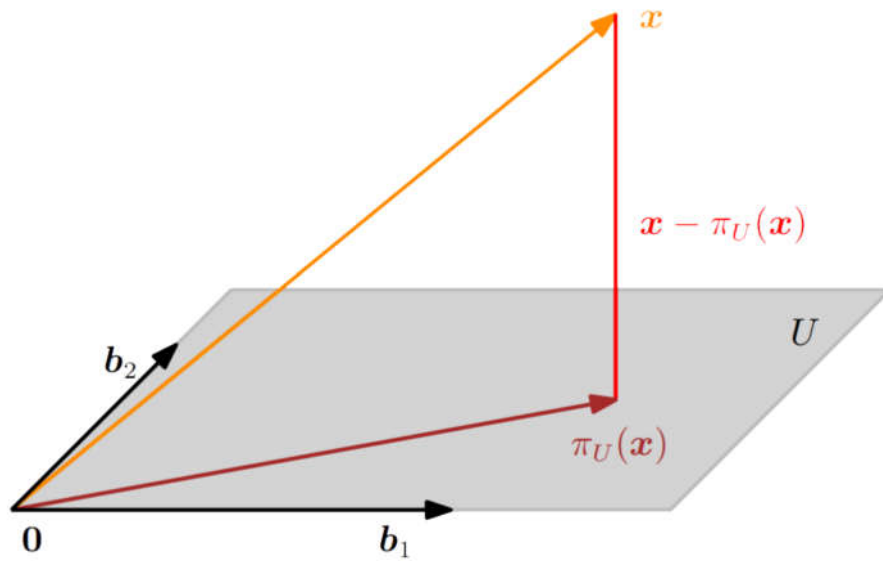
Projection et symétrie



$$\langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0$$

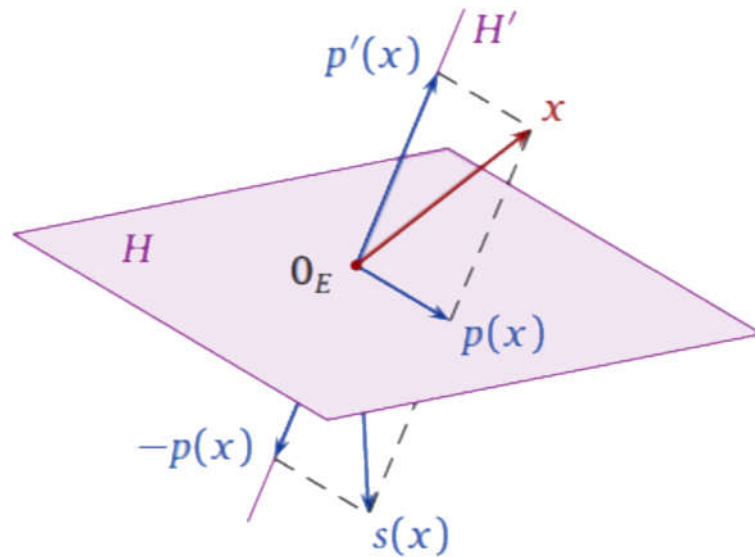
$$\pi_U(\mathbf{x}) = \lambda \mathbf{b}, \quad \lambda \in \mathbb{R}.$$

Une projection



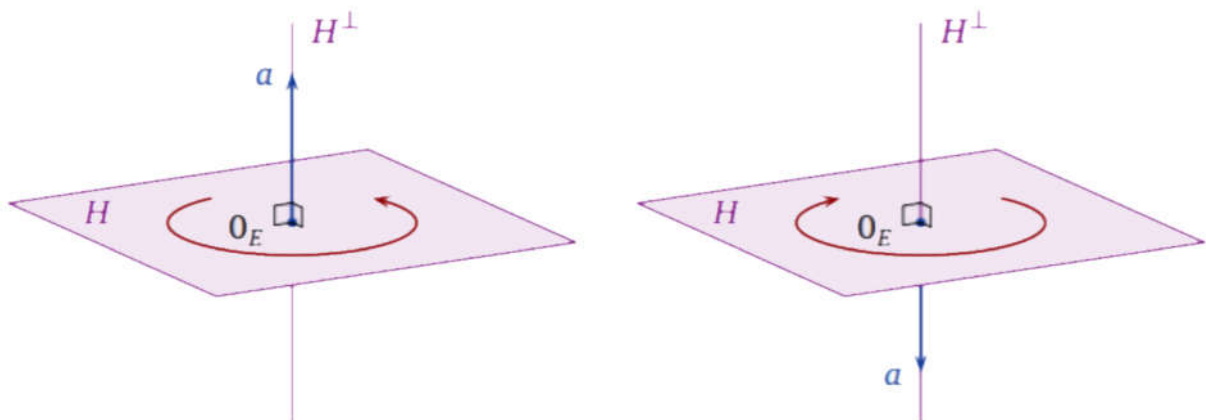
Projection orthogonale sur un plan

Projection sur le plan et sur son espace orthogonal et symétrie :



projection et symétrie (ici orthogonales)

