Propriétés d'approximation des réseaux de neurones profonds

Florian Tilliet et Perrine Jouteur

Octobre 2020

Réseaux de neurones

Définition : Réseau de neurones

Soit $\sigma : \mathbb{R} \to \mathbb{R}$ une fonction d'activation.
Réseaux de neurones : $((T_1, \alpha_1), ..., (T_L, \alpha_L))$ avec, pour $1 \leq l \leq L$:

$$T_l : \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l} \text{ (linéaire)}$$

$$\alpha_l : \mathbb{R}^{N_l} \to \mathbb{R}^{N_l} \text{ (non linéaire)}$$

Réseaux de neurones

Définition : Réseau de neurones

Soit $\sigma : \mathbb{R} \to \mathbb{R}$ une fonction d'activation.
Réseaux de neurones : $((T_1, \alpha_1), ..., (T_L, \alpha_L))$ avec, pour $1 \leq l \leq L$:

$$T_l : \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l} \text{ (linéaire)}$$

$$\alpha_l : \mathbb{R}^{N_l} \to \mathbb{R}^{N_l} \text{ (non linéaire)}$$

Définition : Réalisation

La réalisation d'un réseau de neurones est la fonction :

$$R := \alpha_L \circ T_L \circ ... \circ \alpha_1 \circ T_1$$

Complexité d'un réseau de neurones

Définition : Profondeur, nombre de neurones et de connections

Soit $\phi := ((T_1, \alpha_1), ..., (T_L, \alpha_L))$ avec, pour $1 \leq l \leq L$, $T_l : \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l}$

- Profondeur : $L(\phi) := L$
- Nombre de neurones cachés : $N(\phi) := \sum_{l=1}^{L-1} N_l$
- Nombre de connections := Nombre de coefficients non nuls dans les matrices A_l tels que $T_l(x) = A_l x + b_l$

Notations

Soient $\Omega \subset \mathbb{R}^d$, borélien et de mesure non nulle et σ, une fonction d'activation.

- $\mathtt{NN}^{\sigma,d,k}_{W,L,N}(\Omega)$: les réalisations de réseaux de neurones généralisés, de Ω dans $\mathbb{R}^k$, avec

$$W(\phi) \leq W, \ N(\phi) \leq N \text{ et } L(\phi) \leq L$$

- $\mathtt{SNN}^{\sigma,d,k}_{W,L,N}$: les réalisations de réseaux de neurones stricts, de Ω dans $\mathbb{R}^k$, avec

$$W(\phi) \leq W, \ N(\phi) \leq N \text{ et } L(\phi) \leq L$$

Définition : famille d'approximation

Soit $(X, ||\cdot||)$ un quasi-Banach, dont les éléments sont des fonctions $f : \Omega \to \mathbb{R}^k$.

- Fonction de croissance en profondeur : $\mathfrak{L} : \mathbb{N} \to \mathbb{N}$, croissante.
- Familles d'approximation : pour $n \in \mathbb{N}^*$,

$$\Sigma_n := \mathtt{NN}^{\sigma,d,k}_{n,\mathfrak{L}(n),+\infty}(\Omega) \cap X$$

$$S\Sigma_n := \mathtt{SNN}^{\sigma,d,k}_{n,\mathfrak{L}(n),+\infty}(\Omega) \cap X$$

Définition : classes d'approximation

- (Future) quasi-norme :

$$||f||_{W_q^\alpha} := \left(\sum_{n=1}^{+\infty} (n^\alpha E(f, \Sigma_{n-1}))^q \frac{1}{n} \right)^{1/q}$$

$$||f||_{W_\infty^\alpha} := sup_{n \geq 1} \left(n^\alpha E(f, \Sigma_{n-1}) \right)$$

où $E(f, \Sigma_n) := inf_{g \in \Sigma_n} ||f - g||$

- Classe d'approximation :

$$W_q^\alpha(X, \sigma, \mathfrak{L}) := \{ f \in X, ||f||_{W_q^\alpha} < +\infty \}$$

Équivalence des réseaux généralisés et stricts

Théorème

∘ Si Ω est borné, σ continue et différentiable en un point, de différentielle non nulle ;
∘ Ou si σ peut représenter l'identité id : $\mathbb{R} \to \mathbb{R}$;

