Correction de l'erreur sur la formule de dérivation

# TD4


Soit $X_0$ proche du min

## Ex. 1

$$g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

$$g = \begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix}$$

$\downarrow$

Jacobien

$$Jg = \begin{array}{c} \\ i \end{array} \begin{pmatrix} & \vdots^{\,j} & \\ \cdots & \bullet & \cdots \\ & \vdots & \end{pmatrix} \; \nabla g_i^t$$

$$\frac{\partial g_i}{\partial x_j}$$

$\nabla g$..

$$f(\vec{x}) = f(x_1 \dots x_d)$$

$$f : \mathbb{R}^d \longrightarrow \mathbb{R}$$

$\downarrow$

gradient

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{pmatrix} \quad \mathbb{R}$$

$$w_1 x_1 + \dots + w_d x_d \qquad \frac{\partial}{\partial x_i} \rightsquigarrow w_i$$

$w_i x_i$

$$f_1(\vec{x}) = \langle \vec{w}, \vec{x} \rangle = \sum_j w_j \cdot x_j$$

$$= \vec{w}^t \times \vec{x}$$

$f_1 : \mathbb{R}^d \longrightarrow \mathbb{R}$

linéaire

$\nabla f_1$ ?

$$= \begin{array}{c} \\ i \\ \\ \end{array} \begin{pmatrix} \vdots \\ \frac{\partial f_1}{\partial x_i} \\ \vdots \end{pmatrix}$$

$$\frac{\partial f_1}{\partial x_i} = w_i$$

$$\nabla f_1 = \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix} = \vec{w}$$

**Dériv. 1 — linéaire (1)**

$$\nabla\left(\langle \vec{w}, \vec{x} \rangle\right) = \nabla\left(\vec{w}^t \times \vec{x}\right)$$

$$= \vec{w}$$

$$f_2(\vec{x}) = \vec{x}^{\,t} \times \underbrace{\overset{\boxed{\text{symétrique}}}{B}}_{\text{mat}} \times \vec{x} \longrightarrow \text{poly. d° 2 exactement}$$

forme quadratique

$\downarrow$

forme à reconnaître (2)

forme quadratique

$\equiv$

poly. d° 2 exactement

ex : $\vec{x} \in \mathbb{R}^3$, $\vec{x} = (x, y, z)$

$$B = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & \frac{1}{2} \\ 0 & \frac{1}{2} & 3 \end{pmatrix}$$

$$\vec{x}^{\,t} . B . \vec{x} = (x \;\; y \;\; z) \underbrace{\begin{pmatrix} 1 & 1 & \cdot \\ 1 & 2 & \frac{1}{2} \\ \cdot & \frac{1}{2} & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}} = (x \;\; y \;\; z) \begin{pmatrix} x + y \\ x + 2y + \frac{1}{2} z \\ \frac{1}{2} y + 3z \end{pmatrix}$$

$$= x(x+y) + y\left(x + 2y + \frac{1}{2} z\right) + z\left(\frac{1}{2} y + 3z\right)$$

$$= x^2 + 2y^2 + 3z^2 + 2\,xy + yz$$

$$\begin{array}{c|ccc|} & xy & & \\ \hline & 1 & 1 & 0 \\ yx & 1 & 2 & \frac{1}{2} \\ & 0 & \frac{1}{2} & 3 \end{array}$$

Poly d° 2 exactement $\xleftrightarrow{\text{unique}}$ matrice B symétrique

$P(x)$ $=$ ${}^tX . B . X$

$$B \text{ m } \begin{array}{c|ccc|} & x_1 & \dots & x_d \\ \hline x_1 & & & \\ \vdots & & & \\ x_d & & & \end{array}$$

$x_i x_j$ sym

termes carrés.

ex :

$1 \quad x \quad y$

non sym

$$\begin{array}{c|c}
 & \begin{array}{cc} x & y \end{array} \\
\begin{array}{c} x \\ y \end{array} & \begin{pmatrix} & 1/2 \; 1, 0, 2, \dots \\ 1/2 \; 0, 1, -1, \dots & \end{pmatrix}
\end{array}$$

$$f_2(\vec{x}) = \vec{x}^t B \vec{x} = \sum_{i=1}^{d} \sum_{j=1}^{d} b_{ij} x_i x_j$$

$$\frac{\partial f_2}{\partial x_k} \qquad k \in 1 \dots d$$

$\frac{\partial}{\partial x_k}$

$i$ et $j \neq k$: $\frac{\partial}{\partial x_k} = 0$

$i = j = k$ (terme carré $\Leftrightarrow$ diag)

