Ph.D. research topic

- Title of the proposed topic: **Accounting for uncertainty in machine learning: generative, Bayesian, conformal?**
- Research axis of the 3IA: Axis 1
- **Supervisor (name, affiliation, email):** Pierre-Alexandre Mattei (Inria, Université Côte d’Azur, \url{pierre-alexandre.mattei@inria.fr})
- Potential co-supervisors (name, affiliation): Charles Bouveyron (Inria, Université Côte d’Azur), Aude Sportisse (Inria, Université Côte d’Azur)
- Potential collaborations: Jes Frellsen (Technical U. of Denmark, Copenhagen), Olivier Humbert (Hospital of Nice, Centre Antoine Lacassagne, Université Côte d’Azur)
- The laboratory and/or research group: Inria, Maasai team (Sophia-Antipolis)

Apply by sending an email directly to the supervisor and the co-supervisor.
The application will include:

- Letter of recommendation of the supervisor indicated above
- Curriculum vitæ.
- Motivation Letter.
- Academic transcripts of a master’s degree(s) or equivalent.
- At least, one letter of recommendation.
- Internship report, if possible.

Context and research challenges

The many recent successes of machine learning (ML) have revolutionized several sectors, medical or industrial, by the development of data-driven ML methods. These advances span from supervised learning to protein structure prediction, from unsupervised learning for identifying subgroups of patients with common characteristics for personalized medicine, to semi-supervised learning for diagnostic imaging with a small amount of labeled data. A global strategy for digital health has been recently stated by the World Health Organization [8], recognizing the disruptive role of the digital transformation of healthcare. However, one of the main challenges of ML algorithms is the quantification of uncertainty. Several methods already exist, mostly in predictive contexts: Bayesian models (such as Gaussian processes [7]), ensemble methods (such as deep ensembles [6] or gradient boosting [2]), generative models

(such as energy-based models [4] or variational autoencoders [5]), and conformal prediction [1]. It remains unclear which one of these methods works best, and under which metrics.
Indeed, several metrics for assessing uncertainty exist (coverage, calibration using scoring rules, see e.g. [3]), and it not always obvious how they can benefit practitioners. The goal of this PhD will be to investigate the usefulness of these metrics in different application contexts, and try to understand which methods to quantify uncertainty perform better.
Another important goal would be to study uncertainty beyond predictive contexts, for example for missing data imputation, or in the context of generative models.

Expected skills

The candidate should have a Masters in statistics / machine learning / applied maths, with a strong background in computer science and mathematics. The balance between theory and practice will depend on the profile of the candidate, but a liking for both mathematical proofs and programming (in Python, notably Pytorch and potentially R), is a expected.

References

[1] Angelopoulos and S. Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. arXiv preprint arXiv:2107.07511, 2021.

[2] J. Brophy and D. Lowd. Instance-based uncertainty estimation for gradient-boosted regression trees. Advances in neural information processing systems, 2022.

[3] T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. Journal of the American statistical Association, 2007.

[4] F. Gustafsson, M. Danelljan, R. Timofte, and T. Schön. How to train your energy-based model for regression. In British Machine Vision Conference (BMVC), 2020

[5] S. Kohl et al. A probabilistic U-net for segmentation of ambiguous images. Advances in neural information processing systems, 31, 2018.

[6] B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in neural information processing systems, 2017

[7] C. K. Williams and C. E. Rasmussen. Gaussian processes for machine learning. MIT press Cambridge, MA, 2006

[8] World Health Organisation. Global strategy on digital health 2020-2025, 2020.
https://apps.who.int/iris/bitstream/handle/10665/344249/9789240020924-eng.pdf,
Last accessed on 2022-10-10.