Propriétés d'approximation des réseaux de neurones profonds

Hostein Matthias - Macé Sébastien

Décembre 2022

Neurone

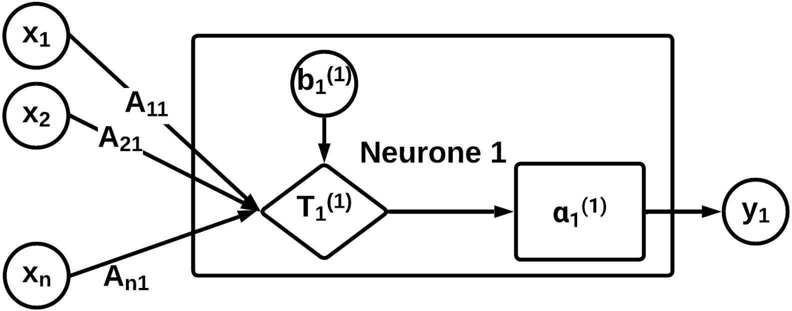

Figure 1: neurone

Couche de neurones

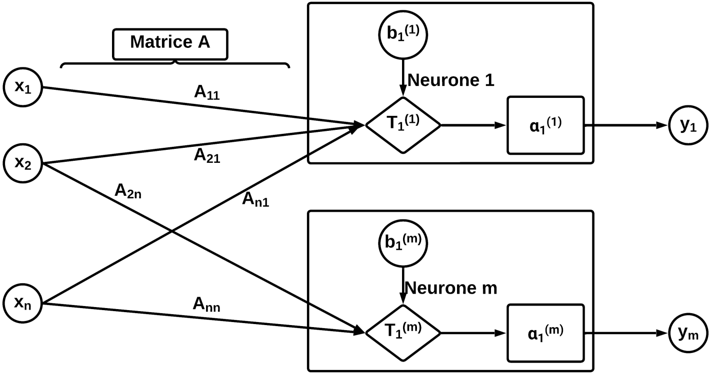

Figure 2: Couche neurone

Réseau profond

Définition (Réseau profond)

On appelle réseau de neurone profond tout L-uplet

$$\Phi = ((T_1, \alpha_1), ..., (T_L, \alpha_L)) \text{ , } où$$

- L *est la profondeur que l'on noteras également* $L(\Phi)$,
- T_l *est une fonction affine* $T_l : \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l}, x \mapsto A_l x + b_l$,
- α_l *est une fonction de* $\mathbb{R}^{N_l}$ *dans lui-même,*
- $d_{in}(\Phi) = N_0$ *et* $d_{out}(\Phi) = N_L$ *sont les dimensions d'entrée et de sortie.*

Réalisation

Définition (Réalisation d'un réseau de neurones)

La réalisation d'un tel réseau profond est la fonction :

$$R(\Phi) = \alpha_L \circ T_L \circ ...\alpha_1 \circ T_1 : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}.$$

Rq : *Cette notion permet de dissocier les réseaux de neurones qui auraient la même réalisation.*

Objectif général

On se donne

- une fonction $f : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$,
- un jeu de données d'entraînement fini $(x_i, f(x_i))_{1 \leq i \leq m}$,
- une fonction de coût L : $\left(\mathbb{R}^{N_L}\right)^2 \to \mathbb{R}$,
- éventuellement une pénalisation $\mathcal{P}$, et un coefficient $\lambda > 0$.

On cherche à résoudre

$$\min_{\Phi} \sum_{i=1}^{m} \mathrm{L}(\mathrm{R}(\Phi)(x_i), f(x_i)) + \lambda \mathcal{P}(\Phi).$$

Objectif de l'exposé

- Définir des espaces d'approximation par les réalisations des réseaux de neurones profonds,
- Étudier leurs propriétés,
- Voir quels espaces de fonctions sympathiques on peut plonger continûment dans ces espaces d'approximation (et faire mieux que le théorème d'approximation universelle).

ρ-réseau

Définition

Soit une fonction $\rho : \mathbb{R} \to \mathbb{R}$.

On construit un réseau $\Phi = ((T_1, \alpha_1), \ldots, (T_L, \alpha_L))$ avec $\alpha_L = id_{\mathbb{R}^{N_L}}$ et $\alpha_l = \bigotimes_{j=1}^{N_l} \rho_j^{(l)}$ si $l \neq L$ où $\rho_j^{(l)} \in \{\rho, id_{\mathbb{R}}\}$.

