Outils Mathématiques pour l'Intelligence Artificielle

Les Fondamentaux

Dr HONTINFINDE Valérie Ing MAMLANKOU Charbel

École d'Été pour l'Intelligence Artificielle (EEIA)

Édition 2025

Plan

1-1 Définitions

Fonction numérique

Soient A et B deux sous-ensembles non vides de $\mathbb{R}$. On appelle fonction numérique de la variable réelle, notée f, toute relation qui à un élément $x \in A$ associe **au plus** une image $f(x) \in B$.

$$\begin{aligned} f : A &\to B \\ x &\mapsto f(x) \end{aligned}$$

A est l'ensemble de **départ**, B est l'ensemble d'**arrivée**.

Application et Ensemble de Définition

On appelle **ensemble de définition** de f, noté D_f, l'ensemble des réels pour lesquels $f(x)$ existe.

$$D_f = \{x \in A \mid f(x) \text{ existe}\}$$

Une fonction f est une **application** si son ensemble de départ A est égal à son ensemble de définition D_f.

Lien avec l'IA

En IA, les **fonctions de coût** (ou *loss functions*) sont des fonctions qui mesurent l'erreur d'un modèle. Notre but est de minimiser cette fonction. Par exemple, l'erreur quadratique moyenne $L(y, \hat{y}) = (y - \hat{y})^2$ est une fonction de la prédiction $\hat{y}$.

Exercices : Fonctions et Applications

À vous de jouer !

Pour chacune des fonctions suivantes, déterminez s'il s'agit d'une application.

1. $$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto x^3 - 5x + 2 \end{aligned}$$

2. $$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{x-3} \end{aligned}$$

3. $$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \sqrt{x-1} \end{aligned}$$

4. $$\begin{aligned} f : [1, +\infty[&\longrightarrow \mathbb{R} \\ x &\longmapsto \sqrt{x-1} \end{aligned}$$

5. $$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \ln(x^2+1) \end{aligned}$$

6. $$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \tan(x) \end{aligned}$$

7. $$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{e^x}{x^2+4} \end{aligned}$$

8. $$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \frac{x+y}{x-y} \end{aligned}$$

1-2 Notions de limites

Définition intuitive

On dit que la fonction $f(x)$ tend vers une limite L quand x tend vers a, si on peut rendre $f(x)$ aussi proche de L que l'on veut, en choisissant x suffisamment proche de a. On note : $\lim_{x \to a} f(x) = L$.

Formes Indéterminées à connaître

Les quatre formes d'indétermination classiques sont :

- $\frac{0}{0}$
- $\frac{\infty}{\infty}$
- $0 \times \infty$
- $\infty - \infty$

Lien avec l'IA

La notion de limite est fondamentale pour comprendre la **convergence des algorithmes d'optimisation**. Quand on entraîne un modèle, on espère que la fonction de coût tend vers une limite (un minimum) au fur et à mesure des itérations.

Limites de fonctions usuelles

Limites en $\pm\infty$

Fonction $f(x)$	$\lim_{x\to+\infty} f(x)$	$\lim_{x\to-\infty} f(x)$
x^n (n pair > 0)	$+\infty$	$+\infty$
x^n (n impair > 0)	$+\infty$	$-\infty$
$\frac{1}{x^n}$ ($n > 0$)	0	0
$\sqrt{x}$	$+\infty$	(n'existe pas)
e^x	$+\infty$	0
$\ln(x)$	$+\infty$	(n'existe pas)

Conseils pratiques

1. **Toujours commencer par essayer la substitution directe.** Si le résultat est un nombre fini, c'est la limite !
2. Si vous obtenez une **forme indéterminée** ($\frac{0}{0}$, $\frac{\infty}{\infty}$, etc.), il faut transformer l'expression :
   - Factoriser et simplifier (pour les polynômes).
   - Multiplier par la quantité conjuguée (pour les racines carrées).
3. Pour les limites en $\pm\infty$ d'une fraction rationnelle, comparez les termes de plus haut degré.
4. N'oubliez pas les limites usuelles (e^x, $\ln(x)$, $\frac{\sin(x)}{x}$...).

