



Institut de Recherche
en Informatique de Toulouse
CNRS - INP - UT3 - UT1 - UT2J

**Institut de Recherche en
Informatique de Toulouse**

Équipe Signaux & Communications
<http://sc.enseiht.fr/>

Contact : Henrique Goulart
henrique.goulart@irit.fr

Master 2 internship proposal – Spring 2022 (5 to 6 months) : Performance of tensor-based machine learning methods for large-scale data

Keywords : tensor models, unsupervised machine learning, random tensors.

Funding This internship will be fully funded by a CIMI (Centre International de Mathématiques et d'Informatique) grant.

Subject description Several machine learning problems can be addressed by leveraging tensor methods, especially in unsupervised settings [1]–[3]. This approach typically relies on estimating a low-rank tensor model from a noisy dataset, which is usually a challenging task. In general, it is difficult to anticipate the best (or the actual) estimation performance that can be attained. Nevertheless, recent years saw substantial progress in this direction, with many authors studying the attainable performance of estimators of such models under the assumption that the dimensions of the observed data tensor are large [4]–[6]. This setting is particularly relevant for large-scale (also known as “big data”) scenarios, where a large number of observations is available.

The primary goal of this internship is to explore the implications of these recent results for some selected practical machine learning problems such as community detection in hypergraphs [7], latent variable model estimation [2] and high-order co-clustering [8]. The intern will thus perform computer simulations aimed at understanding the behavior of estimation algorithms in these target problems, whose performance will be confronted to the existing theoretical predictions. New algorithms and strategies for dealing with these problems may be developed based on the the experimental findings. Scientific dissemination of these developments will be encouraged, via publication of papers and/or participation in scientific events.

Scientific environment The internship will be hosted by the Signal and Communications team (SC) of IRIT, whose members have a strong expertise in signal processing and machine learning methods. He/she will benefit from a rich scientific environment within a large research institution, with opportunities for exchange and interaction with other members. The internship will be supervised by Henrique Goulart, Assistant Professor at Toulouse INP, and co-supervised (in remote collaboration) by Rodrigo Cabral, Assistant Professor at Polytech Nice Sophia Antipolis.

Candidate profile We look for strongly motivated candidates with a solid background on mathematics and statistics, having good programming skills in scientific computing languages (Python, Matlab, Julia).

Practical information

- The intern will be hosted at the ENSEEIHT site of IRIT, located at 2 rue Charles Camichel, Toulouse, in a lively neighborhood close to the city center.
- The monthly internship gratification is of about 600€.

Application procedure Please send an email to Henrique Goulart (henrique.goulart@irit.fr) and Rodrigo Cabral (cabral@i3s.unice.fr) including

- an up-to-date CV;
- official transcripts from each institution that you have attended (in French or English);
- a motivation letter.

You will be contacted if your profile meets the expectations. Review of applications will be closed once the position is filled.

References

- [1] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, *et al.*, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.
- [2] A. Anandkumar, D. Hsu, S. M. Kakade, *et al.*, "Tensor decompositions for learning latent variable models," *Journal of Machine Learning Research*, vol. 15, pp. 2773–2832, 2014.
- [3] S. Rabanser, O. Shchur, and S. Günnemann, "Introduction to tensor decompositions and their applications in machine learning," *arXiv preprint arXiv:1711.10781*, 2017.
- [4] A. Jagannath, P. Lopatto, and L. Miolane, "Statistical thresholds for Tensor PCA," *The Annals of Applied Probability*, vol. 30, no. 4, pp. 1910–1933, 2020.
- [5] A. Perry, A. S. Wein, and A. S. Bandeira, "Statistical limits of spiked tensor models," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 56, no. 1, pp. 230–264, 2020.
- [6] A. Montanari and E. Richard, "A statistical model for tensor PCA," *Advances in Neural Information Processing Systems*, vol. 4, pp. 2897–2905, 2014. eprint: 1411.1076.
- [7] C. Kim, A. S. Bandeira, and M. X. Goemans, "Community detection in hypergraphs, spiked tensor models, and sum-of-squares," in *2017 International Conference on Sampling Theory and Applications (SampTA)*, Tallinn, Estonia, Jul. 2017, pp. 124–128.
- [8] E. E. Papalexakis, N. D. Sidiropoulos, and R. Bro, "From k-means to higher-way co-clustering: Multilinear decomposition with sparse latent factors," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 493–506, 2013.