M2 internship offer
Neural differential systems for learning with continuous medical data

Keywords. Neural differential equations, continuous-time models, epidemiology, point process theory, survival analysis, theoretical deep learning.

Team. This internship will be supervised by Prof. Agathe Guilloux (DR Inria) and Linus Bleistein (PhD student at Inria Paris). The successful candidate will work in the HeKA team (Inria Paris), a multidisciplinary team regrouping doctors, mathematicians and biostatisticians working on the new challenges arising at the crossroad of machine learning and healthcare.

Duration. 4 to 6 months, starting in spring 2023.

Contact. Agathe Guilloux (agathe.guilloux@inria.fr) and Linus Bleistein (linus.bleistein@inria.fr). Please send a CV and a short paragraph detailing your motivations.

Subject. While numerous methods exists for analysing static medical data, ordered data (often coming in the form of time series) is hard to handle. A variety of approaches have been developed in the recent years, ranging from sophisticated feature extraction methods [2] to deep-learning algorithms [3]. However, these methods often fail when being given irregular and heterogeneous data [11]. This type of data is ubiquitous in the medical context: patients are often observed at different frequencies during different timespans. They also often lack theoretical and statistical guarantees.

A recent line of research has handled this data by embedding it into systems of ordinary differential equations. Heuristically, a $\mathbb{R}^d$-valued time series $x = (x_1, \ldots, x_n)$ is then seen as the discretization of an ODE

$$\dot{x}_t = F(x_t)$$

with unknown vector field $F$. Parametrizing $F$ with a neural network has led to a broad field of research on neural differential equations [1]. One can extend approaches of this type to other categories of differential equations such as controlled or stochastical differential equations [8, 5, 7]. This very general technique has allowed for highly efficient models, that can handle irregular data for various tasks such as online regression or classification. It also allows to leverage the rich theory of differential equations to provide theoretical guarantees for this class of models. As a side note, this continuization approach has also shed new theoretical light on classical optimization algorithms and deep learning architectures [5, 4] - in this case, the successive states (of a RNN) or iterates (of a descent algorithm) are seen as discretizations of continuous processes.

In collaboration with Adeline Fermanian (post-doc at Mines ParisTech) and Anne-Sophie Jannot (MCU-PH at AP-HP and Inria Paris), we have recently worked on a special case of these models for real-time regression in the medical context. By leveraging the rich theory of controlled differential equations, we have provided theoretical guarantees on a linearized learning procedure of the unknown vector field $F$. Our method has been successfully applied to artificial data and epidemiological COVID data.

The goal of this internship is to extend our findings in real-time regression, where the outcome is a time series with values in $\mathbb{R}^p$, to the case where the outcomes are time-to-events [6, 10, 9].
Ideal candidate.

- Strong coding abilities in Python. Working on medical data often requires intensive pre-processing, and deploying the models we have in mind can involve complex and time-consuming coding.
- Solid background in time series, theory of differential equations, analysis and statistics.
- Knowledge of deep learning algorithms.
- Eager to work on medical data and to interact with doctors and healthcare practitioners.

No prior knowledge in medicine is required. A previous research experience is a plus. Candidates not meeting these requirements but showing real interest for the subject are strongly encouraged to apply.

References