Université Jean Monnet, ED SIS, Institut Camille Jordan

Proposition de sujet de thèse en statistique : **Optimalité minimax des réseaux neuronaux profonds sous conditions de dépendance**

Par William Kengne

Cette proposition de sujet s'inscrit dans le cadre des contrats doctoraux 2025 à l'Université Jean Monnet à Saint-Etienne.

Résumé: Ce projet de thèse vise à établir la propriété d'optimalité minimax des réseaux neuronaux profonds sur des observations dépendantes. Un cadre général, incluant régression et classification, sera considéré pour l'apprentissage profond des processus mélangeants et ψ-faiblement dépendant. Cette thèse développera des méthodes/algorithmes robustes ainsi que des fonctions de pénalités pour la régularisation explicite des réseaux neuronaux profonds et étudiera leurs propriétés minimax.

Mots clés : Réseaux de neurones profonds, optimalité minimax, estimation non paramétrique, estimation robuste, excès de risque, vitesse de convergence, algorithme robuste, régularisation explicite.

1 Cadre de l'apprentissage profond

On se place dans le cadre de l'apprentissage supervisé où on souhaite prédire une variable aléatoire Y à partir de l'entrée X, où Y et X sont respectivement à valeur dans $\mathcal{X} \subset \mathbb{R}^{d_x}$ (espace des entrées) et $\mathcal{Y} \subset \mathbb{R}^{d_y}$ (espace des sorties), avec $d_x, d_y \in \mathbb{N}^*$. Soient $D_n = \{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ des observations de (X, Y) et $\ell : \mathbb{R}^{d_y} \times \mathcal{Y} \to [0, \infty)$ une fonction de perte.

Soient $L \in \mathbb{N}$, $\mathbf{p} = (p_0, p_1, \cdots, p_{L+1}) \in \mathbb{N}^{L+2}$ et $\sigma : \mathbb{R} \to \mathbb{R}$ une fonction d'activation. Pour un réseau de neurones profond (RNP) h d'architecture $(L, \mathbf{p})$, on pose $\text{depth}(h) = L$ et $\text{width}(h) = \max_{1 \le \ell \le L} p_\ell$ et pour $L, N > 0$, on considère :

$$\mathcal{H} := \mathcal{H}_{\sigma, d_x, d_y}(L, N) = \big\{ \text{RNP } h, \ \text{depth}(h) \le L, \ \text{width}(h) \le N \big\}.$$

Pour tout $h \in \mathcal{H}_{\sigma, d_x, d_y}(L, N)$, on considère le risque et le risque empirique de h,

$$R(h) = \mathbb{E}\big[\ell\big(h(X), Y\big)\big] \quad \text{et} \quad \widehat{R}_n(h) = \frac{1}{n} \sum_{i=1}^{n} \ell\big(h(X_i), Y_i\big).$$

Soit un oracle $h^* \in \underset{h\in\mathcal{F}}{\operatorname{argmin}} R(h)$ (où $\mathcal{F} := \mathcal{F}(\mathcal{X}, \mathcal{Y})$ désigne l'ensemble des fonctions de $\mathcal{X}$ vers $\mathcal{Y}$), une cible sur $\mathcal{H}$, $h_{\mathcal{H}} \in \underset{h\in\mathcal{H}}{\operatorname{argmin}} R(h)$, on a la décomposition de l'excès de risque :

$$\underbrace{R(\widehat{h}_n) - R(h^*)}_{\text{excès de risque}} = \underbrace{R(h_{\mathcal{H}}) - R(h^*)}_{\text{erreur d'approximation}} + \underbrace{R(\widehat{h}_n) - R(h_{\mathcal{H}})}_{\text{erreur d'estimation}},$$

pour tout prédicteur $\widehat{h}_n$ construit à partir de l'échantillon d'apprentissage D_n.

Objectif d'apprentissage: Construire à partir des RNPs, un algorithme ayant un excès de risque le plus faible possible. Cette thèse s'intéresse en particulier à la propriété optimalité minimax de l'excès de risque sur une classe de fonctions cible $\mathcal{H}^*$. Cela consiste à établir une borne supérieure $\varphi(n)$ de $\sup_{h^*\in\mathcal{H}^*} \big(R(\widehat{h}_n) - R(h^*)\big)$ (où $\widehat{h}_n$ est l'estimateur par RNP) de telle sorte que, $\inf_{\widehat{h}_n} \sup_{h^*\in\mathcal{H}^*} \big(R(\widehat{h}_n) - R(h^*)\big) \geq \varphi(n)$ à une constante ou un facteur logarithmique près, où le inf est pris pour tout estimateur basé sur l'échantillon D_n.

2 Contexte

Les récents travaux en apprentissage statistique ont permis de grandes avancées sur les propriétés théoriques des prédicteurs/estimateurs par réseaux de neurones profonds (RNPs); voir par exemple [14], [12], [13], [5], [6]. En particulier, le problème de régression non paramétrique par des RNPs a retenu beaucoup l'attention ces cinq dernières années, voir par exemple [1], [9], [7], [10] ainsi que les références qui s'y trouvent.

Mais, les résultats obtenus dans la littérature présentent une ou plusieurs des caractéristiques suivantes : (i) établis pour des observations indépendantes ([14], [12], [5]), (ii) développés à partir des données bornées et/ou sous-gaussiennes ([14], [12]), (iii) établis pour les cas spécifiques de régression ou classification ([11], [15]), (iv) les bornes de l'excès de risque établies ne sont pas optimales ([9], [7]).

