Optimisation

3 séances d'optimisation

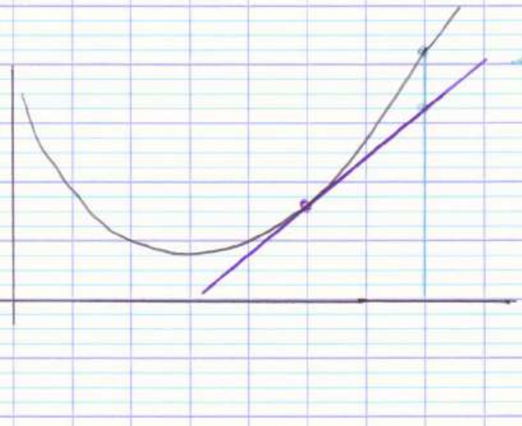
- exemple de problèmes d'optimisation :
  soit une fonction $f : \mathbb{R}^n \to \mathbb{R}$.
  On cherche $x \in \mathbb{R}^n$ tel que $\forall y \in \mathbb{R}^n$, $f(x) \leq f(y)$
  C'est le problème : $\min\limits_{x \in \mathbb{R}^n} f(x)$

- autre exemple : les moindres carrés (quadratique convexe)

- <u>Définition</u> :
  une fonction f est convexe si :

$$\forall x \in \mathbb{R}^n, \forall y \in \mathbb{R}^n, \forall t \in [0,1] :$$

$$f(tx + (1-t)y) \leq t f(x) + (1-t) f(y)$$



<u>Proposition</u> :
Si f est convexe et dérivable, alors $\forall y \in \mathbb{R}^p, \forall x \in \mathbb{R}^p$,
$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$

démonstration :

soit $x \in \mathbb{R}^p$ et $y \in \mathbb{R}^p$, et soit $t \in [0,1]$.

$f(x+t(y-x)) = f(x) + t\langle \nabla f(x), y-x\rangle + o(t)$ car dérivable

où $o(t)$ vérifie $\lim_{t\to 0} \frac{o(t)}{t} = 0$

$f(x+t(y-x)) = f(ty+(1-t)x) \leq t f(y) + (1-t) f(x)$ convexité

donc : $t f(y) + (1-t) f(x) \geq f(x) + t\langle \nabla f(x), y-x\rangle + o(t)$

$f(y) \geq f(x) + \langle \nabla f(x), y-x\rangle + \frac{o(t)}{t}$ (soustrait $f(x)$, divise par t)

Je fais tendre t vers 0 :

$f(y) \geq f(x) + \langle \nabla f(x), y-x\rangle$

Théorème :

Si f est dérivable et x^* est un minimiseur de f, alors $\nabla f(x^*) = 0$

démonstration :

$\forall y \in \mathbb{R}^p$, $f(y) \geq f(x^*)$ (car x^* minimiseur)

$f(x^*+t(y-x^*)) = f(x^*) + t\langle \nabla f(x^*), y-x^*\rangle + o(t)$, $\forall t \in \mathbb{R}$

$f(x^*) + t\langle \nabla f(x^*), y-x^*\rangle + o(t) \geq f(x^*)$

$\langle \nabla f(x^*), y-x^*\rangle + \frac{o(t)}{t} \geq 0$ (soustrait $f(x^*)$, divise par t)

Je fais tendre vers 0 : j'obtiens que $\forall y \in \mathbb{R}^p$,

$\langle \nabla f(x^*), y-x^*\rangle \geq 0$

Je choisis $y = x^* - \nabla f(x^*)$:

$-\|\nabla f(x^*)\|^2 \geq 0$

et donc $\nabla f(x^*) = 0$

Théorème :

si f est convexe et dérivable, alors :

x^* minimiseur de $f \Leftrightarrow \nabla f(x^*) = 0$

(autrement dit, il suffit de trouver un point tel que $\nabla f(x) = 0$ pour avoir un minimum).

démonstration :

- si x^* est minimiseur, alors $\nabla f(x^*) = 0$ (par ce qui précède).
- si f est convexe :

$\forall y \in \mathbb{R}^p, f(y) \geq f(x^*) + \langle \underbrace{\nabla f(x^*)}_{=0}, y - x \rangle \geq f(x^*)$

donc x^* minimiseur de f.

Revenons aux moindres carrés :

$f(\beta) = \frac{1}{2} \|X\beta - y\|_2^2$

$\nabla f(\beta) = X^T(X\beta - y)$

$\beta^* \in \underset{\beta}{\arg\min} f(\beta) \Leftrightarrow \nabla f(\beta^*) = 0 \Leftrightarrow X^T X \beta^* = X^T y$

Analyse en composante principale :

$\min_{M \in \mathbb{R}^{p \times n}} \|A - M^T M\|^2 \quad , \quad A \in \mathbb{R}^{n \times n}$

Lasso :

$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|X\beta - y\|_2^2 + \lambda \|\beta\|_1$

résolu avec l'opérateur proximal :

$\operatorname{prox}_{|\cdot|}(\beta) = \underset{b \in \mathbb{R}}{\arg\min} |b| + \frac{1}{2}(\beta - b)^2$ (seuillage doux)

NMF :

$A \in \mathbb{R}^{n \times p}$

$\min_{U \in \mathbb{R}^{n \times q}, V \in \mathbb{R}^{q \times p}} \|A - UV\|_F^2 \quad U_{ij} \geq 0 \ \forall i,j \quad V_{jl} \geq 0 \ \forall j,l$

non convexe avec contraintes

Régression logistique :

modèle linéaire : $y_i = x_i^T w + w_0 + \varepsilon_i$

classification : $y_i \in \{-1, 1\}$

classifieur : $h : x \mapsto \operatorname{sign}(\langle x, w \rangle + w_0) \quad (w \in \mathbb{R}^p, w_0 \in \mathbb{R})$

$\rightarrow$

problème d'optimisation :

$$\min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} \log(1 + \exp($$

__fonctions composées :__

$f : \mathbb{R}^p \to \mathbb{R}^q$

$g : \mathbb{R}^q \to \mathbb{R}^m$

$g \circ f : \mathbb{R}^p \to \mathbb{R}^m$

$\quad x \to g(f(x))$

Jacobienne : $J_{g \circ f}(x) = J_g(f(x)) \times J_f(x)$

__Réseaux de neurones :__

$L(y, \hat{y}) = \sum_{i=1}^{n} \| y_i - \hat{y}_i \|^2$

$\sigma : \mathbb{R} \to \mathbb{R}$

$N : \mathbb{R}^H \times \mathbb{R}^H \to \mathbb{R}$

$\quad (w, V) \to \sum_{i=1}^{H} w_i V_i$

$V_i : \mathbb{R}^H \to \mathbb{R}^H$

$z \to \begin{pmatrix} v_1(z_1) \\ \vdots \\ v_H(z_H) \end{pmatrix}$

$X : W \to \left(\sum_{j=1}^{p} W_{ij} x_j \right)_{1 \leq i \leq H}$

$F(w, W) = L(y, \sigma(N(w, v(X(N)))))$