On nomme ρ-réseau un réseau de cette forme.

Il est dit strict si $\rho_j^{(l)} = \rho$.

L'opérateur $\bigotimes$ désigne l'évaluation composantes par composantes :

$$\bigotimes_{j=1}^{n} g_j : \mathbb{R}^n \to \mathbb{R}^n, (x_1, \ldots, x_n) \mapsto (g_1(x_1), \ldots, g_n(x_n)).$$

Notations

- Nombre de neurones cachés : $N(\Phi) = \sum_{l=1}^{L-1} N_l$
- nombre de connections de la couche l : $\|T_l\|_{l^0}$, le nombre de valeurs non nulles dans A_l
- nombre de connections : $W(\Phi) = \sum_{l=1}^{L-1} \|T_l\|_{l^0}$

Proposition

- $L(\Phi) \leq 1 + N(\Phi)$
- $W(\Phi) \leq (N_0 + N(\Phi))(N(\Phi) + N_L)$

Réseau constant

Lemme

Soit $\Phi = ((T_1, \alpha_1), \ldots, (T_L, \alpha_L))$ un réseau.
S'il existe $l \in \{1, \ldots, L\}$ tel que $\|T_l\|_{l^0} = 0$ alors $R(\Phi)$ est constante.

Corollaire

Si $W(\Phi) < L(\Phi)$ alors $R(\Phi)$ est constante.

Familles de réseaux

Soient $L \in \mathbb{N}^* \cup \{\infty\}$, $\rho : \mathbb{R} \to \mathbb{R}$, $W, N \in \mathbb{N} \cup \{\infty\}$ et $\Omega \subset \mathbb{R}^d$ un sous-ensemble non vide.

- $\mathcal{NN}^{\rho,d,k}_{W,L,N}$ désigne l'ensemble des ρ-réseau Φ dont l'entrée est de dimension d, la sortie de dimension k avec $N(\Phi) \leq N$, $L(\Phi) \leq L$ et $W(\Phi) \leq W$.
- $\mathcal{SNN}^{\rho,d,k}_{W,L,N}$ désigne le sous ensemble des ρ-réseaux strict.
- $\mathtt{NN}^{\rho,d,k}_{W,L,N}(\Omega) := \left\{ \mathtt{R}(\Phi)_{|\Omega} \, : \, \Phi \in \mathcal{NN}^{\rho,d,k}_{W,L,N} \right\}$ désigne alors l'ensemble des fonctions de $\mathbb{R}^d$ dans $\mathbb{R}^k$ pouvant être réalisées par des ρ-réseaux avec $N(\Phi) \leq N$, $L(\Phi) \leq L$, $W(\Phi) \leq W$.
- De même, on note $\mathtt{SNN}^{\rho,d,k}_{W,L,N}(\Omega) := \left\{ \mathtt{R}(\Phi)_{|\Omega} \, : \, \Phi \in \mathcal{SNN}^{\rho,d,k}_{W,L,N} \right\}$.

Opérations sur les réseaux

Lemme

Soient $\rho : \mathbb{R} \to \mathbb{R}$, $d, k \in \mathbb{N}^*$, $c = \min(d, k)$, $\Phi \in \mathcal{NN}^{\rho,d,k}$ *et* $L_0 \in \mathbb{N}^*$.
Il existe $\Psi \in \mathcal{NN}^{\rho,d,k}$ *tel que* $\mathcal{R}(\Psi) = \mathcal{R}(\Phi)$, $L(\Psi) = L(\Phi) + L_0$, $N(\Psi) = N(\Phi) + cL_0$ *et* $W(\Psi) = W(\Phi) + cL_0$.

Opérations sur les réseaux

Soient $d, k, n \in \mathbb{N}^*$, $c \in \mathbb{R}$, $\rho : \mathbb{R} \to \mathbb{R}$ et $k_i \in \mathbb{N}$ pour $i = 1, \dots, n$.

