Examen Master M2A (08/04/2022), Analyse numérique et réseaux de neurones. Tous documents autorisés. Le sujet est constitué de deux problèmes indépendants.

---

1 Algèbre des fonctions inversibles

• Soit une fonction $f : \mathbb{R}^m \to \mathbb{R}^m$ encodée dans un réseau de neurones à p couches cachées, avec la même taille m pour les entrées et les sortiess

$$f = f_p \circ g_{p-1} \cdots \circ g_2 \circ g_1 \circ g_0, \qquad \text{avec } g_r = L \circ f_r \text{ pour } L = \text{fonction LReLU} . \tag{1}$$

Les fonctions f_r sont linéaires sous la forme $f_r(x) = W_r x + b_r$, $x \in \mathbb{R}^m$ et $W_r \in \mathcal{M}_m(\mathbb{R})$, c'est à dire que toutes les couches ont le même nombre de neurones m. Nous supposons que toutes les matrices carrés $(W_r)_{0\leqslant r\leqslant p}$ sont **inversibles**.

1. Rappeler la forme standard de la fonction LReLU L et montrer que $L^{-1}(x) = -2L(-x)$.
2. Montrer que f est une fonction inversible de $\mathbb{R}^m$ dans lui-même.
   Nous montrons dans la suite que la fonction inverse f^{-1} s'exprime sous la même forme
   $$f^{-1}(x) = \widehat{f}_p \circ \widehat{g}_{p-1} \cdots \circ \widehat{g}_2 \circ \widehat{g}_1 \circ \widehat{g}_0, \qquad \text{avec } \widehat{g}_r = L \circ \widehat{f}_r$$
   pour des fonctions linéaires $\widehat{f}_r(x) = \widehat{W}_r x + \widehat{b}_r$ que nous allons déterminer.
3. Pour une couche cachée $p = 1$, calculer $(\widehat{W}_1, \widehat{b}_1, \widehat{W}_0, \widehat{b}_0)$ en fonction de (W_1, b_1, W_0, b_0).
4. Pour deux couches cachées $p = 2$, calculer $(\widehat{W}_2, \widehat{b}_2, \widehat{W}_1, \widehat{b}_1, \widehat{W}_0, \widehat{b}_0)$ en fonction de $(W_2, b_2, W_1, b_1, W_0, b_0)$.
5. Donner la formule générale pour $(\widehat{W}_p, \widehat{b}_p, \dots, \widehat{W}_0, \widehat{b}_0)$ en fonction de $(W_p, b_p, \dots, W_0, b_0)$.

2 Autour d'un autoencoder (linéaire)

Les autoencoders sont des réseaux de neurones dont l'objectif est de détecter la dimension dite latente de données qui vivent dans un espace de plus grande dimension.

Nous formalisons le problème en considérant une entrée $x \in \mathbb{R}^m$, une variable cachée $\theta \in \mathbb{R}^p$ avec $p \ll m$ et une sortie égale à l'entrée $y = x \in \mathbb{R}^m$. Nous faisons l'hypothèse (très) simplificatrice que la fonction u qui va de la variable cachée vers l'entrée est linéaire. Il existe une matrice M telle que

$$x = M\theta, \quad M \in \mathcal{M}_{mp}(\mathbb{R}).$$

Nous ajoutons la condition non dégénérescence $\text{rank}(M) = p$.
L'objectif d'un autoencoder est de détecter p par une procédure automatique.

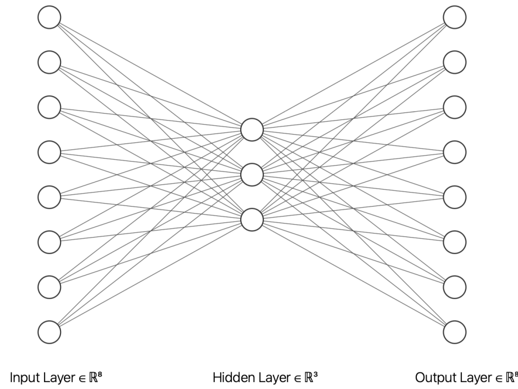

Figure 1: Structure d'un autoencoder simple

1. Pour cela on considère un réseau à une couche cachée, sans biais, avec une fonction d'activation linéaire. Montrer que la fonction correspondante peut s'écrire $f(x) = W_1 W_0 x$ avec deux matrices $W_1 \in \mathcal{M}_{mg}(\mathbb{R})$ et $W_0 \in \mathcal{M}_{gm}(\mathbb{R})$.

2. Soit la fonction coût $J(W_0, W_1) = \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i|^2$.
   Montrer que $J(W_0, W_1)$ est minimum si et seulement si
   $$M - W_1 W_0 M = 0. \tag{2}$$

3. Montrer que si $g < p$, alors l'équation (2) n'a pas de solution.

4. Que dire pour $g \geqslant p$?

5. Soit l'exemple suivant qui a été construit à partir de tests numériques. On prend $m = 150$, $p = 7$ et $x = M\theta$ avec $M \in \mathcal{M}_{150,7}(\mathbb{R})$. On a fait en sorte que le rang de M est $p = 7$. Dans la table qui suit, on affiche le résultat en fin de plusieurs training. Que peut-on dire?

g	1	2	3	4	5	6	7	8
$J(W_*)_{\text{num}}$	4.136	2.418	1.462	0.654	0.210	0.066	$7.04e-09$	$7.51e-09$