

**Internship offer -
Advanced discrete optimization meets machine learning**

Supervisors and location:

- *Teams:* LIMOS (Université Clermont Auvergne) / MAP5 (Université de Paris)
- *Supervisor:* Renaud Chicoisne, Maître de conférences, Recherche Opérationnelle, Laboratoire LIMOS, ISIMA, INP-UCA, 1 rue de la Chebarde, 63178 Aubière.
- *Co-Supervisor:* Pierre Latouche, Professeur des Universités, Statistiques, Machine Learning et IA, Laboratoire MAP5, Université de Paris, 45 rue des Saints-Pères, 75006 Paris.
- *Location:* LIMOS, ISIMA, INP-UCA, 1 rue de la Chebarde, 63178 Aubière.

Contact : renaud.chicoisne@uca.fr, pierre.latouche@math.cnrs.fr

Context and project: In this internship, we aim at addressing the problem of Bayesian variable selection for high-dimensional linear and nonlinear regression. We consider a generative model that uses a spike-and-slab-like prior distribution obtained by multiplying a deterministic binary vector, which traduces the sparsity of the problem, with a random Gaussian parameter vector. The originality of this work is to propose to rely on recent efficient discrete optimization techniques to optimize the type-II log-likelihood of the model. In [1], a simple relaxation along with an EM algorithm were employed for inference. Competitive results on simulated and real data were obtained in comparison with state-of-the-art techniques. We plan here to address the optimization task directly with no relaxation in order to improve the selection of the variables even further.

The resulting maximum-likelihood problem can be cast as a high-dimensional, mixed-integer nonlinear optimization program that can be solved via a tailored cutting plane algorithm combined with piecewise-linear approximation techniques. The two supervisors have already written elements of solutions to the problem.

In this internship, the student will develop and also implement (in C++, R, or Python) solutions and algorithms for discrete optimization in Bayesian variable selection. The starting point is the Bayesian linear regression model but extensions to the non-linear case with deep neural networks is also of strong interest.

This internship is particularly recommended for students who are interested in doing a PhD on the subject after the internship. Generally speaking, at the core of this research field is the will to rely on strong optimization techniques to tackle existing and relevant models for machine learning and artificial intelligence.

Expected skills: The candidate should be a master 2 student in a statistics / machine learning / optimization program, with a strong background in mathematics and computer science.

References:

- [1] Latouche, P., Mattei, P. A., Bouveyron, C., & Chiquet, J. (2016). Combining a relaxed EM algorithm with Occam's razor for Bayesian variable selection in high-dimensional regression. *Journal of Multivariate Analysis*, 146, 177-190.
- [2] Bouveyron, C., Latouche, P., & Mattei, P. A. (2018). Bayesian variable selection for globally sparse probabilistic PCA. *Electronic Journal of Statistics*, 12(2), 3036-3070.
- [3] Bouveyron, C., Latouche, P., & Mattei, P. A. (2020). Exact dimensionality selection for Bayesian PCA. *Scandinavian Journal of Statistics*, 47(1), 196-211.