1. Si $\Phi \in \mathtt{NN}^{\rho,d,k}$ alors $c \cdot \mathtt{R}(\Phi) = \mathtt{R}(\Psi)$ où $\Psi \in \mathtt{NN}^{\rho,d,k}$ avec $W(\Psi) \leq W(\Phi)$, $L(\Psi) = L(\Phi)$ et $N(\Psi) = N(\Phi)$. De même pour $\mathtt{SNN}^{\rho,d,k}$.
2. Si $\Phi_i \in \mathtt{NN}^{\rho,d,k_i}$ pour $i = 1, \dots, n$ alors $(\mathtt{R}(\Phi_1), \dots, \mathtt{R}(\Phi_n)) = \mathtt{R}(\Psi)$ où $\Psi \in \mathtt{NN}^{\rho,d,k}$ avec $W(\Psi) \leq \delta + \sum_{i=1}^{n} W(\Phi_i)$, $L(\Psi_i) = \max_{1 \leq i \leq n} L(\Phi_i)$ et $N(\Psi) \leq \delta + \sum_{i=1}^{n} N(\Phi_i)$ et $c = \min(d, K - 1)$ et $\delta = c(\max_{1 \leq i \leq n} L(\Phi_i) - \min_{1 \leq i \leq n} L(\Phi_i))$.
3. Si $\Phi_i \in \mathtt{NN}^{\rho,d,k_i}$ pour $i = 1, \dots, n$ alors $\sum_{i=1}^{n} \mathtt{R}(\Psi_i) = \mathtt{R}(\Psi)$ où $\Psi \in \mathtt{NN}^{\rho,d,k}$ avec $W(\Psi) \leq \delta + \sum_{i=1}^{n} W(\Phi_i)$, $L(\Psi_i) = \max_{1 \leq i \leq n} L(\Phi_i)$ et $N(\Psi) \leq \delta + \sum_{i=1}^{n} N(\Phi_i)$ et $c = \min(d, K - 1)$ et $\delta = c(\max_{1 \leq i \leq n} L(\Phi_i) - \min_{1 \leq i \leq n} L(\Phi_i))$.

Opérations sur les réseaux

Soient $d, d_1, d_2, k, k_1 \in \mathbb{N}^*$ et $\rho : \mathbb{R} \to \mathbb{R}$.

1. Si $\Phi \in \mathtt{NN}^{\rho,d,k}$ et $P : \mathbb{R}^{d_1} \to \mathbb{R}^d$, $Q : \mathbb{R}^k \to \mathbb{R}^{k_1}$ sont deux fonctions affines alors $Q \circ \mathtt{R}(\Phi) \circ P = \mathtt{R}(\Psi)$ où $\Psi \in \mathtt{NN}^{\rho,d_1,k_1}$ avec $W(\Psi) \leq \|Q\|_{l^{0,\infty}} W(\Phi) \|P\|_{l^{0,\infty}}$, $L(\Psi) = L(\Phi)$ et $N(\Psi) = N(\Phi)$. De même pour $\mathtt{SNN}^{\rho,d,k}$.
2. Si $\Phi_1 \in \mathtt{NN}^{\rho,d,d_1}$ et $\Phi_2 \in \mathtt{NN}^{\rho,d_1,d_2}$ alors $\mathtt{R}(\Phi_2) \circ \mathtt{R}(\Phi_1) = \mathtt{R}(\Psi)$ où $\Psi \in \mathtt{NN}^{\rho,d,d_2}$ avec $W(\Psi) = W(\Phi_2) + W(\Phi_1)$, $L(\Psi) = L(\Phi_2) + L(\Phi_1)$ et $N(\Psi) = N(\Phi_2) + N(\Phi_1) + d_1$.
3. sous les mêmes conditions, il existe $\Psi' \in \mathtt{NN}^{\rho,d,d_2}$ tel que $\mathtt{R}(\Phi_2) \circ \mathtt{R}(\Phi_1) = \mathtt{R}(\Psi)$ avec $W(\Psi') \leq W(\Phi_1) + \max(W(\Phi_1), d) W(\Phi_2)$, $L(\Psi) = L(\Phi_2) + L(\Phi_1) - 1$ et $N(\Psi) = N(\Phi_2) + N(\Phi_1) + d_1$. De même pour $\mathtt{SNN}^{\rho,d,k}$.