Exercices : Calcul de Limites

Calculez les limites suivantes

1. $\lim_{x\to 3}(x^2 - 2x + 1)$
2. $\lim_{x\to +\infty}(3x^3 - 5x^2)$
3. $\lim_{x\to 2} \frac{x^2-4}{x-2}$
4. $\lim_{x\to +\infty} \frac{2x^2+1}{5x^2-x}$
5. $\lim_{x\to 0} \frac{\sqrt{x+1}-1}{x}$
6. $\lim_{x\to -\infty} e^x + 5$
7. $\lim_{x\to 0} \frac{\sin(3x)}{x}$
8. $\lim_{x\to +\infty} \frac{x+7}{x^3-4}$
9. $\lim_{x\to 1} \frac{1}{x-1}$
10. $\lim_{x\to +\infty}(\sqrt{x^2 + x} - x)$

1-3 Dérivées de fonctions

Définition

Le nombre dérivé de f en x_0, noté $f'(x_0)$, est la limite (si elle existe et est finie) :

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

Géométriquement, $f'(x_0)$ est la **pente de la tangente** à la courbe de f au point d'abscisse x_0.

Lien avec l'IA

La dérivée nous indique **comment varie une fonction**. Si la dérivée est positive, la fonction croît. Si elle est négative, elle décroît. En IA, on "suit" la pente (la dérivée) pour trouver le minimum d'une fonction de coût : c'est le principe de la **descente de gradient**.

1-3 Dérivées des fonctions usuelles

Dérivées de base

Fonction	Dérivée
k (constante)	0
x^n	nx^{n-1}
$\frac{1}{x}$	$-\frac{1}{x^2}$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
e^x	e^x
$\ln(x)$	$\frac{1}{x}$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$

Opérations

Opération	Dérivée
$(u+v)'$	$u'+v'$
$(ku)'$	$k\cdot u'$
$(uv)'$	$u'v+uv'$
$(\frac{u}{v})'$	$\frac{u'v-uv'}{v^2}$
$(g\circ u)'$	$u'\times(g'\circ u)$

Exercices : Calcul de Dérivées

Calculez la fonction dérivée pour chacune des fonctions suivantes

1. $f(x) = 4x^3 - 7x^2 + 5x - 10$
2. $g(x) = x^2 e^x$
3. $h(x) = \frac{2x+1}{x^2+1}$
4. $k(x) = (3x^2 - 1)^5$
5. $p(x) = \ln(x^2 + 4)$
6. $q(x) = \cos(5x + \pi)$
7. $r(x) = e^{-3x} \sin(2x)$
8. $s(x) = \frac{\ln(x)}{x}$
9. $t(x) = \sqrt{x^2 + e^x}$
10. $u(x) = x\sqrt{1 - x^2}$

2-1 Introduction

Définition

Une fonction de plusieurs variables est une application f qui associe à un vecteur de n variables réelles $(x_1, \ldots, x_n)$ un unique nombre réel.

$$f : D \subset \mathbb{R}^n \to \mathbb{R}$$
$$(x_1, \ldots, x_n) \mapsto f(x_1, \ldots, x_n)$$

Exemple

La fonction $f(x_1, x_2) = x_1^2 + x_2^2$ associe à chaque point (x_1, x_2) du plan sa distance au carré par rapport à l'origine.

Lien avec l'IA

Un réseau de neurones est une fonction de **milliers, voire de milliards de variables** (les poids et les biais). La fonction de coût dépend de toutes ces variables. Nous sommes donc en plein dans le domaine des fonctions de plusieurs variables.

2-2 Dérivées partielles et Gradient

Dérivée partielle

La dérivée partielle de f par rapport à une variable x_i, notée $\frac{\partial f}{\partial x_i}$, mesure la variation de f lorsque **seule la variable** x_i **change**, les autres étant considérées comme des constantes.

$$\frac{\partial f}{\partial x_i}(a_1, \ldots, a_n) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h}$$

Exemple

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

- Pour calculer $\frac{\partial f}{\partial x}$, on traite y comme une constante : $\frac{\partial f}{\partial x} = 2xy$.
- Pour calculer $\frac{\partial f}{\partial y}$, on traite x comme une constante : $\frac{\partial f}{\partial y} = x^2 + 3y^2$.

2-3 Le Gradient

Définition du Gradient

Le **gradient** de f en un point est le vecteur qui contient toutes ses dérivées partielles. On le note ∇f.

$$\nabla f(x_1, \ldots, x_n) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Signification et lien avec l'IA

Le gradient a une propriété fondamentale :

- Il pointe dans la **direction de la plus forte croissance** de la fonction.
- L'opposé du gradient, $-\nabla f$, pointe dans la direction de la **plus forte décroissance**.

