

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'ap-
proximation

Principe de
l'estimation

Mise en
pratique

Propriétés d'approximation et d'estimation des réseaux de neurones à une couche cachée

Arthur Maritch-Roy, Martin Molter

Table des matières

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'ap-
proximation

Principe de
l'estimation

Mise en
pratique

- 1 Présentation du modèle
- 2 Contrôle de l'erreur d'approximation
- 3 Principe de l'estimation
- 4 Mise en pratique

Objectif général

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'approximation

Principe de
l'estimation

Mise en
pratique

Nos données

Échantillon $((X_i, Y_i)_{1 \leq i \leq N})$ i.i.d. de loi $\mathbb{P}_{(X, Y)}$.

- $X_i \in \mathbb{R}^d, \|X_i\| \leq 1$ de loi $\mathbb{P}_X =: \mu$
- $Y_i \in I, \lambda(I) = b > 0$.

Objectif

Approximer f continue telle que :

$$f(X_i) = Y_i$$

Réseau de neurones à une couche cachée

Propriétés
d'approximation
et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'ap-
proximation

Principe de
l'estimation

Mise en
pratique

Définition

$$f_n(x, \theta) = c_0 + \sum_{k=1}^n c_k \varphi(\langle a_k, x \rangle + b_k)$$

Avec φ sigmoïdale et ν -lipschitzienne, $a_k \in \mathbb{R}^d$, $b_k, c_k, c_0 \in \mathbb{R}$.

Théorème d'approximation universelle

Soit $\varepsilon > 0$. Il existe n et θ tels que

$$\|f_n - f\|_{2,\mu} \leq \varepsilon.$$

Trop théorique, on ne sait pas comment trouver n et θ .

Objectifs de notre présentation

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'approximation

Principe de
l'estimation

Mise en
pratique

On fixe n le nombre de neurones de la couche cachée.

Deux objectifs

- Estimer θ par $\hat{\theta}$ (construit sur l'échantillon)
- Avoir des bornes sur l'erreur

$$\mathbb{E} \left(\|f - f_n(\cdot, \hat{\theta})\|^2 \right)$$

Majoration de l'erreur d'approximation

Propriétés
d'approximation
et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'ap-
proximation

Principe de
l'estimation

Mise en
pratique

Théorème (Barron, 1993)

On suppose que $C_f = \int \|\omega\|_1 |\tilde{f}(\omega)| d\omega < \infty$. Alors il existe θ tel que :

$$\|f - f_n(\cdot, \theta)\|_{2, \mu} \leq \frac{C_f}{\sqrt{n}}.$$

On peut de plus imposer, pour $C \geq C_f$: $\sum_1^n |c_k| \leq C$,
 $|c_0 - f(0)| \leq C$ et $|b_k| \leq \|a_k\|_1$.

Idée de la preuve

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'ap-
proximation

Principe de
l'estimation

Mise en
pratique

Lemme

Soient H un Hilbert, $G \subset H$, avec $\forall g \in G, \|g\| \leq b$. Alors, pour $c' > b^2 - \|f\|^2$ et $\bar{f} \in \overline{\text{Conv}(G)}$, il existe $f_n \in \text{Conv}_n(G)$ telle que :

$$\|\bar{f} - f_n\|^2 \leq \frac{c'}{n} .$$

→ Pour nous : $G = G_\varphi = \{\gamma\varphi(\langle a, \cdot \rangle + b), |\gamma| \leq 2C\}$.

Choix de l'estimateur

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'approximation

Principe de
l'estimation

Mise en
pratique

Indice de résolvabilité

Pour $\Theta = \Theta_n$ discret :

$$R_{n,N} = \min_{\theta \in \Theta_n} \left(\|f - f_n(\cdot, \theta)\|^2 + \lambda \frac{L_{n,N}(\theta)}{N} \right)$$

pour un λ que l'on fixera.

Estimateur des moindres carrés avec pénalisation

$$\hat{f}_{n,N}(x) = \bar{f}_n(x, \hat{\theta}_{n,N}) ,$$

$$\hat{\theta}_{n,N} = \operatorname{argmin}_{\theta \in \Theta_n} \left(\frac{1}{N} \sum_1^N (Y_i - \bar{f}_n(X_i, \theta))^2 + \lambda \frac{L_{n,N}(\theta)}{N} \right)$$

Majoration du risque

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'approximation

Principe de
l'estimation

Mise en
pratique

Théorème (Barron, 1990)