Représentation

Définition

On dit que $\rho : \mathbb{R} \to \mathbb{R}$ *peut représenter une fonction* $f : \mathbb{R} \to \mathbb{R}$ *en* n *termes si* $f \in \mathtt{NN}^{\rho,1,1}_{\infty,2,n}$. *Il existe ainsi* $c, a_1, b_1, c_1, \dots, a_n, b_n, c_n \in \mathbb{R}$ *tels que :*

$$f(x) = c + \sum_{i=1}^{n} a_i \cdot \rho(b_i x + c_i), \text{ pour tout } x \in \mathbb{R}.$$

Lemme

Pour tout $r \in \mathbb{N}$, $\rho_r : x \mapsto (\max(0, x))^r$ *peut représenter tout polynôme de degré* $\leq r$ *en* $2r + 2$ *termes.*

Lemme

Si ρ peut représenter l'identité en n termes alors pour tout $d, k \in \mathbb{N}^*$, $W, N \in \mathbb{N}$ et $L \in \mathbb{N}^* \cup \{\infty\}$, on a

$$\mathtt{NN}^{\rho,d,k}_{W,L,N} \subset \mathtt{SNN}^{\rho,d,k}_{n^2 W,L,nN}.$$

Lemme

Si $\rho : \mathbb{R} \to \mathbb{R}$ peut représenter tous les polynômes de degré 2 en n termes alors

1. *pour tout $d \geq 2$, la fonction produit M_d sur $\mathbb{R}^d$ vérifie*

$$M_d \in \mathtt{NN}^{\rho,d,1}_{6n(2^j-1),2j,(2n+1)(2^j-1)-1} \text{ où } j = \lceil \log_2 d \rceil.$$

2. *pour tout $k \in \mathbb{N}$, la fonction multiplication m_k sur $\mathbb{R} \times \mathbb{R}^k$ satisfait*

$$m_k \in \mathtt{NN}^{\rho,k+1,k}_{6kn,2,2kn}.$$

Quasi-norme, quasi-espace de Banach

Définition

Soit X un $\mathbb{K}$-espace vectoriel. On dit qu'une application $\|\cdot\| : X \longrightarrow \mathbb{R}^+$ est une quasi-norme si elle vérifie les propriétés suivantes :

1. *Pour* $x \in X$, $\|x\| = 0 \Rightarrow x = 0_X$,
2. $\forall \lambda \in \mathbb{K},\ \forall x \in X, \quad \|\lambda x\| = |\lambda| \|x\|$,
3. $\exists K > 0,\ \forall x, y \in X, \quad \|x + y\| \leq K(\|x\| + \|y\|)$.

L'application $\|\cdot\|$ induit donc une topologie et une notion de suites de Cauchy sur X. $(X, \|\cdot\|)$ est alors appelé quasi-espace de Banach si les suites de Cauchy convergent pour cette topologie.

Définitions

Soient X un quasi-espace de Banach muni d'une quasi-norme $\|\cdot\|_X$ et $f \in X$.
On appelle erreur de meilleure approximation de f par un ensemble non vide $\Gamma \in X$ la quantité :

$$E(f,\Gamma)_X = \inf_{g\in\Gamma} \|f - g\|_X \in [0,\infty).$$

Pour une famille $\Sigma = (\Sigma_n)_{n\in\mathbb{N}}$ d'ensemble non vide de X, on définit pour tout $f \in X$, $\alpha \in (0,\infty)$ et $q \in (0,\infty]$ la quantité :

$$\|f\|_{A_q^\alpha(X,\Sigma)} = \begin{cases} \left(\sum_{n=1}^{\infty} [n^\alpha E(f,\Sigma_{n-1})_X]^q \frac{1}{n}\right)^{\frac{1}{q}} & \text{si } q < \infty \\ \sup_{n\in\mathbb{N}^*} [n^\alpha E(f,\Sigma_{n-1})_X] & \text{si } q = \infty. \end{cases}$$

Conditions

Dans la suite, on s'intéressera aux familles Σ vérifiant les conditions :

1. $\Sigma_0 = \{0\}$
2. $\Sigma_n \subset \Sigma_{n+1}$ pour tout $n \in \mathbb{N}$
3. $a \cdot \Sigma_n = \Sigma_n$ pour tout $a \in \mathbb{R} \setminus \{0\}$ et tout $n \in \mathbb{N}$
4. il existe une constante $c \in \mathbb{N}^*$ tel que $\Sigma_n + \Sigma_n \subset \Sigma_{cn}$ pour tout $n \in \mathbb{N}$
5. $\Sigma_\infty = \bigcup_{n \in \mathbb{N}} \Sigma_n$ est dense dans X
6. pour tout $n \in \mathbb{N}$, tout $f \in X$ a une meilleur approximation par Σ_n (i.e. le inf sur Σ_n est en fait un min).