L'algorithme de la **descente de gradient** consiste à mettre à jour les paramètres (poids) du modèle en se déplaçant dans la direction de $-\nabla f$ pour minimiser la fonction de coût. C'est le moteur de l'apprentissage profond (Deep Learning).

Exercices : Calcul de Gradient

Calculez le gradient ∇f pour les fonctions suivantes

1. $f(x,y) = 2x^3 + 4y^2 - 5$
2. $f(x,y) = x^2y + 3xy^3$
3. $f(x,y) = e^{2x}\sin(y)$
4. $f(x,y) = \ln(x^2 + y^2)$
5. $f(w_1, w_2) = (w_1x_1 + w_2 - y)^2$
   (où x_1, y sont des constantes)
6. $f(x,y,z) = x^2y + y^2z + z^2x$

2-4 Dérivées d'ordre supérieur et Matrice Hessienne

Dérivées secondes

On peut dériver une seconde fois les dérivées partielles. On obtient 4 types de dérivées secondes pour une fonction de 2 variables $f(x, y)$:

$$\frac{\partial^2 f}{\partial x^2}, \quad \frac{\partial^2 f}{\partial y^2}, \quad \frac{\partial^2 f}{\partial y \partial x}, \quad \frac{\partial^2 f}{\partial x \partial y}$$

Application : Appliquer aux 4 premières fonctions de l'exercice précédent.

Matrice Hessienne

La matrice Hessienne regroupe toutes les dérivées secondes. Pour $f(x, y)$:

$$H_f = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

Lien avec l'IA

La Hessienne est utilisée dans des méthodes d'optimisation plus avancées (dites du second ordre), comme la méthode de Newton. Elle permet de trouver le minimum plus rapidement en tenant compte de la courbure de la fonction de coût.

3-1 Généralités sur les matrices

Définition

Une matrice A de taille $m \times n$ (ou d'ordre (m, n)) est un tableau de nombres à m lignes et n colonnes.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

- Si $m = n$, la matrice est **carrée**.
- Si $n = 1$, c'est un **vecteur colonne**.
- Si $m = 1$, c'est un **vecteur ligne**.

Lien avec l'IA

Les matrices sont omniprésentes en IA :

- Un **batch de données** est une matrice (chaque ligne est un exemple, chaque colonne une feature).
- Les **poids d'une couche de neurones** sont stockés dans une matrice.
- Une **image** peut être représentée comme une matrice de pixels.

3-2 Opérations sur les matrices

Addition

$C = A + B \implies c_{ij} = a_{ij} + b_{ij}$
(les matrices doivent avoir la même taille)

Multiplication par un scalaire

$B = \lambda A \implies b_{ij} = \lambda a_{ij}$

Produit matriciel

Si A est $m \times n$ et B est $n \times p$, alors $C = AB$ est $m \times p$.

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

Intuitivement : l'élément c_{ij} est le produit scalaire de la **ligne i** de A par la **colonne j** de B.
Attention : Le produit matriciel n'est **pas commutatif** ($AB \neq BA$ en général).

Lien avec l'IA

Le calcul effectué par une couche de neurones denses est un produit matriciel $Y = XW^T + b$, où X est la matrice des entrées, W la matrice des poids et b le vecteur de biais. L'efficacité des GPU en IA vient de leur capacité à paralléliser massivement ces produits matriciels.

Exercices : Opérations sur les Matrices

Effectuez les opérations suivantes (si possible)

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 4 \end{pmatrix} \qquad C = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 3 & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} 0 & -2 & 1 \\ -3 & 5 & 2 \end{pmatrix} \qquad D = \begin{pmatrix} 5 & 1 \\ -2 & 3 \end{pmatrix}$$

1. Calculez $A + 2B$.
2. Calculez le produit AC et CA.
3. Comparez les matrices obtenues en (2). Que pouvez-vous conclure ?

3-3 Matrices spéciales et déterminant

Matrices spéciales

- **Matrice identité I_n :** Matrice diagonale avec des 1 sur la diagonale. $AI_n = I_nA = A$.
- **Matrice transposée A^T :** Les lignes deviennent des colonnes (et vice-versa). $(a_{ij})^T = a_{ji}$.
- **Matrice inverse A^{-1} :** Si A est carrée, son inverse (s'il existe) vérifie $AA^{-1} = A^{-1}A = I_n$.