On considère l'estimation d'un réseau de neurone, par une fonction $\hat{f}_{n,N}$ comme précédemment. Alors, si $\lambda > 5b^2/3$:

$$\mathbb{E}\|f - \hat{f}_{n,N}\|^2 \leq \gamma R_{n,N}(f) + \frac{2\gamma\lambda}{N},$$

avec $\gamma = (3\lambda + b^2)/(3\lambda - 5b^2)$. On retiendra :

$$\mathbb{E}\|f - \hat{f}_{n,N}\|^2 \leq O(R_{n,N}(f)).$$

Preuve du théorème

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'approximation

Principe de
l'estimation

Mise en
pratique

Une inégalité (Craig, 1933)

Soient U_i des v.a. indépendantes vérifiant pour un $h > 0$:

$$\mathbb{E}|U_i - \mathbb{E}(U_i)|^k \leq \frac{\text{var}(U_i)}{2} k! h^{k-2} \quad \forall k \geq 2. \text{ Alors :}$$

$$\mathbb{P} \left(\bar{U} - \mathbb{E}\bar{U} \geq \frac{\tau}{N\varepsilon} + \frac{N\varepsilon \text{var}(\bar{U})}{2(1-c)} \right) \leq e^{-\tau}$$

Pour $0 < \varepsilon h \leq c < 1$ et $\tau > 0$.

Discrétisation de l'ensemble des paramètres

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'approximation

Principe de
l'estimation

Mise en
pratique

→ Objectif : rendre Θ fini (a priori, $\Theta = \mathbb{R}^{n(d+2)+1}$)

Rendre Θ borné

$\Theta_{n,\tau,C} = \{\theta \in \Theta, \sum_1^n |c_k|, |c_0 - f(0)| \leq C, |b_k|, \|a_k\| \leq \tau\}$
Alors il existe $f_n = f_n(x, \theta)$ avec $\theta \in \Theta_{n,\tau_n,C}$ (avec τ_n bien choisi) tel que :

$$\|f - f_n\| \leq \frac{2C_f}{\sqrt{n}} .$$

Discrétisation de l'ensemble des paramètres

Propriétés
d'approximation
et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'approximation

Principe de
l'estimation

Mise en
pratique

Rendre Θ discret

Soit $\Theta_{n,\varepsilon,\tau,C}$ le sous ensemble (discret) de $\Theta_{n,\tau,C}$ qui le recouvre à ε -près. Alors pour tout $\theta \in \Theta_{n,\tau,C}$, il existe $\theta^* \in \Theta_{n,\varepsilon,\tau,C}$ tel que :

$$\|f_n(\cdot, \theta) - f_n(\cdot, \theta^*)\| \leq 4Cv\varepsilon .$$

Pour ε_n bien choisi, il existe $\theta^* \in \Theta_{n,\varepsilon_n,\tau,C}$ tel que :

$$\|f - f_n(\cdot, \theta^*)\| \leq O\left(\frac{C}{\sqrt{n}}\right) .$$

→ On conserve le résultat de Barron 93.

Obtention de la borne

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'approximation

Principe de
l'estimation

Mise en
pratique

Choix de la pénalisation

On peut prendre une pénalisation constante :

$$L_{n,N}(\theta) = \log(|\Theta_{n,\varepsilon,\tau,C}|)$$

Pour avoir

$$R_{n,N} = \|f - \hat{f}_{n,N}\|^2 + \lambda \frac{L_{n,N}}{N}$$

Majoration de l'erreur

$$\mathbb{E} \left(\|f - \hat{f}_{n,N}\|^2 \right) \leq \|f - f_n(\cdot, \theta^*)\|^2 + \frac{\lambda}{N} \log(|\Theta_{n,\varepsilon,\tau,C}|)$$

Obtention de la borne

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'approximation

Principe de
l'estimation

Mise en
pratique

Majoration du cardinal

$$\log(|\Theta_{n,\varepsilon,\tau,c}|) \leq (n(d+2) + 1) \log\left(\frac{2e(1+\tau_n)}{\varepsilon}\right)$$

Donc

$$\mathbb{E}\left(\|f - \hat{f}_{n,N}\|^2\right) \leq \|f - f_n(\cdot, \theta^*)\|^2 + \frac{\lambda}{N} m_n \log\left(\frac{2e(1+\tau_n)}{\varepsilon}\right)$$

Optimisation en ε .

Le résultat final

Propriétés
d'approximation et
d'estimation
des réseaux de
neurones à
une couche
cachée

Arthur
Maritch-Roy,
Martin Molter

Présentation
du modèle

Contrôle de
l'erreur d'approximation

Principe de
l'estimation

Mise en
pratique

Théorème (Barron)

L'estimateur des moindres carrés avec pénalisation $\hat{f}_{n,N,C}$, construit sur l'espace de paramètres $\Theta_{n,\varepsilon,\tau,C}$, satisfait :

$$\mathbb{E} \left(\|f - \hat{f}_{n,N}\|^2 \right) \leq O \left(\frac{C^2}{n} \right) + O \left(\frac{nd}{N} \log(CndN) \right)$$