L'hypothèse d'indépendance n'est pas vérifiée dans de nombreuses applications de la vie réelle, telles que : le traitement du signal, les observations météorologiques, la prévision du nombre de nouvelles hospitalisations,... L'hypothèse sous-gaussienne ou d'observations bornées exclut les modèles à queues épaisses et pose le problème de la robustesse de tels estimateurs. La propriété d'optimalité minimax des RNPs est très peu étudiée dans le cadre dépendant. Les récents travaux de [1], [11], établissent cette propriété pour l'apprentissage profond des chaînes de Markov et des processus β-mélangeant.

La régularisation des RNPs reste encore très peu comprise et peu étudiée de nos jours ; c'est le cas en particulier de la régularisation implicite. Des avancées ont récemment été réalisées pour le cas de la régularisation explicite, avec des pénalités de type L_1, voir [12], [9], [8].

3 Résultats attendus

Cette thèse considérera des observations dépendantes, i.e. $D_n = \{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ est une trajectoire d'un processus $\{Z_t = (X_t, Y_t), t \in \mathbb{Z}\}$. Un cadre général, incluant régression non paramétrique et classification, est considéré pour l'apprentissage profond des processus mélangeants et ψ-faiblement dépendants, et étudiera pour la classe des prédicteurs/estimateurs par RNPs, des bornes de l'excès de risque et leur optimalité minimax. A l'issue de cette thèse, des algorithmes robustes ainsi que de nouvelles fonctions de pénalité, pour la régularisation explicite des RNPs seront développés. Plus précisément, les problématiques suivantes seront étudiées :

- Optimalité minimax des RNPs sur des observations fortement mélangeants ;
- Optimalité minimax des RNPs robustes pour les données β-mélangeants ;
- Effets de la queue de distribution sur la vitesse de l'excès de risque d'estimateurs par RNP dans un cadre dépendant ;
- Optimalité minimax des RNPs avec des procédures de régularisation de type Tikhonov.

Des exemples d'applications sur des modèles autorégressifs, ainsi qu'aux modèles affines causaux avec ou sans covariables exogènes (voir [3], [2], [4]) seront considérés. Des applications aux données de qualité de l'air, données de pollution, données de nombre d'hospitalisations quotidiennes pour maladies respiratoires, seront aussi considérées.

4 Encadrement et candidature

Cette thèse se déroulera à l'Université Jean Monnet à Saint-Etienne, dans l'équipe Probabilités, Statistique, Physique Mathématique de l'Institut Camille Jordan et sera dirigée par William Kengne.

Le candidat doit avoir une formation de base en mathématiques, et de bonnes connaissances en statistique mathématique et en apprentissage statistique. Pour candidater, envoyer **avant le 10 mars 2025**, un CV, les relevés de notes de L3 (ou équivalent), M1 (ou équivalent) et M2 (ou équivalent) ou notes de M2 déjà disponibles à : **william.kengne@univ-st-etienne.fr**

References

[1] Alquier, P., and Kengne, W. Minimax optimality of deep neural networks on dependent data via pac-bayes bounds. *arXiv preprint arXiv:2410.21702* (2024).

[2] Bardet, J.-M., Kamila, K., and Kengne, W. Consistent model selection criteria and goodness-of-fit test for common time series models. *Electronic Journal of Statistics 14*, 1 (2020), 2009–2052.

[3] Bardet, J.-M., Kare, K., and Kengne, W. Efficient and consistent model selection procedures for time series. *Bernoulli 29*, 4 (2023), 2652–2690.

[4] Diop, M. L., and Kengne, W. Inference and model selection in general causal time series with exogenous covariates. *Electronic Journal of Statistics 16*, 1 (2022), 116–157.

[5] Fan, J., Gu, Y., and Zhou, W.-X. How do noise tails impact on deep relu networks? *The Annals of Statistics 52*, 4 (2024), 1845–1871.

[6] Kengne, W. Excess risk bound for deep learning under weak dependence. *Mathematical Methods in the Applied Sciences* (2025).

[7] Kengne, W., and Wade, M. Deep learning for ψ-weakly dependent processes. *Journal of Statistical Planning and Inference* (2024), 106163.

[8] Kengne, W., and Wade, M. Deep learning from strongly mixing observations: Sparse-penalized regularization and minimax optimality. *arXiv preprint arXiv:2406.08321* (2024).

[9] Kengne, W., and Wade, M. Sparse-penalized deep neural networks estimator under weak dependence. *Metrika* (2024), 1–32.

[10] Kengne, W., and Wade, M. Robust deep learning from weakly dependent data. *Neural Networks, to appear* (2025).

[11] Kurisu, D., Fukami, R., and Koike, Y. Adaptive deep learning for nonparametric time series regression. *arXiv preprint arXiv:2207.02546* (2022).

[12] Ohn, I., and Kim, Y. Nonconvex sparse regularization for deep neural networks and its optimality. *Neural Computation 34*, 2 (2022), 476–517.

[13] Padilla, O. H. M., Tansey, W., and Chen, Y. Quantile regression with relu networks: Estimators and minimax rates. *Journal of Machine Learning Research 23*, 247 (2022), 1–42.

[14] Schmidt-Hieber, J. Nonparametric regression using deep neural networks with relu activation function. *The Annals of Statistics 48*, 4 (2020), 1875–1897.

[15] Zhang, Z., Shi, L., and Zhou, D.-X. Classification with deep neural networks and logistic loss. *Journal of Machine Learning Research 25*, 125 (2024), 1–117.