Large-width functional asymptotics for deep Gaussian neural networks

Daniele Bracale[1], Stefano Favaro[1,2], Sandra Fortini[3], Stefano Peluchetti[4]

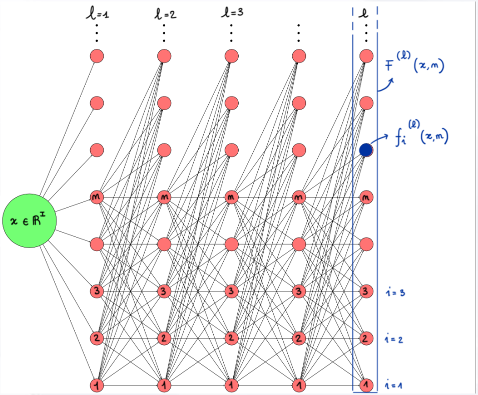
[1] *University of Torino*, [2] *Collegio Carlo Alberto*, [3] *Bocconi University*, [4] *Cogent Labs*

March 20, 2021

Introduction

	Input	Black box	Output
Classic problem	$\{x_1, \dots, x_K\} \subset \mathbb{R}^I$	Neural Network (with random weights)	$K \times I$- dim. random variable
Functional approach	$\mathbb{R}^I$	Neural Network (with random weights)	Stochastic process on $\mathbb{R}^I$

Fully connected feed-forward deep NN



Why an infinite-width NN? Gaussian Process (GP). Applications in NTK, functional-data analysis, Bayesian NN ecc, where the smoothness of the limiting GP provides satisfactory inference performance.

Assumptions and definitions

(I) $L \geq 1$ layers

(II) n goes to infinity jointly on network's layers

(III) **Wheights**: $\omega_{i,j}^{(l)}$ are iid $N(0, \sigma_\omega^2)$

(IV) **Biases**: $b_i^{(l)}$ are iid $N(0, \sigma_b^2)$

(V) **Activation function**: $\phi : \mathbb{R} \rightarrow \mathbb{R}$ Lipschitz non-linearity

(V) **Node i on a layer l**, is a random process

$$x \mapsto f_i^{(l)}(x, n) = \begin{cases} \sum_{j=1}^{\mathcal{I}} \omega_{i,j}^{(1)} x_j + b_i^{(1)}, & l = 1 \\ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \omega_{i,j}^{(l)} \phi(f_j^{(l-1)}(x, n)) + b_i^{(l)}, & l > 1 \end{cases}$$

Remark: $f_i^{(l)}(n) \in C(\mathbb{R}^{\mathcal{I}}; \mathbb{R})$ and are Lipschitz.
Aim: What about the limiting process? Does the NN converges on a function space? Take the limit as n goes to infinity.

Step 1

The weak-limit process $f_i^{(l)}$ exists. Compute the limit of the **fdd**.
Tool: **Levy thm**: $\varphi_{f_i^{(l)}(\mathbf{X},n)}(\mathbf{t}) \to \varphi_{f_i^{(l)}(\mathbf{X})}(\mathbf{t})$ for all $\mathbf{X} = (x^{(1)}, \dots, x^{(k)})$

Step 2

$f_i^{(l)}$ **is a.s. continuous and locally γ-Hölder continuous $\forall \gamma \in (0,1)$.**
Tool: **Kolmogorov-Chentsov thm**: $\mathbb{E}|f_i^{(l)}(x) - f_i^{(l)}(y)|^\alpha \leq H|x-y|^{\mathcal{I}+\beta}$

Step 3

$f_i^{(l)}(n)$ **and** $f_i^{(l)}$ **define probability measures on** $C(\mathbb{R}^{\mathcal{I}}; \mathbb{R})$.
Tool: **Kolmogorov extension thm**: consistency property.

Convergence in function space

$f_i^{(l)}(n) \xrightarrow{d} f_i^{(l)}$ **on** $C(\mathbb{R}^{\mathcal{I}}; \mathbb{R})$.
Proof: Tightness $f_i^{(l)}(n)$ is tight in $C(\mathbb{R}^{\mathcal{I}}; \mathbb{R})+$ Steps 1-3.

Limit for a layer / jointly on the nodes

$$\begin{cases} \mathbf{F}^{(1)}(x) = \left[f_1^{(1)}(x), f_2^{(1)}(x), \dots\right]^T \in \mathbb{R}^\infty \\ \mathbf{F}^{(l)}(x,n) = \left[f_1^{(l)}(x,n), f_2^{(l)}(x,n), \dots\right]^T \in \mathbb{R}^\infty \end{cases}$$

Results for $f_j^{(l)}(\mathbf{X}, n)$ imply results for $F^{(l)}(\mathbf{X}, n)$, i.e. we have

- 1) $\exists$ the weak-limit process $\mathbf{F}^{(l)}$ such that $\mathbf{F}^{(l)}(n) \xrightarrow{f_d} \mathbf{F}^{(l)}$.
- 2) $\mathbf{F}^{(l)}(1), \mathbf{F}^{(l)}(2), \dots$ and the limit process $\mathbf{F}^{(l)}$ are continuous.
- 3) $\mathbf{F}^{(l)}(1), \mathbf{F}^{(l)}(2), \dots$ and the limit process $\mathbf{F}^{(l)}$ define prob. measures on $C(\mathbb{R}^{\mathcal{I}}; \mathbb{R}^\infty)$.
- 4) **functional weak convergence**: $\mathbf{F}^{(l)}(n) \xrightarrow{d} \mathbf{F}^{(l)}$ on $C(\mathbb{R}^{\mathcal{I}}; \mathbb{R}^\infty)$.

Discussion

We proved that a **large-width deep NN with Gaussian wheights** induces a continuous GP on $C(\mathbb{R}^{\mathcal{I}}; S)$ with $S = \mathbb{R}$ or $\mathbb{R}^{\infty}$. More precisely:

- **Classic approach**: the fd sequence of stochastic processess defined on $\mathbb{R}^{\mathcal{I}}$ converges weakly to a Gaussian process on $\mathbb{R}^{\mathcal{I}}$;
- **Smoothness**: the limiting process is locally γ-Hölder continuos for each $\gamma \in (0,1)$;
- **Functional convergence**: the sequence of stochastic processes converges weakly to a GP on $C(\mathbb{R}^{\mathcal{I}}; S)$.

E.g. of application: consider a continuous map $g : C(\mathbb{R}^{\mathcal{I}}; \mathbb{R}^{\infty}) \to \mathbb{R}$. By continuous mapping theorem $g(\mathbf{F}^{(l)}(n)) \xrightarrow{d} g(\mathbf{F}^{(l)})$, and under uniform integrability $\mathbb{E}[g(\mathbf{F}^{(l)}(n))] \to \mathbb{E}[g(\mathbf{F}^{(l)})]$.