Ethan Cohen

PC 3 - Questions de Cours

Exercice 7

Par double implication :

$\Leftarrow$) Si X^TX est non-singulière, alors elle est inversible et $w^* = (X^TX)^{-1}X^Ty$ est l'unique solution de $X^TXw = X^Ty$.

$\Rightarrow$) On suppose qu'on a une unique solution de $X^TXw = X^Ty$ notée w^*.

Soit $a \in \mathrm{Ker}(X^TX)$. On a : $X^TXa = 0$

D'où $X^TXa = (X^Ty - X^TXw^*)$.

$\Rightarrow X^TX(a + w^*) = X^Ty$.

Par unicité de la solution, $a + w^* = w^* \Rightarrow a = 0$ Donc $\mathrm{Ker}(X^TX) = \{0\}$.

Donc X^TX est non singulière (i.e est inversible).

D'où le résultat.

Exercice 8

Dans l'algorithme de Newton, $g(w) = J^T(w)r(w)$ et $H^k s^k = -g^k$.

En remplaçant H^k par $(J^k)^T(J^k)$, on a :

$(J^k(w_k))^T J(w_k)\, s^k = -(J^k(w_k))^T r^k$.

Avec la décomposition QR, $J^k = Q^k R^k$.

D'où $\cancel{(R^k(w_k))^T}\underbrace{(Q^k(w_k))^T Q^k(w_k)}_{Id} R^k(w_k) s^k = -\cancel{(R^k(w_k))^T}(Q^k(w_k))^T r^k$

D'où : $R^k(w_k)s^k = -Q^k(w_k)^T r^k$

Ce qui est exactement l'étape 4 de l'algorithme de Gauss-Newton. D'où le résultat.

Exercice 9

Soit $g : t \longmapsto J((1-t)v + tw)$ pour $t \in [0,1]$. On a g dérivable et $g'(t) = \langle \nabla J((1-t)v + tw, v - w \rangle$

Et $g(1) = g(0) + \int_0^1 g'(t)\,dt$

Donc $J(w) = J(v) + \int_0^1 \langle \nabla J((1-t)v + tw, v - w\rangle\, dt \quad (*)$

Par caractère L-lipschitz : $\forall t \in [0,1]$,
$\| \nabla J((1-t)v+tw) - \nabla J(v)\| \leq Lt\|w-v\|$.

Donc $\langle \nabla J((1-t)v+tw), w-v\rangle = \langle \nabla J((1-t)v+tw) - \nabla J(v), w-v\rangle + \langle \nabla J(v), w-v\rangle$

$$\underset{\text{(Cauchy-Schwarz)}}{\leq} \|\nabla J[(1-t)v+tw] - \nabla J(v)\| \cdot \|w-v\| + \langle \nabla J(v), w-v\rangle$$

$$\leq Lt\|w-v\|^2 + \langle \nabla J(v), w-v\rangle.$$

En remplaçant dans (*) :

$$J(w) \leq J(v) + \int_0^1 \left(\langle \nabla J(v), w-v\rangle + Lt\|w-v\|^2\right) dt$$
$$\leq J(v) + \langle \nabla J(v), w-v\rangle + \frac{L}{2}\|w-v\|^2$$

Exercice 10

On a : $z^{(\ell)} = W^{(\ell)} a^{(\ell-1)} + b^{(\ell)}$ et $a^{(\ell-1)} = \sigma(z^{(\ell-1)})$

De plus, $\forall j$, $\sigma(z^{(\ell-1)})_j = \sigma(z_j^{(\ell-1)})$

On veut $\left(\frac{\partial z^{(\ell)}}{\partial z^{(\ell-1)}}\right)_{ij} = \left(\frac{\partial z_i^{(\ell)}}{\partial z_j^{(\ell-1)}}\right)$.

Et $z_i^{(\ell)} = \sum_k W_{ik}^{(\ell)} a_k^{(\ell-1)} + b_i^{(\ell)}$ avec $a_k^{(\ell-1)} = \sigma(z_k^{(\ell-1)})$

Il vient $\left(\frac{\partial z^{(\ell)}}{\partial z^{(\ell-1)}}\right)_{ij} = \sum_k W_{ik}^{(\ell)} \frac{\partial a_k^{(\ell-1)}}{\partial z_j^{(\ell-1)}}$

$$= W_{ij}^{(\ell)} \frac{\partial a_j^{(\ell-1)}}{\partial z_j^{(\ell-1)}} = W_{ij}^{(\ell)} \sigma'(z_j^{(\ell-1)})$$

D'où le résultat.

Exercice 11

- Soit $D \subset \mathbb{R}^d$ un compact et $P \subset C(D)$ les polynômes

- Soit $x \in D$, on peut trivialement avoir $P \in P$ tq $P(x) \neq 0$
- Soient $\tilde{x} \neq \tilde{y} \in D$, pour $P : x \mapsto x - \tilde{x}$, on a $P(\tilde{x}) \neq P(\tilde{y})$.
- Les polynômes sont stables par addition, multiplication et multiplication par un scalaire.