Définition

On définit la classe d'approximation

$$A_q^\alpha(X,\Sigma) = \{f \in X \ : \ \|f\|_{A_q^\alpha(X,\Sigma)} < \infty\}.$$

Proposition

Si les cinq premières conditions sont vraies alors les classes $\left(A_q^\alpha(X,\Sigma), \|\cdot\|_{A_q^\alpha(X,\Sigma)}\right)$ *sont des quasi-Banach vérifiant* $A_q^\alpha(X,\Sigma) \hookrightarrow X$ *et*

$$A_q^\alpha(X,\Sigma) \hookrightarrow A_s^\beta(X,\Sigma) \text{ si } \alpha > \beta \text{ ou } (\alpha = \beta \text{ et } q \leq s).$$

Fonction de profondeur croissante

Définition

Une fonction de profondeur croissante est une fonction $\mathcal{L} : \mathbb{N} \to \mathbb{N} \cup \{\infty\}$ *croissante.*

Définition

Soit ρ *une fonction d'activation, une fonction de profondeur croissante* $\mathcal{L}$*, un sous-ensemble* $\Omega \subset \mathbb{R}^d$ *et un quasi-Banach* X *composé de fonctions de* Ω *dans* $\mathbb{R}^k$*. On défini* $\mathtt{N}_0(X, \rho, \mathcal{L}) = \mathtt{W}_0(X, \rho, \mathcal{L}) := \{0\}$ *et*

- $\mathtt{W}_n(X, \rho, \mathcal{L}) := \mathtt{NN}^{\rho,d,k}_{n,\mathcal{L}(n),\infty} \cap X$,
- $\mathtt{N}_n(X, \rho, \mathcal{L}) := \mathtt{NN}^{\rho,d,k}_{\infty,\mathcal{L}(n),n} \cap X$,

et de même pour les ρ*-réseaux stricts :* $\mathtt{SW}_n(X, \rho, \mathcal{L})$ *et* $\mathtt{SN}_n(X, \rho, \mathcal{L})$*.*

Approximation

Définition

On définit les espaces d'approximation :

- $SW_q^\alpha(X,\rho,\mathcal{L}) = A_q^\alpha(X,\Sigma)$ où $\Sigma = (\mathtt{SW}_n(X,\rho,\mathcal{L}))_{n\in\mathbb{N}}$,
- $SN_q^\alpha(X,\rho,\mathcal{L}) = A_q^\alpha(X,\Sigma)$ où $\Sigma = (\mathtt{SN}_n(X,\rho,\mathcal{L}))_{n\in\mathbb{N}}$.

Plongement

Proposition

Soient ρ une fonction d'activation, $\mathcal{L}$ une fonction de profondeur croissante, un sous-ensemble $\Omega \subset \mathbb{R}^d$ et un quasi-Banach X composé de fonctions de Ω dans $\mathbb{R}^k$. Pour tout $\alpha > 0$ et tout $q \in (0, \infty]$, on a :

$$\|\cdot\|_{W_q^\alpha(X,\rho,\mathcal{L})} \leq \|\cdot\|_{SW_q^\alpha(X,\rho,\mathcal{L})} \text{ et } \|\cdot\|_{N_q^\alpha(X,\rho,\mathcal{L})} \leq \|\cdot\|_{SN_q^\alpha(X,\rho,\mathcal{L})}.$$

On a ainsi

$$SW_q^\alpha(X,\rho,\mathcal{L}) \hookrightarrow W_q^\alpha(X,\rho,\mathcal{L}) \text{ et } SN_q^\alpha(X,\rho,\mathcal{L}) \hookrightarrow N_q^\alpha(X,\rho,\mathcal{L})$$

.

Égalité

Proposition

On a l'égalité des ensembles et l'équivalence des normes si l'une des deux conditions suivantes est vérifiée:

- Ω *est bornée,* ρ *est continue et dérivable en un certain* $x_0 \in X$ *avec* $\rho'(x_0) \neq 0$
- ρ *peut représenter l'identité en* m *termes pour un certain* $m \in \mathbb{N}$.

