Master thesis position in mathematical statistics and machine learning

Theoretical guarantees for Gaussian and deep Gaussian processes and their approximations

Can be followed by a PhD thesis (already funded)

Project description

Gaussian processes provide very efficient Bayesian models in statistics and machine learning [5, 6], with applications in many fields such as computer experiments, geostatistics, robotics and image analysis. In the last 15 years, their theoretical understanding has increased considerably, see [3] and references therein.

Further theoretical understanding is nevertheless required for two recent extensions of Gaussian processes: deep Gaussian processes and constrained Gaussian processes. Deep Gaussian processes enable to model more irregular functions in practice, and with richer compositional structures [2]. Constrained Gaussian processes enable to take expert knowledge into account, improving prediction and uncertainty quantification drastically [4]. In both cases, open questions remain regarding posterior contraction rates, typically when these extended priors are used in regression, density estimation and classification. In addition, there exist several largely employed approximations for Gaussian processes dedicated to large data sets, in particular variational inference [1], for which the existing theory is not yet complete.

The goal of the master thesis is to study the recent literature on these topics, in order to identify one specific open problem that can be addressed.

The master thesis project is part of the GAP project that is funded by the French national research agency (ANR) in 2021-2025 and hosted at Institut de Mathématiques de Toulouse, with François Bachoc as principal investigator (PI) and Agnès Lagnoux as co-investigator. The master thesis can be followed by a PhD thesis, also funded by the GAP project.

Candidate profile

We are seeking for candidates with a degree in mathematics, with a specialization in probability, statistics, machine learning or applied mathematics. Solid theoretical skills are expected.

Details

- Supervisors: François Bachoc and Agnès Lagnoux (Institut de Mathématiques de Toulouse).
- Start and duration: lasts 4 to 7 months and takes place in Spring and Summer 2022.
- Location: Institut de Mathématiques de Toulouse (Toulouse, France).
- Can be followed by a PhD thesis that would start in Fall 2022.

How to apply

Applications will be considered until the position is filled. The candidates should send a CV, application letter and grade transcripts (master level) to François Bachoc (francois.bachoc@math.univ-toulouse.fr) and Agnès Lagnoux (lagnoux@univ-tlse2.fr).

References

[1] D. R. Burt, C. E. Rasmussen, and M. van der Wilk. Convergence of sparse variational inference in Gaussian processes regression. *Journal of Machine Learning Research*, 21:1–63, 2020.

[2] G. Finocchio and J. Schmidt-Hieber. Posterior contraction for deep Gaussian process priors. *arXiv preprint arXiv:2105.07410*, 2021.

[3] S. Ghosal and A. Van der Vaart. *Fundamentals of nonparametric Bayesian inference*, volume 44. Cambridge University Press, 2017.

[4] A. F. López-Lopera, F. Bachoc, N. Durrande, and O. Roustant. Finite-dimensional Gaussian approximation with linear inequality constraints. *SIAM/ASA Journal on Uncertainty Quantification*, 6(3):1224–1255, 2018.

[5] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, 2006.

[6] M. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.