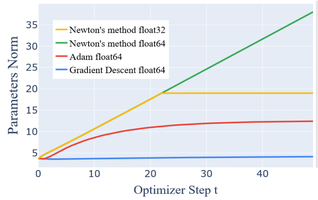
Pay attention to your loss:
understanding misconceptions about 1-Lipschitz neural networks

Louis BETHUNE, Thibaut BOISSIN,
Mathieu SERRURIER, Franck MAMALET,
Corentin FRIEDRICH, Alberto GONZALEZ-SANZ

Why study 1-Lipschitz networks?



Conventional networks optimization leads to uncontrollable growth of their Lipschitz constant.

This causes vulnerabilities to adversarial attacks. 1-Lipschitz networks provide robustness certificates against such attacks. However their expressiveness in classification is often overlooked.

Lipschitz constant $L(f)$:

$$\|f(x) - f(z)\|_2 \leq L(f)\|x - z\|_2$$

Conventional networks can be made 1-Lipschitz:

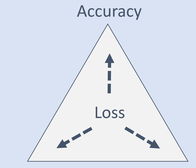
$$g^* = \arg\min_{g \in C(\mathcal{X}, \mathbb{R}^K)} \mathbb{E}_{x,y}\mathcal{L}(g(x), y) \qquad f^* = \frac{1}{L(g^*)} g^*$$

But $L(g^*)$ is often high, and finding it is NP-hard.

1-Lipschitz functions are approximated by constraining the weights of each layers. This is done in practice with **Deel-Lip** library:

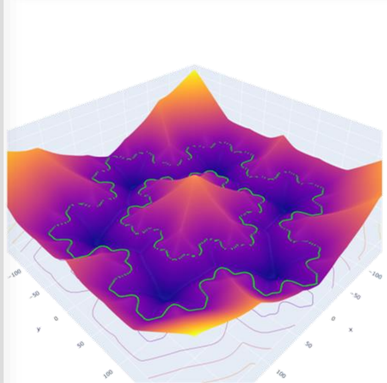
$$f^* = \arg\min_{f \in \mathrm{Lip}_1(\mathcal{X}, \mathbb{R})} \mathbb{E}_{x,y}\mathcal{L}(f(x), y)$$

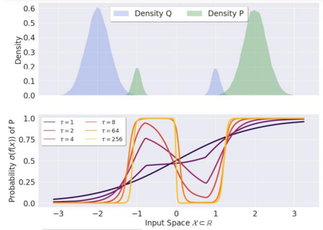
The choice of loss controls the tradeoff between accuracy, robustness and generalization.



In our experiment, 1-Lipschitz Network reached **99.9% accuracy** on Cifar-100 with **random labels.**

Theorem: if classes are separable, zero error is always achievable. However the hyper-parameters of the loss **must be tuned**.





Sigmoid Cross-entropy temperature tuning in 1D classification.

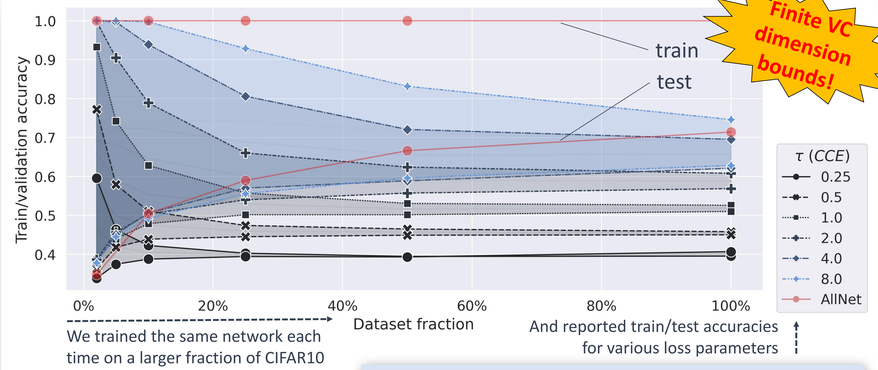
Fractal decision boundary in 2D with Von Koch snowflake.

Accuracy

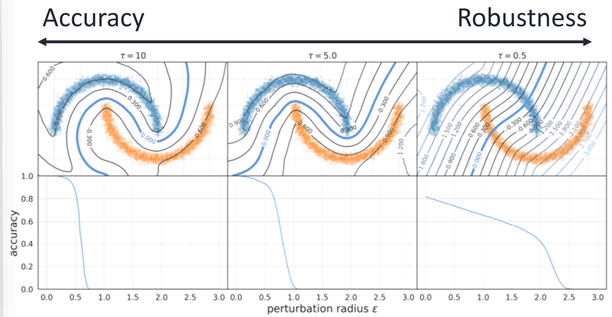
Generalization

Robustness

$$\mathrm{MCR}_{XY}(f) = \mathbb{E}_{x,y}[\mathbb{1}\{yf(x) > 0\}|f(x)|] + \mathbb{E}_{x,y}[-\mathbb{1}\{yf(x) < 0\}|f(x)|] = \mathbb{E}_{x,y}yf(x)$$



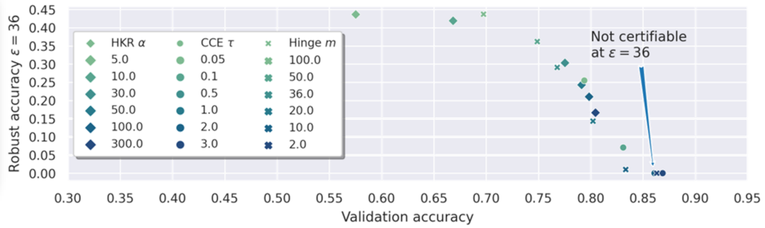
Theorem: 1-Lipschitz networks are ***consistent estimators***; adding more samples closes the train/test gap. Whereas conventional networks can always over-fit regardless of the size of the training set.



Classifiers with highest Mean Certifiable Robustness (MCR):

- Are WGAN discriminators,
- Have usually low accuracy,
- Correspond to low cross-entropy temperature.

Moving along Accuracy-Robustness Pareto front on Cifar-10 by tuning hyper-parameters of the loss.



DEEL DEpendable & Explainable Learning · IVADO · SAINT EXUPÉRY · CRIAQ · ANITI ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE · IRIT

DEEL LIP LIPSCHITZ KERAS LAYERS

NEURAL INFORMATION PROCESSING SYSTEMS November 28 – December 2 - 2022