Orientation d'un hyperplan



Deux orientations possibles

Relation entre les types de matrices

Matrice symétrique : matrice de réels égale à sa transposée

Matrice adjointe

$$M^* = {}^t \overline{M} = \overline{{}^t M}$$

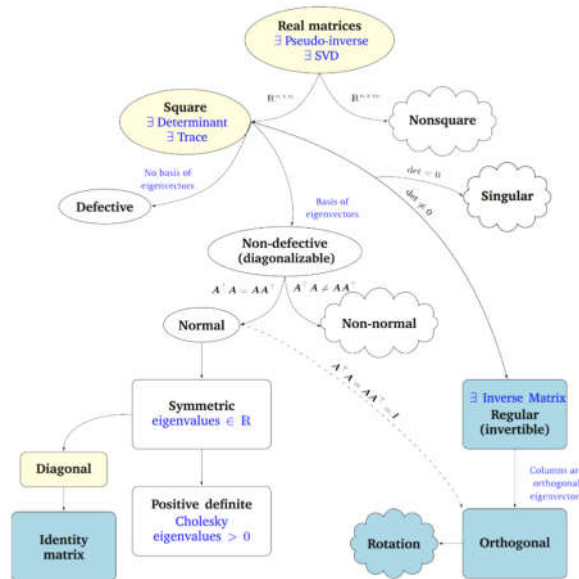
Matrice hermitienne : matrice de complexes égale à son adjointe

Matrice normale : commute avec sa matrice adjointe (et donc transposée si réelle)

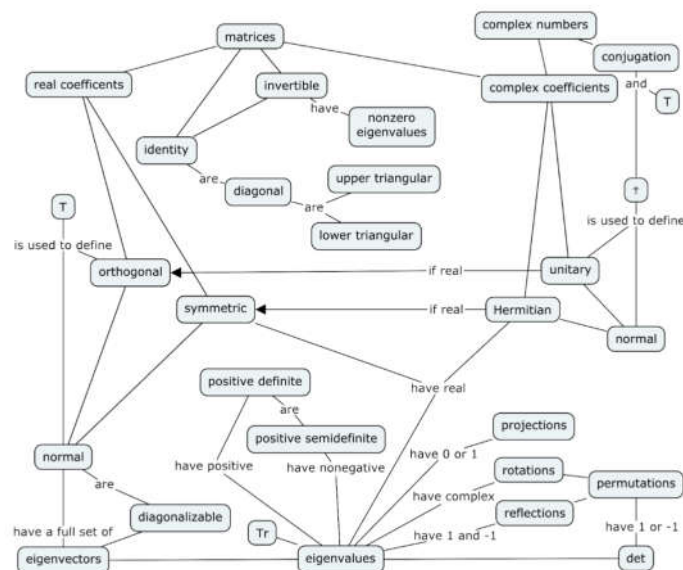
Matrice (...) définie positive : matrice symétrique (ou hermitienne) avec valeurs propres strictement positives.

Inverse généralisé d'une matrice : $A A^+ A = A$ et $A^+ A A^+ = A^+$

Pseudo inverse de Penrose : inverse généralisée avec 2 autre conditions



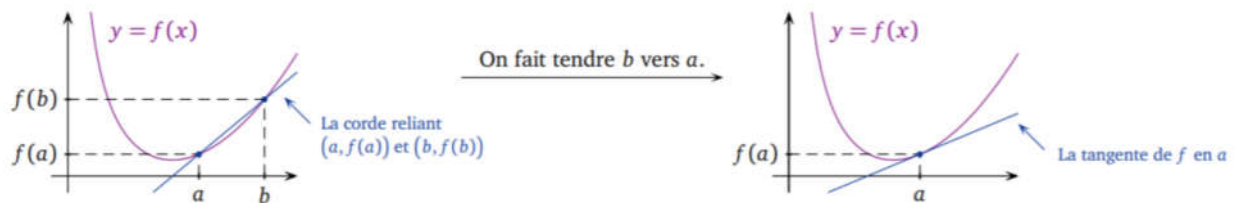
relation entre les types de matrices réelles