Lemme

Soient $\rho : \mathbb{R} \to \mathbb{R}$ et $\mathcal{L}$ une fonction de profondeur croissante. Les ensembles Σ satisfont les conditions 1 à 4 avec $c = 2 + \min(d, k)$.

Lemme

Soient $\Omega \subset \mathbb{R}^d$ un borélien de mesure non nulle, $p \in (0, +\infty]$ et $k \in \mathbb{N}^*$.

L'ensemble $L_p(\Omega, \mathbb{R}^k) = \begin{cases} \{f_{|\Omega} : f \in L^p(\mathbb{R}^d, \mathbb{R}^k)\} & \text{si } p < \infty \\ \{f_{|\Omega} : f \in C_0(\mathbb{R}^d, \mathbb{R}^k)\} & \text{si } p = \infty. \end{cases}$ *muni de la quasi-norme* $\|\cdot\|_{L^p}$ *est un quasi-espace de Banach.*

Dégénérescence

Définition

Un fonction d'activation $\rho : \mathbb{R} \to \mathbb{R}$ est dite non dégénérée si elle est borélienne, localement bornée, continue en dehors d'un fermé de mesure nulle et non polynomiale presque partout.

Rq : *Sous cette hypothèse supplémentaire, on évite des cas indésirables où la propriété 5 n'est pas vérifiée.*

Théorème de densité

Soit ρ une fonction d'activation borélienne localement bornée, $\mathcal{L}$ une fonction de profondeur croissante et $p \in (0, \infty]$. Notons $L = \sup\{\mathcal{L}(n) : n \in \mathbb{N}\}$.

1. Soit $\Omega \subset \mathbb{R}^d$ un borélien de mesure non nulle et supposons $L \geq 2$:
   - Si $p \in (0, \infty)$ alors $\mathtt{NN}^{\rho,d,k}_{\infty,\infty,\infty} \subset L_p(\Omega, \mathbb{R}^k)$ et si ρ est non dégénérée alors $\Sigma_\infty(L_p(\Omega, \mathbb{R}^k))$ est dense dans $L_p(\Omega, \mathbb{R}^k)$,
   - si $p = \infty$ et si ρ est continue alors les propriétés sont également vraies.
2. Supposons que la fermeture de $\mathtt{NN}^{\rho,d,1}_{\infty,L,\infty} \cap L_p(\mathbb{R}^d, \mathbb{R})$ dans L_p contienne un élément g telle que :
   - il existe μ décroissante et positive telle que $\int\limits_{\mathbb{R}^d} \mu(|x|)dx < \infty$ et $g \leq \mu \circ |\cdot|$
   - $\int\limits_{\mathbb{R}^d} g(x)dx \neq 0$

   Alors $\Sigma_\infty(L_p(\Omega, \mathbb{R}^k))$ est dense dans $L_p(\Omega, \mathbb{R}^k)$, pour tout $\Omega \subset \mathbb{R}^d$ borélien de mesure non nulle et tout $k \in \mathbb{N}$.

Corollaires

Corollaire

La condition 5 est vérifiée pour tout $\Omega \subset \mathbb{R}^d$ borélien borné de mesure non nulle et tout $p \in (0, \infty]$ dès que $L \geq 2$ et ρ est continue et non polynomiale.

Corollaire

La condition 5 est vérifiée pour tout $\Omega \subset \mathbb{R}^d$ borélien de mesure non nulle et tout $p \in (0, \infty]$ dès que $L \geq 2$ et $\mathtt{NN}^{\rho,d,1}_{\infty,L,\infty}$ contient une fonction g positive, non nulle à support compact et bornée.

Théorème

Soit ρ *une fonction d'activation,* $\mathcal{L}$ *une fonction de profondeur croissante,* $d \in \mathbb{N}$*,* $p \in (0, \infty]$ *et* $\Omega \subset \mathbb{R}^d$ *un borélien de mesure non nulle. Notons* $L = \sup\{\mathcal{L}(n) : n \in \mathbb{N}\}$*. Supposons que l'une des deux conditions suivantes est vérifiée :*

- ρ *est continue et non polynomial,* $L \geq 2$ *et* Ω *est bornée*
- $\mathcal{NN}^{\rho,d,1}_{\infty,L,\infty} \cap L_p(\mathbb{R}^d, \mathbb{R})$ *contient une fonction g positive, non nulle à support compact et bornée.*

