PhD opportunity at the University of Lille:
Risk bounds for neural networks

PhD advisor: Nicolas Wicker (statistics)

PhD co-advisor: Philippe Heinrich (probability)

Place: Department of mathematics, University of Lille

Subject: The PhD thesis objective is to refine known bounds on capacity measures (Rademacher complexity, covering numbers) in the framework of neural networks. The general idea is to bound the theoretical risk $R(f)$ that a function f, picked in $\mathcal{F}$, using the empiral risk $\hat{R}(f)$ observed on a training sample $S = \{x_1, \ldots, x_m\}$ and a term $\mathcal{R}_m(\mathcal{F})$ measuring the complexity of the class of functions $\mathcal{F}$. This capacity measure is not bounded in a satisfactory way for neural networks.

More generally [Mohri et al., 2012], the iconical inequality in the field states that with probability larger than $1 - \delta$, we have:

$$R(f) \leq \hat{R}(f) + \mathcal{R}_m(\mathcal{F}) + \sqrt{\frac{\log(1/\delta)}{2m}}$$

where $\mathcal{R}_m(\mathcal{F})$ is the Rademacher complexity of the class of functions $\mathcal{F}$. This complexity is defined as the expectation over m-size samples S distributed according to some distribution D^m, of the empirical complexity $\hat{\mathcal{R}}_S(\mathcal{F})$ observed on S:

$$\mathcal{R}_m(\mathcal{F}) = \mathbb{E}_{S \sim D^m}[\hat{\mathcal{R}}_S(\mathcal{F})] \quad \text{avec} \quad \hat{\mathcal{R}}_S(\mathcal{F}) = \mathbb{E}_\sigma \sup_{f \in \mathcal{F}} \left[\frac{1}{m} \sum_{i=1}^{m} \sigma_i f(x_i) \right]$$

where $\sigma = \{\sigma_1, \ldots, \sigma_m\}$ is a set of uniform variables $\sigma_i = \pm 1$ called Rademacher variables.

The estimation of $\hat{\mathcal{R}}_S(\mathcal{F})$ in the literature [Anthony, 2005, Neyshabur et al., 2015, Bartlett et al., 2017] is not really satisfactory as far as neural networks are concerned. Our objective is consequently to review existing results and refine them. A first track is to start from the observation that when decomposing the Rademacher complexity on the last layer of a neural network, that is:

$$\hat{\mathcal{R}}_S(\mathcal{F}) = \mathbb{E}_\sigma \left[\sup_{\substack{\alpha_j \in \mathbb{R} \\ f_j \in \mathcal{F}_j}} \frac{1}{m} \sum_{i=1}^{m} \sigma_i \sum_{j=1}^{k} \alpha_j f_j(x_i) \right]$$

in most of the proves we know, the dependency between the classes of functions $\mathcal{F}_1, \ldots, \mathcal{F}_k$ is not taken into account. If they were, bounds would certainly better. Another approach, using covering numbers and Dudley chaining technique is also considered.

Eligibility:

To be eligible for this opportunity you must:

- possess or in the way to possess by September 2021 a master degree in either mathematics or statistics (with a strong interest in theoretical statistics)
- be fluent in either French or English (the PhD thesis can be written in English)

Benefits:

Stipend of about 1700€ per month.

Contact info:

For further details, contact: nicolas.wicker@univ-lille.fr.

How to apply ?

Send to nicolas.wicker@univ-lille.fr:

- CV
- motivation letter
- transcript of marks and diploma certificates for the last 3 years
- recommendation letter.

References

M. Anthony. Connections between neural networks and boolean functions, 2005.

P.L. Bartlett, D.J. Foster, and M. Telgarsky. Spectrally-normalised margin bounds for neural networks. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017.

M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of machine learning*. MIT Press, 2012.

B. Neyshabur, R. Tomioka, and N. Srebro. Norm-based capacity control in neural networks. In *JMLR: Workshop and conference proceedings*, volume 40, pages 1–26, 2015.