



Offre n°2024-07273

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

Type de contrat : Fixed-term contract

Niveau de diplôme exigé : Graduate degree or equivalent

Fonction : PhD Position

A propos du centre ou de la direction fonctionnelle

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 regional 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.

Contexte et atouts du poste

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.

Mission confiée

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.

Principales activités

The objectives of this thesis are the following:

- To devise generalization guarantees in probability, e.g., probably-approximately- correct (PAC); or in expectation, e.g., generalization error, for which fundamental 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.

Compétences

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.

Avantages

- 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)

Rémunération

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)

Informations générales

- **Thème/Domaine** : Optimization, machine learning and statistical methods
Statistics (Big data) (BAP E)
- **Ville** : Sophia Antipolis
- **Centre Inria** : [Centre Inria d'Université Côte d'Azur](#)
- **Date de prise de fonction souhaitée** : 2024-10-01
- **Durée de contrat** : 3 years
- **Date limite pour postuler** : 2024-04-28

Contacts

- **Équipe Inria** : [NEO](#)
- **Directeur de thèse** :

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

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

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.