Alors : $\forall k \in \mathbb{N}$, $\forall p, q \in]0, \infty]$, $\forall \alpha > 0$, $\forall \mathfrak{L}$ fonction de croissance en profondeur,

$$SW_q^\alpha(L^p(\Omega), \sigma, \mathfrak{L}) = W_q^\alpha(L^p(\Omega), \sigma, \mathfrak{L})$$

Et il existe $C > 0$, telle que

$$|| \cdot ||_{W_q^\alpha} \leqslant || \cdot ||_{SW_q^\alpha} \leqslant C|| \cdot ||_{W_q^\alpha}$$

Complétude des classes d'approximation

Propriété

Si la famille d'approximation $(\Sigma_n)_n$ vérifie les 5 conditions ci-dessous, alors la classe d'approximation $W_q^\alpha(X, \sigma, \mathfrak{L})$, munie de $||\cdot||_{q,\alpha}$, est un quasi-Banach :

1. $\Sigma_0 = \{0\}$
2. Pour tout $n \geq 0$, $\Sigma_n \subset \Sigma_{n+1}$
3. Pour tout $n \geq 1$, pour tout $a \in \mathbb{R}^*$, $a \cdot \Sigma_n = \Sigma_n$
4. Il existe $c \in \mathbb{N}$ tel que pour tout $n \in \mathbb{N}$, $\Sigma_n + \Sigma_n \subset \Sigma_{cn}$
5. $\Sigma_\infty := \bigcup_{n\geq 0} \Sigma_n$ est dense dans X

Approximation universelle : cas continu

Théorème

Soit σ une fonction d'activation non dégénérée. Alors pour toute fonction $f : \mathbb{R}^d \to \mathbb{R}$, continue, il existe une suite de fonctions $(g_n)_n$ dans $\mathtt{NN}^{\sigma,d,1}_{n,2,+\infty}(\mathbb{R}^d)$ qui converge uniformément sur tout compact vers f.

Démonstration : cf l'exposé de Maxence et Antoine sur l'introduction aux réseaux de neurones artificiels.

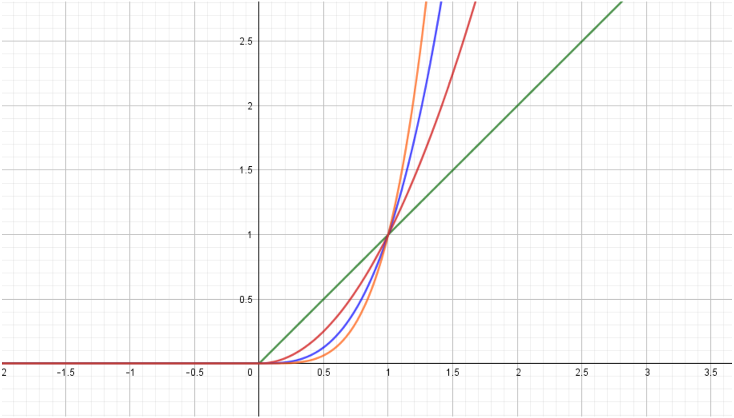
Approximation universelle : cas L^p

Théorème

Soient σ une fonction d'activation non dégénérée, Ω un borélien de mesure non nulle, $p > 0$ et $\mathfrak{L}$ une fonction de croissance en profondeur, avec $L := \sup_{n\in\mathbb{N}} \mathfrak{L}(n)$.

- Si Ω est borné et $L \geq 2$;
- Ou s'il existe $g \in \overline{\mathtt{NN}^{\sigma,d,1}_{\infty,L,\infty} \cap L^p(\mathbb{R}^d)}$ telle que :
  - $\exists\ \mu : \mathbb{R}_+^* \to \mathbb{R}_+^*$ croissante, telle que $\int_{\mathbb{R}^d} \mu(|x|)dx < \infty$ et $g \leqslant \mu(|\cdot|)$;
  - $\int_{\mathbb{R}^d} g(x)dx \neq 0$;

Alors $\Sigma_\infty(L^p(\Omega, \mathbb{R}^k), \sigma, \mathfrak{L})$ est dense dans $L^p(\Omega, \mathbb{R}^k)$.