- Montrons que les réseaux de neurones à une couche ne vérifient pas le point 3 (subalgebra) :

Ethan Cohen

Même si il y a fermeture sous l'addition et la multiplication par un scalaire, il n'y a pas fermeture par multiplication.

Soient $G_1(x) = \sum_{j=1}^{N_1} a_j^{(1)} \sigma(w_j^{(1)} x + b_j^{(1)})$

$G_2(x) = \sum_{j=1}^{N_2} a_j^{(2)} \sigma(w_j^{(2)} x + b_j^{(2)})$

Or quand on fait $G_1(x) \cdot G_2(x)$, on a des termes en $\sigma(w_j^{(1)} x + b_j^{(1)}) \sigma(w_k^{(2)} x + b_k^{(2)})$ et $\sigma(t)$ n'est pas fermé pour la multiplication.

On peut le voir par exemple en prenant $\sigma(t) = \frac{1}{1+e^{-t}}$.

$\sigma(x)\sigma(y) = \frac{1}{(1+e^{-x})(1+e^{-y})} \neq \frac{\alpha}{1+e^{-A}}$. d'où le résultat

Exercice 12 On procède par récurrence sur n.

Ini: $n=0$. $k \in \{0,1\}$.

Pour $k=0$, on a $H(0) = 0 - \sum_{n \geq 1} \frac{g_n(0)}{4^n}$ par lemme 1.

$= 0$

Pour $k=1$, $H(1) = H(0) + 1 - 2 \times 0 = 1$ (par (11)).

Donc pour $n=0$ on a le résultat.

Héré: Soit $n \geq 0$ tq $H\left(\frac{k}{2^n}\right) = \left(\frac{k}{2^n}\right)^2$, $0 \leq k \leq 2^n$.

Soit $k \in [0, 2^{n+1}]$ et $x = \frac{k}{2^{n+1}}$.

On a $x = \frac{1}{2}\frac{k}{2^n}$

- Si $k \in [0, 2^n]$, alors on utilise que:

$H(x) = H\left(\frac{1}{2}\frac{k}{2^n}\right) = \frac{H\left(\frac{k}{2^n}\right)}{4}$ par (10)

$= \left(\frac{k}{2^{n+1}}\right)^2$ par H. R.

- Sinon $k \in]2^n, 2^{n+1}]$ et $1 - x = \frac{2^{n+1}-k}{2^{n+1}}$,

Comme $k > 2^n$, $\frac{2^{n+1}-k}{2^n} < 1$ ie $2(1-x) < 1$ et $2^{n+1}-k \in [0, 2^n]$

Donc on a: $H(1-x) = \frac{H(2(1-x))}{4}$ par (10)

$= (1-x)^2$ par H. R

Et par 11, $H(x) = H(1-x) + 2x - 1 = (1-x)^2 + 2x - 1$

$= x^2$.

Donc dans tous les cas, $M(x) = x^2$.

Ce qui achève la récurrence et montre le résultat.

Exercice 13 On remarque que pour $x \leq -2$, $\psi(x) = 0$.

De même pour $x \geq 2$, $\psi(x) = 0$.

Donc $\phi_\ell(x) \neq 0$ pour $\left|3L\left(x - \frac{\ell}{L}\right)\right| < 2$

i.e $\frac{\ell}{L} - \frac{2}{3L} < x < \frac{\ell}{L} + \frac{2}{3L}$

i.e $Lx - \frac{2}{3} < \ell < Lx + \frac{2}{3}$

Soient $L \geq 1$ et $x \in [0,1]$. $(\phi_\ell(x) = \psi(3Lx - 3\ell))$

On a : $\sum_{\ell=0}^{L} \phi_\ell(x) = \sum_{\ell=0}^{L} \left[\sigma(3Lx - 3\ell + 2) - \sigma(3Lx - 3\ell - 1)\right] + \sum_{\ell=0}^{L} \left[\sigma(3Lx - 3\ell - 2) - \sigma(3Lx - 3\ell + 1)\right]$

$= \sum_{\ell=0}^{L} \sigma(3Lx - 3\ell + 2) - \sum_{\ell=1}^{L+1} \sigma(3Lx - 3\ell + 2) + \sum_{\ell=1}^{L+1} \sigma(3Lx - 3\ell + 1) - \sum_{\ell=0}^{L} \sigma(3Lx - 3\ell + 1$

$= \sigma(3Lx + 2) - \sigma(3L(x-1) - 1) + \sigma(3L(x-1) - 2) - \sigma(3Lx + 1)$

Or $x \leq 1$ donc $3L(x-1) - 1 \leq 0$ et $3L(x-1) - 2 < 0$ et $\left|\begin{array}{l} 3Lx + 2 > 0 \\ 3Lx - 1 > 0 \end{array}\right.$

Donc $\sum_{\ell=0}^{L} \phi_\ell(x) = \sigma(3Lx + 2) - \sigma(3Lx + 1)$

$= 3Lx + 2 - (3Lx + 1)$

$= 1$.

D'où le résultat.