Offer #2024-07273

PhD Position F/M [DOCT2024-NEO] Generalization Capabilities of Machine Learning Algorithms

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

About the research centre or Inria department

The Inria centre at Université Côte d'Azur includes 37 research teams and 8 support services. The centre's staff (about 500 people) is made up of scientists of different nationalities, engineers, technicians and administrative staff. The teams are mainly located on the university campuses of Sophia Antipolis and Nice as well as Montpellier, in close collaboration with research and higher education laboratories and establishments (Université Côte d'Azur, CNRS, INRAE, INSERM …), but also with the regiona economic players.
With a presence in the fields of computational neuroscience and biology, data science and modeling, software engineering and certification, as well as collaborative robotics, the Inria Centre at Université Côte d'Azur is a major player in terms of scientific excellence through its results and collaborations at both European and international levels.

Context

The PhD candidate is hosted by INRIA at Sophia Antipolis within the research team NEO. The PhD degree is granted by the Université Côte d'Azur (UniCA) and it develops within an international collaboration network under the supervision of Samir M. Perlaza.

Assignment

In [1, 2], we have introduced the notion of worst-case data-generating (WCDG) probability measure, which has been a key instrument to the study of generalization capabilities of machine learning algorithms. In particular, the WCDG probability measure is the key mathematical tool to characterize the generalization error of any machine learning algorithm [3]. This is a major result as such a characterization was known exclusively for one algorithm: the Gibbs algorithm [4]. This characterization of the generalization error has led to the conclusion that algorithms with zero-generalization error are impossible to build, unless the model selection is made independently of the training datasets. From this perspective, some strictly positive generalization error is to be admitted in the design of machine learning algorithms, which is reminiscent to a no-free lunch theorem in optimization theory. On the other hand, algorithms that exhibit the largest generalization error are those that deterministically choose a model based on the training dataset, e.g., empirical risk minimization (ERM) [5]. This analysis has put into the spotlight the Gibbs algorithm as a potential alternative to current machine learning paradigms, e.g., gradient-based algorithms, etc. Nonetheless, the impact of the hyper-parameters of such an algorithm, remains uncharted territory in the realms of statistics, information theory, and statistical machine learning.

References

[1] X. Zou, S. M. Perlaza, I. Esnaola, and E. Altman, "Generalization analysis of machine learning algorithms via the worst-case data-generating probability measure," in Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, Canada, Feb. 2024.

[2] X. Zou, S. M. Perlaza, I. Esnaola, and E. Altman„ "The worst-case data-generating probability measure," INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France, Tech. Rep. RR-9515, Aug. 2023.

[3] X. Zou, S. M. Perlaza, I. Esnaola, E. Altman, and H. V. Poor, "An exact characterization of the generalization error of machine learning algorithms," INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France, Tech. Rep. RR-9539, Jan. 2024.

[4] G. Aminian, Y. Bu, L. Toni, M. Rodrigues, and G. Wornell, "An exact characterization of the generalization error for the Gibbs algorithm," Advances in Neural Information Processing Systems, vol. 34, pp. 8106–8118, Dec. 2021.

[5] S. M. Perlaza, G. Bisson, I. Esnaola, A. Jean-Marie, and S. Rini, "Empirical risk minimization with relative entropy regularization," IEEE Transactions on Information Theory, in press 2024.

[6] F. Daunas, I. Esnaola, S. M. Perlaza, and H. V. Poor, "Empirical risk minimization with f-divergence

regularization in statistical learning," INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France, Tech. Rep. RR-9521, Oct. 2023.

[7] F. Daunas, I. Esnaola, S. M. Perlaza, and H. V. Poor, "Analysis of the relative entropy asymmetry in the regularization of empirical risk minimization," in Proceedings of the IEEE International Symposium on Information Theory (ISIT), Taipei, Taiwan, Jun. 2023.

[8] F. Daunas, I. Esnaola, S. M. Perlaza, and H. V. Poor, "Empirical risk minimization with relative entropy regularization type-II," INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France, Tech. Rep. RR-9508, May. 2023.

[9] S. M. Perlaza, G. Bisson, I. Esnaola, A. Jean-Marie, and S. Rini, "Empirical risk minimization with generalized relative entropy regularization," INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France, Tech. Rep. RR-9454, Feb. 2022.

Main activities

The objectives of this thesis are the following:

- To device generalization guarantees in probability, e.g., probably-approximately- correct (PAC); or in expectation, e.g., generalization error, for which funda- mental limits can be provided. The starting point in this task is the mathematical framework developed upon the WCDG probability measure.

- To characterize the impact of the hyper-parameters on the above generalization guarantees of algorithms obtained from empirical risk minimization subject to f-divergence regularizations [6, 7, 8], e.g., the Gibbs algorithm.

- To provide algorithm-design guidelines based on the developed theoretical framework and implement a testbed to compare with existing alternatives.

Skills

Candidates are expected to have a strong background in mathematics. Previous knowledge on information theory, and game theory is desirable. Abilities in algorithm design and computer programming are also essential. The candidate must have a provable level of written and spoken english. Skills in french language are not required.

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Contribution to mutual insurance (subject to conditions)

Remuneration

Duration: 36 months
Location: Sophia Antipolis, France
Gross Salary per month: 2100€ brut per month (year 1 & 2) and 2190€ brut per month (year 3)

General Information

- **Theme/Domain** : Optimization, machine learning and statistical methods
  Statistics (Big data) (BAP E)
- **Town/city** : Sophia Antipolis
- **Inria Center** : Centre Inria d'Université Côte d'Azur
- **Starting date** : 2024-10-01
- **Duration of contract** : 3 years
- **Deadline to apply** : 2024-04-28

Contacts

- **Inria Team** : NEO
- **PhD Supervisor** :

Medina Perlaza Samir / samir.perlaza@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Before applying, it is strongly recommended that you contact the Scientific manager beforehand.

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.