Déterminant (pour matrices carrées)

Le déterminant, noté $\det(A)$ ou $|A|$, est un nombre scalaire.
Pour une matrice 2×2 :

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Propriété cruciale : Une matrice carrée A est inversible **si et seulement si** $\det(A) \neq 0$.

Déterminant

Le déterminant d'une matrice carrée A d'ordre n est le nombre noté $det(A)$ ou bien $|A|$, il se calcul de manière récursive :
$det(A) = a_{11}$ si A $= (a_{11})$ est une matrice d'ordre 1.
D'une manière générale, si A $\in \mathbb{K}^{n,n}$

$$det(A) = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} det(A_i),$$

où A_i est la matrice obtenue en rayant la i-ème colonne et la 1ère ligne, ainsi $A_i \in \mathbb{K}^{n-1,n-1}$.
Si n $= 3$, alors,

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$det(A) = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Exemple

$$\begin{vmatrix} 1 & 3 & 0 \\ 2 & 6 & 4 \\ -1 & 0 & 2 \end{vmatrix} = 1 \begin{vmatrix} 6 & 4 \\ 0 & 2 \end{vmatrix} - 3 \begin{vmatrix} 2 & 4 \\ -1 & 2 \end{vmatrix} + 0 \begin{vmatrix} 2 & 6 \\ -1 & 0 \end{vmatrix} = 12 - 3 \times 8 = -12$$

NB : Lorsque le déterminant d'une matrice est $\pm$ 1, on dit que cette matrice est orthogonale.

Inverse d'une matrice

Soit A une matrice carrée d'ordre n. L'inverse de A (notée A^{-1}) est une matrice carrée d'ordre n qui satisfait $A^{-1}A = I_n$, où I_n est la matrice identité $n \times n$.

Si une telle matrice existe, on dit que la matrice A est inversible.

Si $det(A = 0)$, alors A n'est pas inversible.

Comatrice

La comatrice d'une matrice carrée $A = (a_{ij})$ est définie par $C = (C_{ij})$, où chaque C_{ij} est le cofacteur de l'élément a_{ij} de A.

Cofacteur

Le cofacteur de l'élément a_{ij} (noté C_{ij}) est égal à $C_{ij} = (-1)^{i+j} A_{ij}$ où A_{ij} est le déterminant de la matrice obtenue en supprimant la ligne i et la colonne j de A.

Exemple : Déterminez la comatrice de A avec $A = \begin{pmatrix} 7 & -5 & -8 \\ -3 & -7 & -2 \\ 0 & -4 & -8 \end{pmatrix}$;

Pour la ligne 1,

$$A_{11} = \begin{vmatrix} -7 & -2 \\ -4 & -8 \end{vmatrix} = 48; A_{12} = \begin{vmatrix} -3 & -2 \\ 0 & -8 \end{vmatrix} = 24; A_{13} = \begin{vmatrix} -3 & -7 \\ 0 & -4 \end{vmatrix} = 12$$

Pour la ligne 2,

$$A_{21} = \begin{vmatrix} -5 & -8 \\ -4 & -8 \end{vmatrix} = 8; A_{22} = \begin{vmatrix} 7 & -8 \\ 0 & -8 \end{vmatrix} = -56; A_{23} = \begin{vmatrix} 7 & -5 \\ 0 & -4 \end{vmatrix} = -28$$

Pour la ligne 3,

$$A_{31} = \begin{vmatrix} -5 & -8 \\ -7 & -2 \end{vmatrix} = -46; A_{32} = \begin{vmatrix} 7 & -8 \\ -3 & -2 \end{vmatrix} = -56;$$

$$A_{33} = \begin{vmatrix} 7 & -5 \\ -3 & -7 \end{vmatrix} = -64$$

Rappelons que les cofacteurs peuvent être obtenus à partir des mineurs correspondants en les multipliant par 1 ou -1, en fonction de leur position dans la matrice suivante : $\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$ Ainsi on obtient la comatrice suivante

$$\begin{pmatrix} 48 & -24 & 12 \\ -8 & -56 & 28 \\ -46 & 38 & -64 \end{pmatrix}$$

Si A est une matrice inversible, alors son inverse est

$$A^{-1} = \frac{1}{det(A)} C^t$$

Exercices : Déterminant et Inverse de Matrice

À vous de jouer !