$$\frac{\partial}{\partial x_k}\left(b_{kk} x_k^2\right) = 2 b_{kk} x_k$$

$i = k \quad j \neq k$ ou $j = k \quad i \neq k$

$$\frac{\partial b_{kj} x_k x_j}{\partial x_k} = b_{kj} x_j$$

$$\frac{\partial b_{ik} x_i x_k}{\partial x_k} = b_{ik} x_i$$

1 terme

$$\frac{\partial}{\partial x_k} f_2(\vec{x}) = 2 b_{kk} x_k + \underbrace{\sum_{\substack{i=1 \\ i \neq k}}^{d} b_{ki} x_i + \sum_{\substack{i=1 \\ i \neq k}}^{d} b_{ik} x_i}_{\sum_{\substack{i=1 \\ i \neq k}}^{d} (b_{ik} + b_{ki}) x_i}$$

$$\frac{\partial f_2}{\partial x_k} = \sum_{j=1}^{d} (b_{jk} + b_{kj}) x_j$$

$$\nabla f_2(\vec{x}) = i \begin{pmatrix} \vdots \\ \sum_{j=1}^{d} (b_{ji} + b_{ij}) x_j \\ \vdots \end{pmatrix} \text{ — } \frac{\partial}{\partial x_i}$$

$$i \begin{pmatrix} \text{—— } B \text{ ——} \end{pmatrix} \begin{pmatrix} \vdots \\ \vdots \end{pmatrix} \quad \vec{x}$$

$\times$ ligne $i$

$$\begin{pmatrix} \vdots \\ \sum_{j=1}^{d} b_{ij} x_j \\ \vdots \end{pmatrix} i = B \cdot \vec{x}$$

$$= B^t \cdot \vec{x} + B \cdot \vec{x}$$

$$\nabla f_2(\vec{x}) = (B + B^t)\,\vec{x}$$

Si B sym

$$B + B^t = 2B$$

Si B non sym

$B + B^t$ sym

$\boxed{\text{Dériv. 2 — forme quad / poly } d^\circ 2 \quad \nabla({}^tX.B.X) = (B+B^t)X \quad (= 2BX \text{ si } B \text{ sym.})}$

$$f_3(\vec{x}) = \|\vec{x}\|^2 \longrightarrow \vec{x}^t . ? . \vec{x} \qquad \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$$

$= x_1^2 + \dots + x_d^2$ — forme quad - $d^\circ 2$

$$= \langle \vec{x}, \vec{x} \rangle = \vec{x}^t . \vec{x}$$

$$= \vec{x}^t . \underbrace{I}_{\text{sym}} . \vec{x} \qquad I = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

$$\nabla f_3 = 2.I.\vec{x}$$

$$= 2.\vec{x}$$

$$\nabla(\|\vec{x}\|^2) = 2\vec{x}$$

$$f_4(\vec{x}) = \|\underbrace{A\vec{x} + b}_{f}\|^2 \quad \text{— composée . } " \|\cdot\|^2 \circ (Ax+b) "$$

$g : \|\cdot\|^2$

$A\vec{x} + b : \mathbb{R}^d \to \mathbb{R}^d$ ($A$ : $d \times d$, $b$ : $d$)

$J \cancel{X} (A\vec{x}+b)$

$(g \circ f)' = f' \times (g' \circ f)$

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

$" \nabla(g \circ f) = \underbrace{J(f)^t}_{\text{mat}} \times \underbrace{(\nabla g \circ f)}_{\text{vect}}$

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

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

$f : \mathbb{R}^d \to \mathbb{R}^d$

$g : \mathbb{R}^d \to \mathbb{R}$

$$\nabla f_4(\vec{x}) = A^t \times \overbrace{2.(Ax+b)}^{\nabla g(f)}$$

$\nabla g(\vec{x}) = 2\vec{x}$

$\nabla g = 2I$

car

$$\underbrace{J}_{\text{dérivée}}(A\vec{x}+b) = A$$

$$\nabla f_4(\vec{x}) = 2A^t(A\vec{x}+b)$$

dim 1 $\sim (ax)^2 \longrightarrow 2ax \cdot a = 2a^2x$

$f_5(\vec{x}) = \|A\vec{x}\|^2$ — $f_4$ avec $b = \vec{0}$

$$\nabla f_5(\vec{x}) = \cancel{2A^2\vec{x}} \quad 2A^tA\vec{x}$$

$$f_6(\vec{x}) = \|A\vec{x}+b\|_2 = \sqrt{\|Ax+b\|^2}$$

$g \circ f$

$g : \mathbb{R}^{(+)} \longrightarrow \mathbb{R}$

$\sqrt{\ }$

$f(\vec{x}) = \|A\vec{x}+b\|^2$
$: \mathbb{R}^d \to \mathbb{R}$

$$\nabla f_6(\vec{x}) = \nabla(g\circ f) = \nabla f \times g' \circ f \quad (*)$$

