$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} \qquad X^T = (X_1, \dots, X_n)$$

2) $f(w) = \frac{1}{2n} \|y - Xw\|_2^2 + \frac{\lambda}{2} \|w\|_2^2 = \frac{1}{n} \sum_{i=1}^{n} f_i(w)$ où $f_i(w) = \frac{1}{2}\left[(y_i - x_i^T w)^2 + \lambda \|w\|^2\right]$

grad i

$\nabla f_i(w) = -(y_i - x_i^T w) X_i + \lambda w.$

$$\underset{\text{grad}}{\nabla f(w)} := \frac{1}{n} \sum_{i=1}^{n} \underbrace{-X_i}_{\text{vect.}} \underbrace{(y_i - x_i^T w)}_{\text{scalaire.}} + \lambda w = \underbrace{\left[\frac{1}{n} \sum_{i=1}^{n} X_i X_i^T + \lambda I\right]}\, w + \text{cste } \% w.$$

$$\|\nabla f(w) - \nabla f(w')\| \leq \underbrace{\Big|\Big|\Big| \overbrace{\frac{1}{n} \sum_{i=1}^{n} X_i X_i^T + \lambda I}^{M} \Big|\Big|\Big|}_{L} \; \|w - w'\|$$

si M : symmétrique réelle, $M = U^T D U$ $\quad U^T U = I$.

$$\boxed{L = |||M|||} = \sup_u \frac{\|Mu\|}{\|u\|} = \sqrt{\sup_u \frac{u^T M^T M u}{u^T u}} \underset{\uparrow\, M \text{ symmétrique}}{=} \sup \{|\lambda| ;\ \lambda \in Spec(M)\} = \sup_u \frac{|u^T M u|}{u^T u}.$$

Lip

$$\boxed{L = \Big|\Big|\Big| \frac{1}{n} \sum X_i X_i^T + \lambda I \Big|\Big|\Big|}$$

$$= \Big|\Big|\Big| \frac{1}{n} \sum X_i X_i^T \Big|\Big|\Big| + \lambda.$$

si $X_i \in \mathbb{R}^d$, $\quad |||\underbrace{X_i X_i^T}_{M_i}||| = \sup \{\lambda ;\ \lambda \in Spec(M_i)\}$

si $v \in \{X_i\}^\perp$, $\quad (X_i X_i^T) v = 0 \quad$ donc : 0 vp. d'ordre $d-1$.

$(X_i X_i^T) X_i = \|X_i\|^2 X_i \Rightarrow \|X_i\|^2$ vp d'ordre 1.

Donc $\boxed{|||X_i X_i^T||| = \|X_i\|^2.} \Rightarrow |||X_i X_i^T + \lambda I||| = \underbrace{\|X_i\|^2 + \lambda}.$

$$\underset{\text{Lip max}}{L_{max}} = \max [L_{\nabla f_1}, \dots, L_{\nabla f_n}] = \max [\underbrace{|||X_1 X_1^T + \lambda I|||}, \dots, |||X_n X_n^T + \lambda I|||]$$

$$(\|X_1\|^2 + \lambda, \dots, \|X_n\|^2 + \lambda).$$

$(L_1, \dots, L_d)$: Lip. coordinate

$\frac{\partial f}{\partial w_j} = \frac{1}{n} \sum_{i=1}^{n} -X_i[j] (y_i - x_i^T w) + \lambda w[j].$ grad_coordinate.

$$\frac{\partial^2 f}{\partial w_j^2} = \boxed{\frac{1}{n} \sum_{i=1}^{n} X_i^2[j] + \lambda := L_j}$$

Régression Logistique :

$$\ell(y_i | x_i) = \begin{cases} \dfrac{e^{x_i^T w}}{1 + e^{x_i^T w}} = \dfrac{1}{1 + e^{-x_i^T w}} & \text{si } y_i = 1 \\ \dfrac{1}{1 + e^{x_i^T w}} & \text{si } y_i = -1 \end{cases}$$

$$= \frac{1}{1 + e^{-y_i x_i^T w}}.$$

Loss : $f(w) := -\frac{1}{n} \sum_{i=1}^{n} \log \ell(y_i | x_i) + \frac{\lambda}{2} \|w\|^2.$

$$\underset{\text{LE Loss}}{f(w)} := \frac{1}{n} \sum_{i=1}^{n} \underbrace{\log(1 + e^{-y_i x_i^T w})}_{\psi(y_i x_i^T w)} + \frac{\lambda}{2} \|w\|^2.$$

où $\psi(t) = \log(1 + e^{-t})$.