Alors pour tout $k \in \mathbb{N}$*,* $\alpha > 0$*,* $q \in (0, \infty]$ *et* $X = L_p(\mathbb{R}^d, \mathbb{R})$*, on a*

- *Les cinq premières conditions sur les familles* $\Sigma = (\mathcal{SW}_n(X, \rho, \mathcal{L}))_{n\in\mathbb{N}}$ *ou* $\Sigma = (\mathcal{SN}_n(X, \rho, \mathcal{L}))_{n\in\mathbb{N}}$ *sont vérifiées*
- *les espaces* $\left(W^\alpha_q(X, \rho, \mathcal{L}), \|\cdot\|_{W^\alpha_q(X,\rho,\mathcal{L})}\right)$ *et* $\left(N^\alpha_q(X, \rho, \mathcal{L}), \|\cdot\|_{N^\alpha_q(X,\rho,\mathcal{L})}\right)$ *sont quasi-Banach.*

Exemple particulier : ReLU et ses puissances

Théorème

Soit $r \in \mathbb{N}^*$ et soit $\mathcal{L}$ une fonction de profondeur croissante. On pose : $\rho_r : x \mapsto (x_+)^r$. Dans le cas où $X = L_p(\Omega, \mathbb{R}^k)$, on a :

1. $SW_q^\alpha(X, \rho_r, \mathcal{L}) = W_q^\alpha(X, \rho_r, \mathcal{L})$ *et* $SN_q^\alpha(X, \rho_r, \mathcal{L}) = N_q^\alpha(X, \rho_r, \mathcal{L})$ *avec équivalence des normes sur ces espaces.*
2. *Si $\mathcal{L}$ vérifie :*

$$\sup_{n \in \mathbb{N}^*} \mathcal{L}(n) \geq \begin{cases} 2 & \text{si } \Omega \text{ est borné, ou si } d = 1, \\ 3 & \text{sinon} \end{cases}$$

alors les propriétés 1 à 5 sont vérifiées pour $\Sigma = (\mathtt{W}_n(X, \rho_r, \mathcal{L}))_{n \in \mathbb{N}}$ (resp. $\Sigma = (\mathtt{N}_n(X, \rho_r, \mathcal{L}))_{n \in \mathbb{N}}$), faisant de l'espace $W_q^\alpha(X, \rho_r, \mathcal{L})$ (resp. $N_q^\alpha(X, \rho_r, \mathcal{L})$) un quasi-espace de Banach

ReLU et ses puissances : pour pouvoir bien approximer, il faut creuser !

Théorème

Soient $\Omega \subset \mathbb{R}^d$ un ouvert de mesure non-nulle, $p, q \in (0. + \infty]$ et $L \in \mathbb{N}^*$. Alors, en prenant $X = L_p(\Omega, \mathbb{R})$, on a :

- $\mathcal{C}_c^3(\Omega) \cap W_q^\alpha(X, \rho_1, L) \neq \{0\} \Rightarrow \left\lfloor \frac{L}{2} \right\rfloor \geq \frac{\alpha}{2}$.
- $\mathcal{C}_c^3(\Omega) \cap N_q^\alpha(X, \rho_1, L) \neq \{0\} \Rightarrow L - 1 \geq \frac{\alpha}{2}$.

Cela veut dire que, pour espérer plonger des espaces de fonctions classiques dans nos espaces d'approximation, il faut avoir un nombre de couches grand !

ReLU et ses puissances : dans certains cas, on peut espérer.

Proposition

Soient $d, r, L \in \mathbb{N}^*$ *et soit* $\Omega \subset \mathbb{R}^d$ *ouvert de mesure non-nulle. Alors, sous les hypothèses :*

- $r \geq 4$ *et* $L \geq 2$ *ou* $r \in \{2, 3\}$ *et* $L \geq 3$, *si* $d = 1$ *ou*
- $r \geq 4$ *et* $L \geq 3$ *ou* $r \in \{2, 3\}$ *et* $L \geq 5$ *si* $d > 1$, *on a que pour* $\alpha > 0$ *et* $q \in (0, +\infty]$, $N_q^\alpha(X, \rho_r, L) \cap \mathcal{C}_c^3(\Omega) \neq \{0\}$ *et* $W_q^\alpha(X, \rho_r, L) \cap \mathcal{C}_c^3(\Omega) \neq \{0\}$