Soient les trois matrices suivantes :

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 3 \\ 1 & 4 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & -1 & 5 \\ 0 & 3 & -2 \\ 0 & 0 & 4 \end{pmatrix}$$

1. Calculez le déterminant de chaque matrice : $\det(A)$, $\det(B)$ et $\det(C)$.
2. Identifiez les matrices qui sont inversibles. Justifiez votre réponse.
3. En utilisant le résultat précédent, calculez l'inverse de la matrice A, si elle existe.

Exemples de Matrices Particulières

1. Matrice Nulle

Tous les coefficients sont nuls.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

2. Matrice Diagonale

Seuls les coefficients sur la diagonale peuvent être non nuls ($a_{ij} = 0$ si $i \neq j$).

$$\begin{pmatrix} a_{11} & 0 & \cdots & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ \vdots & 0 & a_{33} & 0 & \vdots \\ \vdots & \vdots & 0 & a_{44} & 0 \\ 0 & \cdots & \cdots & 0 & a_{55} \end{pmatrix}$$

4. Matrice Tri-diagonale

Seules la diagonale principale et les deux diagonales adjacentes peuvent être non nulles.

$$\begin{pmatrix} a_{11} & a_{12} & 0 & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & \ddots & \vdots \\ 0 & a_{32} & a_{33} & a_{34} & 0 \\ \vdots & \ddots & a_{43} & a_{44} & a_{45} \\ 0 & \cdots & 0 & a_{54} & a_{55} \end{pmatrix}$$

5. Matrice Identité

Matrice diagonale avec $a_{ij} = 0$ si $i \neq j$ et $a_{ii} = 1$.

6. Matrice Triangulaire Supérieure

Tous les coefficients sous la diagonale principale sont nuls ($a_{ij} = 0$ si $i > j$).

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ 0 & a_{22} & a_{23} & a_{24} & a_{25} \\ 0 & 0 & a_{33} & a_{34} & a_{35} \\ 0 & 0 & 0 & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & a_{55} \end{pmatrix}$$

7. Matrice Symétrique

La matrice est égale à sa transposée ($a_{ij} = a_{ji}$).

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{12} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{13} & a_{23} & a_{33} & a_{34} & a_{35} \\ a_{14} & a_{24} & a_{34} & a_{44} & a_{45} \\ a_{15} & a_{25} & a_{35} & a_{45} & a_{55} \end{pmatrix}$$

3-4 Systèmes d'équations linéaires

On appelle *système de n équations à p inconnues le n-uplet d'équations :*

$$(S) : \begin{cases} a_{11}x_1 + ... + a_{1j}x_j + ... + a_{1p}x_p = b1 \\ a_{21}x_1 + ... + a_{2j}x_j + ... + a_{2p}x_p = b2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots = \dots \\ a_{i1}x_1 + ... + a_{ij}x_j + ... + a_{ip}x_p = bi \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots = \dots \\ a_{n1}x_1 + ... + a_{nj}x_j + ... + a_{np}x_p = bn \end{cases}$$

où les coefficients a_{ij} et b_i sont des nombres réels ou complexes et où les inconnues sont $x_1, x_2, \cdots, x_p$.

Nous pouvons regrouper d'une part les inconnues en un vecteur X et les éléments du second membre en un vecteur B. Ces deux vecteurs sont de dimensions X. D'autre part les coefficients a_{ij} forment la matrice carrée de taille n du système.

Soit :

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_p \end{pmatrix} ; B = \begin{pmatrix} b1 \\ b2 \\ \vdots \\ bi \\ \vdots \\ bp \end{pmatrix} ; A = \begin{pmatrix} a_{11} \cdots a_{1j} \cdots a_{1p} \\ a_{21} \cdots a_{2j} \cdots a_{2p} \\ \vdots \\ a_{i1} \cdots a_{ij} \cdots a_{ip} \\ \vdots \\ a_{n1} \cdots a_{nj} \cdots a_{np} \end{pmatrix}$$

Forme matricielle

Un système de n équations à p inconnues peut s'écrire sous la forme compacte :

$$AX = B$$

où :

- A est la matrice $(n \times p)$ des coefficients.
- X est le vecteur colonne $(p \times 1)$ des inconnues.
- B est le vecteur colonne $(n \times 1)$ des seconds membres.