Les fonctions ReLU pour $r = 1, 2, 3, 4$

Complétude des classes d'approximation

Propriété

Si la famille d'approximation $(\Sigma_n)_n$ vérifie les 5 conditions ci-dessous, alors la classe d'approximation $W_q^\alpha(X, \sigma, \mathfrak{L})$, munie de $|| \cdot ||_{q,\alpha}$, est un quasi-Banach :

1. $\Sigma_0 = \{0\}$
2. Pour tout $n \geq 0$, $\Sigma_n \subset \Sigma_{n+1}$
3. Pour tout $n \geq 1$, pour tout $a \in \mathbb{R}^*$, $a \cdot \Sigma_n = \Sigma_n$
4. Il existe $c \in \mathbb{N}$ tel que pour tout $n \in \mathbb{N}$, $\Sigma_n + \Sigma_n \subset \Sigma_{cn}$
5. $\Sigma_\infty := \bigcup_{n \geq 0} \Sigma_n$ est dense dans X

Degré de régularité d'une fonction

Soit $\Omega := [0,1]^d \subset \mathbb{R}^d$, et $f \in L^p(\Omega)$.

- Opérateurs de différences : pour $r \in \mathbb{N}^*$, $h \in \mathbb{R}^d$, $\Delta_h^r(f)(x) = (f(x+h) - f(x))^r \times \mathbb{1}_\Omega(x)$
- Degré de régularité d'ordre r :

$$\omega_r(f,t)_p := sup_{|h| \leqslant t} ||\Delta_h^r(f)(\cdot)||_p$$

Degré de régularité d'une fonction

Soit $\Omega := [0,1]^d \subset \mathbb{R}^d$, et $f \in L^p(\Omega)$.

- Opérateurs de différences : pour $r \in \mathbb{N}^*$, $h \in \mathbb{R}^d$, $\Delta_h^r(f)(x) = (f(x+h) - f(x))^r \times \mathbb{1}_\Omega(x)$
- Degré de régularité d'ordre r :

$$\omega_r(f,t)_p := sup_{|h| \leqslant t} ||\Delta_h^r(f)(\cdot)||_p$$

Définition : Espace de Besov

$B_q^\alpha(L^p(\Omega))$: fonctions f telles que

$$|f|_{B_q^\alpha(L^p(\Omega))} := \left(\int_0^1 (t^{-\alpha}\omega_r(f,t)_p)^q dt/t \right)^{1/q} < +\infty$$

Lien entre espaces d'approximation et espaces de Besov

Rappel : pour les classes d'approximations $W_q^\alpha(X, \sigma, \mathfrak{L})$, on avait :

$$||f||_{W_q^\alpha} := \left(\sum_{n=1}^{+\infty} (n^\alpha E(f, \Sigma_{n-1}))^q \frac{1}{n} \right)^{1/q}$$

Ce qui ressemble fortement à la norme des espaces de Besov :

$$|f|_{B_q^\alpha(L^p(\Omega))} := \left(\int_0^1 (t^{-\alpha} \omega_r(f, t)_p)^q \frac{dt}{t} \right)^{1/q}$$

Théorème

Soit Ω un domaine de Lipschitz borné, et $\mathfrak{L}$ une fonction de croissance en profondeur, avec $L := sup_{n \geq 1}(\mathfrak{L}(n))$.

- Si $d = 1$ et $L \geqslant 2$, alors $\forall r \in \mathbb{N}$, $\forall p, q > 0$,

  $$B_q^{\alpha}(L^p(\Omega)) \hookrightarrow W_q^{\alpha}(L^p(\Omega), \rho_r, \mathfrak{L})$$

  pour $0 < \alpha < r + \min(1, 1/p)$.
- Si $d > 1$ et $L \geqslant 3$, alors $\forall r \in \mathbb{N}$, $\forall p, q > 0$,

  $$B_q^{d\alpha}(L^p(\Omega)) \hookrightarrow W_q^{\alpha}(L^p(\Omega), \rho_r, \mathfrak{L})$$

  pour $0 < \alpha < \dfrac{\min(1, 1/p)}{d}$

Merci d'avoir (peut-être) écouté !