Définition d'un espace de Besov

Définition (Module de régularité)

Soient $\Omega \subset \mathbb{R}^d$ *un ouvert,* $f \in L^p(\Omega)$ *et* $h \in \mathbb{R}^d$*. On introduit les quantités suivantes :*

- τ_h *désigne l'opérateur de translation par* h*,*
- $\Delta_r^h := (\tau_h - I)^r$*,*
- *Pour* $x \in \mathbb{R}^d$*,* $\Delta_r^h(f, x, \Omega) := \begin{cases} \Delta_r^h f(x) & \text{si } x + ih \in \Omega \text{ pour tout } i \in [\![0, r]\!], \\ 0 & \text{sinon.} \end{cases}$

Le module de régularité, noté alors $\omega_r(f, t, \Omega)_p$ *est défini ainsi :*

$$\omega_r(f, t, \Omega)_p := \sup_{\|h\| \leq t} \|\Delta_r^h(f, \cdot, \Omega)\|_{L^p(\Omega)}$$

Définition d'un espace de Besov

Définition (Espace de Besov)

Soient $r \in \mathbb{N}^*$, $\alpha \in (0, r)$ *et* $p, q \in (0, +\infty]$. *L'espace de fonctions suivant :*

$$B_q^\alpha(L^p(\Omega)) = \{f \in L^p(\Omega) \;/\; |f|_{B_q^\alpha(L^p(\Omega))} < +\infty\}$$

où

$$|f|_{B_q^\alpha(L^p(\Omega))} := \begin{cases} \left(\int_0^1 |t^{-\alpha}\omega_r(f, t, \Omega)_p|^q \frac{dt}{t}\right)^{\frac{1}{q}} & si\ q < +\infty \\ \sup_{t \in [0,1]} |t^{-\alpha}\omega_r(f, t, \Omega)_p| & si\ q = +\infty \end{cases}$$

est appelé espace de Besov. Il est parfois noté $B_{p,q}^\alpha(\Omega)$.

Plongement des espaces de Besov dans nos espaces d'approximation

Définition (Domaine lipschitzien)

Soit $\Omega \subset \mathbb{R}^n$ *un ouvert connexe tel que* $\partial\Omega$ *soit borné.* Ω *est appelé domaine lipschitzien si, pour tout* $p \in \partial\Omega$*, il existe un rayon* $r > 0$ *et un homéomorphisme* $h_p : B_r(p) \longrightarrow B_1(0)$ *tels que :*

1. h_p *et* h_p^{-1} *soient lipschitziennes*
2. $h_p(B_r(p) \cap \partial\Omega) = B_0$
3. $h_p(B_r(p) \cap \Omega) = B_+$

où $B_0 := \{\boldsymbol{x} \in B_1(0) \ / \ x_n = 0\}$ *et* $B_+ := \{\boldsymbol{x} \in B_1(0) \ / \ x_n > 0\}$

Plus simplement, un domaine Lipschitzien est un ouvert qui peut se voir localement comme l'hypographe strict d'une fonction lipschitzienne.

Plongement des espaces de Besov dans nos espaces d'approximation

Théorème

Soit $\Omega \subset \mathbb{R}^d$ un domaine lipschitzien borné de mesure non-nulle et soit $\mathcal{L}$ une fonction de profondeur croissante. Alors :

1. *Si $d = 1$, et si $L := \sup_{n \in \mathbb{N}^*} \mathcal{L}(n) \geq 2$, alors, quelque soit $r \in \mathbb{N}^*$, on a :*

$$\forall p, q \in (0, +\infty],\ 0 < s < \min(1, 1/p), \quad B^s_{p,q}(\Omega) \hookrightarrow W^s_q(L_p(\Omega, \mathbb{R}), \rho_r, \mathcal{L}).$$

2. *Si $d > 1$, et $L \geq 3$, alors, pour tout $r \in \mathbb{N}^*$, en posant $r_0 = r$ si $r \geq 2$ et $L \geq 2 + 2\lceil \log_2(d) \rceil$ et $r_0 = 0$ sinon, on a :*

$$\forall p, q \in (0, +\infty]\ 0 < s < \frac{r_0 + \min(1, 1/p)}{d}, \quad B^{sd}_{p,q}(\Omega) \hookrightarrow W^s_q(L_p(\Omega, \mathbb{R}), \rho_r, \mathcal{L}).$$