Résolution

Si A est une matrice carrée $(n = p)$ et **inversible** $(\det(A) \neq 0)$, alors le système admet une solution unique :

$$X = A^{-1}B$$

Exemple : Résoudre

$$\begin{cases} 2x_1 + 2x_2 + x_3 = 2 \\ x_1 + 2x_2 + x_3 = 1 \\ x_1 - 2x_2 + 2x_3 = 3 \end{cases}$$

Lien avec l'IA

Bien que l'on calcule rarement l'inverse explicitement (trop coûteux), la résolution de systèmes linéaires est au cœur de nombreux problèmes. Par exemple, les **régressions linéaires** se résolvent via un système d'équations (les équations normales). Des algorithmes comme le **pivot de Gauss** sont utilisés pour trouver la solution de manière efficace.

4-1 Formule de Taylor

L'idée

La formule de Taylor permet d'approximer une fonction "compliquée" $f(x)$ au voisinage d'un point x_0 par un polynôme, qui est beaucoup plus simple à manipuler.

Formule de Taylor à l'ordre n

Si f est suffisamment dérivable en x_0, alors pour x proche de x_0 :

$$f(x) \approx \underbrace{f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n}_{\text{Polynôme de Taylor d'ordre n}}$$

- Si $x_0 = 0$, on parle de la formule de **Maclaurin**.
- L'approximation est d'autant meilleure que l'ordre n est élevé et que x est proche de x_0.

Lien avec l'IA

L'approximation de Taylor est la base théorique de la descente de gradient et d'autres méthodes.

- **Ordre 1 :** $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$. C'est une approximation linéaire. La descente de gradient suppose que la fonction de coût est localement linéaire.
- **Ordre 2 :** $f(x) \approx \cdots + \frac{f''(x_0)}{2}(x - x_0)^2$. C'est une approximation quadratique. Les méthodes de Newton utilisent cette information sur la courbure pour converger plus vite.

5-1 Principe

Approcher la dérivée

En revenant à la définition de la dérivée $f'(x_0) = \lim_{h \to 0} \frac{f(x_0+h)-f(x_0)}{h}$, on peut obtenir une approximation en choisissant un h **petit mais non nul**.

$$f'(x_0) \approx \frac{f(x_0+h)-f(x_0)}{h}$$

C'est la formule des **différences finies "avant"** (forward difference).

Pour les dérivées partielles

De même, on peut approximer les dérivées partielles :

$$\frac{\partial f}{\partial x}(x_0, y_0) \approx \frac{f(x_0+h, y_0)-f(x_0, y_0)}{h}$$

$$\frac{\partial f}{\partial y}(x_0, y_0) \approx \frac{f(x_0, y_0+h)-f(x_0, y_0)}{h}$$

Lien avec l'IA : Le "Gradient Checking"

En Deep Learning, l'algorithme de **backpropagation** calcule le gradient exact (analytique) de la fonction de coût. Cet algorithme est complexe et facile à mal implémenter.
Pour vérifier que notre code est correct, on utilise le **gradient checking** :

1. On calcule le gradient avec notre backpropagation.
2. On calcule une **approximation** du gradient avec les différences finies.
3. On compare les deux résultats. S'ils sont très proches, notre implémentation est probablement correcte !

C'est une technique de débogage indispensable.

Exercices : Taylor et Approximations Numériques

Développement de Taylor

1. Soit la fonction $f(x) = e^{2x}$. Calculez les dérivées $f'(x)$, $f''(x)$ et $f'''(x)$.
2. Déterminez le développement de Maclaurin de $f(x) = e^{2x}$ à l'ordre 3.
3. En utilisant le polynôme de Taylor de l'ordre 1, donnez une approximation de $e^{0.2}$ (c'est-à-dire $f(0.1)$).

Approximation de la dérivée

4. En utilisant la formule de la dérivée de la question 1, calculez la valeur **exacte** de $f'(0)$.
5. Utilisez la formule des différences finies "avant" avec $h = 0.1$ pour trouver une valeur **approchée** de $f'(0)$.
6. Comparez la valeur exacte (question 4) et la valeur approchée (question 5). Que remarquez-vous ? Comment pourrait-on améliorer la précision de l'approximation numérique ?

Conclusion

Ce qu'il faut retenir

Les outils mathématiques présentés ne sont pas des concepts abstraits, mais les **briques fondamentales** qui permettent aux algorithmes d'IA de fonctionner et d'apprendre.

- La **dérivée** et le **gradient** sont les guides pour l'optimisation.
- Les **matrices** sont le langage pour manipuler les données et les modèles.
- Les **approximations (Taylor, différences finies)** nous permettent de comprendre et de construire des algorithmes efficaces.