$\psi'(t) = \frac{-e^{-t}}{1 + e^{-t}} = \frac{-1}{1 + e^t}$

$\psi''(t) = \frac{e^t}{(1 + e^t)^2} \leq \frac{1}{4}$

$$\underset{\text{LE Grad}}{\nabla f(w)} = \frac{1}{n} \sum_{i=1}^{n} \psi'(y_i x_i^T w)\, y_i X_i + \lambda w.$$

$\nabla f_i(\omega)$ LR grad i

Donc : $\nabla f(w) = \frac{1}{n}\sum_{i=1}^{n} \frac{-1}{1+e^{y_i x_i^T w}} y_i x_i + \lambda w$; $X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix}$

$$= \frac{1}{n}[x_1,\dots,x_n]\left\{\begin{bmatrix}-y_1\\ \vdots \\ -y_n\end{bmatrix} \underset{\substack{\uparrow \\ \text{multip. terme à terme.}}}{\odot} \left[-\frac{1}{1+e^{y \odot Xw}}\right]\right\} + \lambda w$$

LR grad.

$$\frac{\partial f}{\partial w_j} = \frac{1}{n}\sum x_i[j]\,(\ )$$

$$\nabla^2 f(w) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\frac{e^{y_i x_i^T w}}{(1+e^{y_i x_i^T w})^2}}_{\varphi(y_i x_i^T w)} \underbrace{y_i^2}_{1} x_i x_i^T + \lambda I.$$

$$\frac{u}{(1+u)^2} = \underbrace{\left(\frac{u}{1+u}\right)}_{p}\underbrace{\left(\frac{1}{1+u}\right)}_{1-p} \le 1/4.$$

$$|||\nabla^2 f(w)||| = \sup_{\|u\|=1} u^T \nabla^2 f\, u = \sup_{\|u\|=1} \frac{1}{n}\sum_{i=1}^{n}\varphi(y_i x_i^T w)\, u^T x_i x_i u + \lambda \underbrace{u^T u}_{=1}.$$

$$\le \frac{1}{4}\sup_{\|u\|=1} u^T\left[\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T\right]u + \lambda.$$

$$= \frac{1}{4}\left|\left|\left|\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T\right|\right|\right| + \lambda.$$) LR lip.

$$L = \left|\left|\left|\frac{1}{4}\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T\right)\right|\right|\right| + \lambda.$$

$$L_{max} = \max\left(\frac{1}{4}\|x_1\|^2+\lambda, \dots, \frac{1}{4}\|x_n\|^2+\lambda\right)$$ LR lip. max.

axis=(1) → Σ ligne.

$$(\underbrace{L_1},\dots,\underbrace{L_d}) = \left(\frac{1}{4}\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i^2[1]}_{} + \lambda, \dots, \frac{1}{4}\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i^2[d]}_{} + \lambda\right)$$ LR lip. coord.

axis=0 : somme par colonne.

$w_j \mapsto f(w)$.

AGD : $\alpha^{(k)} = \frac{1}{Lip}$.

(idx_sample) = unif $[\{0,\dots,n-1\}]$
: n x Iter n.

GD : Boucle : k:0 : nb_iter
$\omega \leftarrow \omega - \eta\nabla f(\omega)$

$\omega_{t+1} = \omega_t - \eta\nabla f(\omega_t)$.
$\eta = \frac{1}{L}$.

AGD : Boucle : k:0 : nb_iter

$$\begin{cases} z' = w - \eta\nabla f(w) \\ t' = \frac{1+\sqrt{1+4t^2}}{2} \\ w' = z' + \frac{t-1}{t'} * (z'-z) \end{cases}$$

$z = z'$, $t = t'$, $w = w'$.

$\eta = 1/L$.

Slide 34.

CGD : Boucle : k : 0 → nb_iter $k^{-1} \to d$.   slide 27

pour tout $j \in \{0, .., $ nb_features$\}$.   systématiques.

$$w(j) \leftarrow w[j] - step(j) \frac{\partial f}{\partial w_j}(w)$$

$$\frac{1}{L_j}$$

SGD : Boucle : idx : 0 ↦ nb_iter

[Choisir $i \in \{0, .., n\_samples\}$.

$$w \leftarrow w - \frac{step}{\sqrt{idx+1}} \nabla f_i(w)$$

slide 40 $n = 1$.

$$d_k = \frac{1}{\sqrt{k}}.$$

SAG : Stoch. aver. grad. Desc. boucle idx : 0 → nb_iter.

[Choisir $i \in \{0, .., nb\_sample^{-1}\}$

$$y = y + \frac{\nabla f_i(w)}{n} - \frac{grad\_mem(i)}{n}.$$

$grad\_mem(i) = \nabla f_i(w)$.

$$w \leftarrow w - step\ y.$$

$step = L_{max}^{-1}$.

SVRG : Stoch. Var. Reduced grad. Desc

boucle idx : 0 → nb_iter.

Si idx % nb_sample == 0

[$w\_old = w$.
  $\mu = \nabla f(w)$

Sample i. unif $\{0, .., nb\_sample - 1\}$.

[$z' = \nabla f_i(w)$
  $z = \nabla f_i(w\_old)$.

$step = \frac{1}{L_{max}}$.

$$w \leftarrow w - step(\underbrace{\nabla f_i(w)}_{z'} - \underbrace{\nabla f_i(w\_old) + \mu}_{z})$$