$$= \left(\cancel{2} A^t (A\vec{x}+b)\right) \times \frac{1}{\cancel{2}} \cdot \frac{1}{\underset{\sim}{\sqrt{\|Ax+b\|^2}}}$$

$$\nabla f_6(\vec{x}) = \frac{A^t(A\vec{x}+b)}{\|A\vec{x}+b\|}$$

$/$
$\|A\vec{x}+b\|$

$$\sqrt{x} = x^{1/2} = \frac{1}{2} x^{-\frac{1}{2}}$$

(*) Formule correcte :
$f : \mathbb{R}^d \to \mathbb{R}$
$g : \mathbb{R} \to \mathbb{R}$

$$\nabla(g\circ f) = \underbrace{J(f)^t}_{=\nabla f} \times g' \circ f$$

ex :
$$\nabla(\|\vec{x}\|) = \frac{\vec{x}}{\|\vec{x}\|}$$

$$f_7(\vec{x}) = e^{\langle \vec{w}, \vec{x}\rangle} \longrightarrow g\circ f$$

- $f : \mathbb{R}^d \to \mathbb{R}$, $\vec{x} \longmapsto \langle \vec{w}, \vec{x}\rangle$ — linéaire
- $g : \mathbb{R} \to \mathbb{R}^+$, $t \longmapsto e^t$

intuition $\quad f' \times g' \circ f$

$$\nabla f_7(\vec{x}) = \nabla f(\vec{x}) \times g'(f(\vec{x})) \quad \text{idem}$$

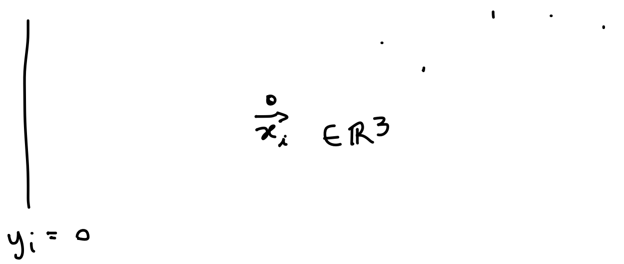
$$= \vec{w} \times e^{\langle \vec{w}, \vec{x}\rangle}$$

## Ex. 2

Pts de données $(\vec{x}_i, y_i)$ — répartis linéairement

$\vec{x}_i$ : $\mathbb{R}^d$ — modèle linéaire

GIG




$\longrightarrow$ scannés sur un ~~plan~~.

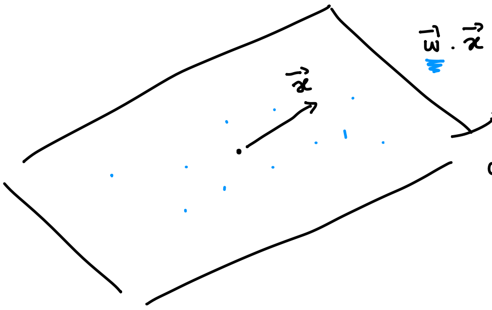
$\notin$ plan

proche d'un plan

Éq. linéaire

?

$$\vec{w} \cdot \vec{x} = y_0$$



$\vec{w} \cdot \vec{x} = 0$ → plan vect. contient $\vec{0}$

**3D** $\vec{x} = (x, y, z)$

$ax + by + cz = 0$

$\begin{pmatrix} a \\ b \\ c \end{pmatrix}^t \times \vec{x} = 0$

w

$\langle \vec{w}, \vec{x} \rangle = 0$

<u>Pb</u> trouver le plan le + proche des pts

$\vec{w}$ ?

