Sorbonne Université — M2A & M2 Stats **Spring 2024**
Modern Topics in Machine Learning

Reference of the article (authors, year, title, journal, volume, pages)
Approximating Lipschitz functions with GS-NN

by Ugo Tanielian, Maxime Sangnier, Gerard Biau

Name of the student(s) :
OULAD ALI Ayoub

What did you know, or think you knew, about the subject of the paper ? [1] (max 10 lines)

Je savais que les fonctions lipschitz (continue) pouvaient être approchées par des réseaux de neurones (par exemple ReLU).

Je connaissais aussi les enjeux d'avoir des fonctions 1-lipchitz en particulier pour lutter contre le data-poisonning.

Je ne savais pas par contre que ReLU etait mauvais lorsqu'on contraint ses poids en pratique.

Who are the authors ? Briefly describe their background, institution, research fields, name of the journal where they publish,... (max 10 lines)

(LPSM), Directeur du SCAI. Ses travaux reposent sur l'études des modèles de ML et leur consistance, Ran

ıt à Sorbonne Universite (LPSM). Ses travaux touchent aux différentes méthodes d'apprentissage, proces

ıt Criteo AI Lab , ses travaux touchent en particulier aux méthodes génératives

1. Question not graded

Outline this article by indicating :

- **the question(s) the authors wanted to address,**
- **the type of modelling used,**
- **the type of experiments used (if present)**
- **the main results (theoretical or experimental)**
- **the take home message of the paper.**
- **the questions mentioned but not addressed.**

(max 1 page)

Cette recherche propose une architecture de réseaux de neurones conçue pour approximer efficacement les fonctions 1-Lipschitz tout en garantissant mathématiquement cette propriété par construction.
Pour y parvenir, les auteurs remplacent les activations classiques comme ReLU par la fonction GS, qui divise les entrées en groupes de taille k pour les trier par ordre décroissant, tout en imposant des contraintes strictes sur les matrices de poids via les normes mixtes (2,inf) et inf.

Sur le plan théorique, il est démontré que cette structure permet non seulement une représentation exacte des fonctions 1-Lipschitz affines par morceaux, mais agit aussi comme un approximateur universel dont la profondeur nécessaire devient indépendante du taux d'erreur à mesure que k augmente.

Les expérimentations, portant notamment sur la régression 1D et l'estimation de la distance de Wasserstein, révèlent que GroupSort surpasse largement les modèles ReLU en préservant mieux la norme du gradient et le signal discriminatif.

cette architecture offre une flexibilité dans le compromis profondeur-largeur, elle reste confrontée au défi structurel du fléau de la dimension, puisque la taille du réseau requise pour maintenir la précision croît de manière double-exponentielle face à l'augmentation de la dimensionnalité.

What are the main assumptions on the model and on the underlying data ? (max 10 lines)

Tout le papier repose sur une Hypothèse et un lemme :

,,

H1 : Contrainte des poids norm_2,inf(V1) <= 1 et max(norm_inf(Vi) i=2,...,q))<=1

et max(norm(ci)) <= K2.

L1 : Sous cette hypothèse on préserve le gradient et les réseaux sont 1-lip

Find a question, not raised in the article, that could be addressed with the model/data (max 5 lines)

Aucun test fait en dimension d>>1 , une question aurait pu être la question du tradeOFF Temps / Tache / Précision surtout dans un contexte ou la principale utilisation des GAN se fait pour de la génération de données en grande dimension.

Among the references cited in the article, which one would you like to read to enhance your understanding of the field ? Indicate why. (max 15 lines)

He, J., Li, L., Xu, J., and Zheng, C. (2018). ReLU
deep neural networks and linear finite elements.
arXiv:1807.03973. (Le papier le plus cité, theoreme de representation des lpcm par GS-NN)

Björck, A. and Bowie, C. (1971). An iterative algorithm
for computing the best estimate of an orthogonal matrix.
SIAM Journal on Numerical Analysis, 8:358–364.

Comprendre comment la normalisation se fait et pq avoir choisi cette algo.

If you were a reviewer for this paper, what are the comments you would make about it ? Please comment on the content rather than the presentation, and avoid critics on the simplicity of the model (max 15 lines).

En tant que relecteur, on pourrait souligner que le papier offre une solution à un problème d'architecture (la conservation du gradient), mais qu'il ne résout pas le problème fondamental de l'échantillonnage en haute dimension. Il serait nécessaire de demander aux auteurs si des structures de poids spécifiques (ex: parcimonie) pourraient atténuer cette croissance exponentielle, ou si l'avantage du GroupSort s'estompe dès que d >>1.

Il serait intéressant de demander une analyse du temps de calcul par rapport au gain de précision, surtout si l'on doit augmenter la largeur du réseau pour compenser la dimensionnalité.

L'absence de tests sur des distributions à queues lourdes ou des données non structurées limite la portée des conclusions expérimentales actuelles.

[OPTIONAL] Propose a short code [2] **that may illustrate one specific item of the paper (model, assumption, main result, open question, at your convenience)** (max 30 lines including comments + a figure (with legends))

https://colab.research.google.com/drive/1jUKJtdJWQZJ7uOm8M7aZfkKhGloiJcW3?usp=sharing

2. This code should include comments that will allow the reader to understand the different steps