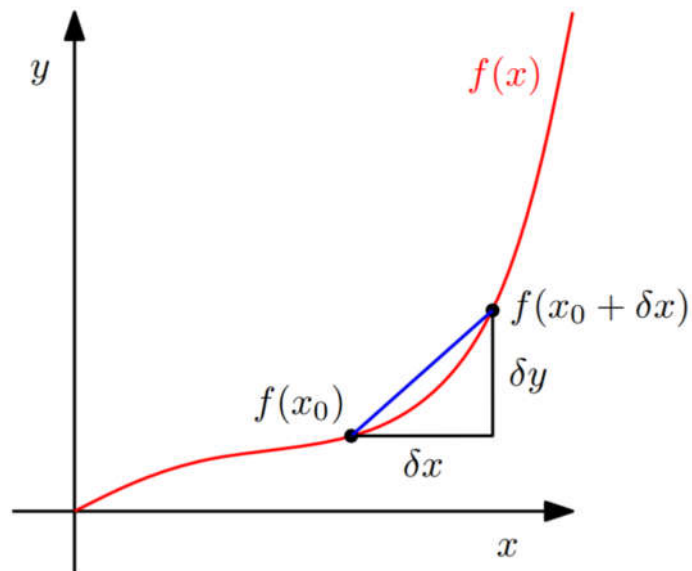
Yet Other Relations ... entre les types de matrices carrées

Autres tips

Dérivée



En route vers la dérivation



dérivée, vers une limite

Différentiation : la règle au centre du ML

$$\frac{\partial(y_1, \dots, y_k)}{\partial(x_1, \dots, x_n)} = \frac{\partial(y_1, \dots, y_k)}{\partial(u_1, \dots, u_m)} \frac{\partial(u_1, \dots, u_m)}{\partial(x_1, \dots, x_n)}$$

$$f(x, y) = (x + 2y^3)^2$$

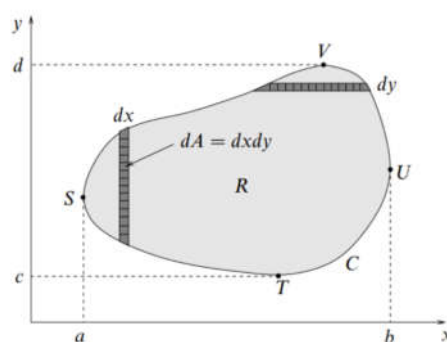
$$\frac{\partial f(x, y)}{\partial x} = 2(x + 2y^3) \frac{\partial}{\partial x}(x + 2y^3) = 2(x + 2y^3)$$

$$\frac{\partial f(x, y)}{\partial y} = 2(x + 2y^3) \frac{\partial}{\partial y}(x + 2y^3) = 12(x + 2y^3)y^2$$

chain rule

Intégration sur un domaine

$$I = \int_R f(x, y) dA,$$

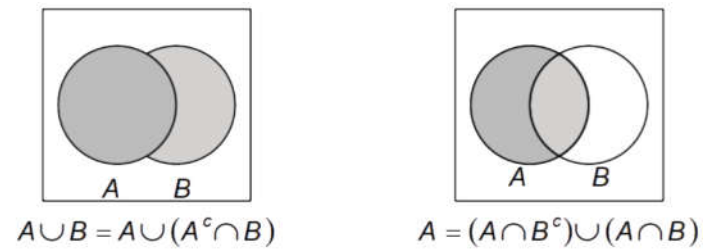


$$I = \iint_R f(x, y) dx dy,$$

Une intégrale double

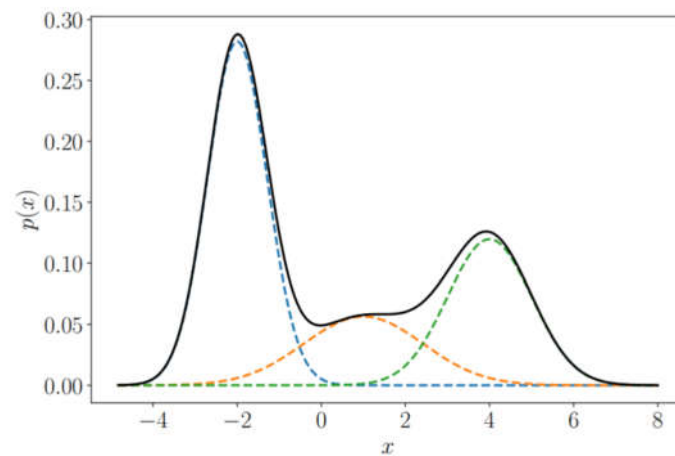
petite réflexion ensembliste/probabilité

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$



Raisonnement de base sur les ensembles

Mixture gaussienne



$$p(x | \theta) = 0.5\mathcal{N}(x | -2, \frac{1}{2}) + 0.2\mathcal{N}(x | 1, 2) + 0.3\mathcal{N}(x | 4, 1)$$

Mixture - différentes populations