éq. $\langle \vec{w}, \vec{x} \rangle = 0$

$\vec{x}_i \rightsquigarrow \vec{w} \cdot \vec{x}_i = \varepsilon_i$ erreur au pt $i$

$d(\vec{x}_i, P_{\vec{w}})$ ✓

moindres carrés

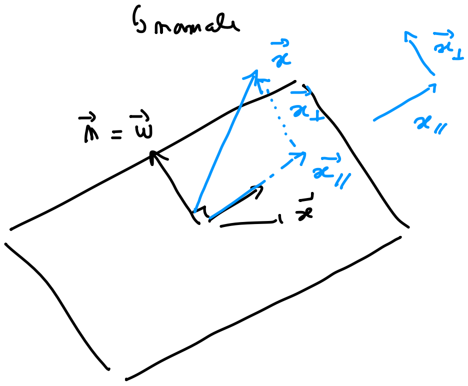
erreur globale $\sum \varepsilon_i^2$

$= \sum (w \cdot \vec{x}_i)^2$

w → min ?

Plan d'éq. $\vec{w} \cdot \vec{x} = 0$

normale




$\vec{w} \cdot \vec{x} = d(\vec{x}, P)$

$\vec{w} \cdot \vec{x} = \vec{w} \cdot (\vec{x}_\parallel + \vec{x}_\perp)$

$= \cancel{\vec{w} \cdot \vec{x}_\parallel} + \vec{w} \cdot \vec{x}_\perp$

0

colinéaires

$\|\vec{w}\| \cdot \|\vec{x}_\perp\|$

$d(\vec{x}, P)$

$= \frac{\vec{w}}{\|\vec{w}\|} \cdot \vec{x}$

$(\vec{x}_i, y_i)$ données

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

suit un modèle ~~linéaire~~ affine

$\sim$ $\vec{w} \cdot \vec{x} = y$ modèle → $\varepsilon_i = \vec{w} \cdot \vec{x}_i - y_i$

0 si pt exactement sur le modèle

erreur moindres carrés.

on cherche le min de $\frac{1}{m} \sum_{i=1}^{m} \left( \langle \vec{w}, \vec{x}_i \rangle - y_i \right)^2$

Régression linéaire ⟷ modèle linéaire cherché aux moindres carrés.

$$\min_{\vec{w}} f(\vec{w}) = \frac{1}{m}\sum_{i=1}^{m} \Big(\overbrace{\langle \vec{w}, \vec{x}_i \rangle - y_i}^{\varepsilon_i}\Big)^2$$

$\vec{x}_i \,/\, y_i$ données.

$\left( \min_{\vec{w}} \right.$ ?

Nouveaux pts $\vec{x}_j,\ y_j$

$$\begin{cases} \to \in \text{échantillon} \xleftarrow{\text{petit}} w\vec{x}_j - y_j \\ \to \notin \xleftarrow{\text{grand}} w\vec{x}_j - y_j \end{cases}$$

i) $X = (\vec{x}_1 \ \dots \ \vec{x}_m)^t$

$\vec{x}_1 \downarrow \begin{pmatrix} \\ \\ \end{pmatrix}$ $d$

$$X : \begin{pmatrix} \vdots \\ -\ x_i^t\ - \\ \vdots \end{pmatrix} \quad m \times d$$

$$\vec{y} : \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

$$f(w) = \sum \underbrace{(\dots)}_{\varepsilon_i}^2 = \|\vec{\varepsilon}\|^2 = \vec{\varepsilon} \cdot \vec{\varepsilon}$$

$$\vec{\varepsilon} = \begin{pmatrix} \vdots \\ \varepsilon_i \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \langle \vec{w}, \vec{x}_i \rangle - y_i \\ \vdots \end{pmatrix}$$

lin / $\vec{w}$

$(\text{mat} \times \vec{w} + \text{vect})_i$

$$\vec{\varepsilon} = \underbrace{\begin{pmatrix} \vdots \\ -\ x_i^t\ - \\ \vdots \end{pmatrix}}_{X} \vec{w} - \underbrace{\begin{pmatrix} \\ y_i \\ \\ \end{pmatrix}}_{\vec{y}}$$

$$f(\vec{w}) = \|\vec{\varepsilon}\|^2 = \|X \cdot \vec{w} - \vec{y}\|^2 \longleftrightarrow f_4$$

ii) $\nabla f(\vec{w}) = 2\, X^t (X\vec{w} - \vec{y})$

## Ex.3

$x_0, y_0, r$ ?

$$\varepsilon_i = (x_i - x_0)^2 + (y_i - y_0)^2 - r^2$$

non linéaire / $x_0, y_0, r$ ....

$$\vec{\varepsilon} = \text{non lin} \,/\, x_0, y_0, r$$