2018-00454 - [Campagne Post-Doctorant 2018/CRI LILLE - Efficient cross-validation for huge data sets (M/F)]

**Level of qualifications required**: PhD or equivalent
**Fonction**: Post-Doctoral Research Visit

---

**About the research centre or Inria department**

The Inria Lille - Nord Europe Research Centre was founded in 2008 and employs a staff of 360, including 300 scientists working in sixteen research teams. Recognised for its outstanding contribution to the socio-economic development of the Nord - Pas-de-Calais Region, the Inria Lille - Nord Europe Research Centre undertakes research in the field of computer science in collaboration with a range of academic, institutional and industrial partners.

The strategy of the Centre is to develop an internationally renowned centre of excellence with a significant impact on the City of Lille and its surrounding area. It works to achieve this by pursuing a range of ambitious research projects in such fields of computer science as the intelligence of data and adaptive software systems. Building on the synergies between research and industry, Inria is a major contributor to skills and technology transfer in the field of computer science.

---

**Context**

**Job environnements**

In real-life applications, two challenging problems arise:

(i) Estimators used by practitioners always depend on unknown parameters that have to be chosen, which is called “parameter tuning”. For instance in statistical learning, Lasso or SVM algorithms crucially depend on a “regularization parameter”, which determines the final performance.

(ii) This tuning step is often performed by use of Cross-Validation (CV) in practice [1].

This is usually time-consuming especially when considering huge datasets, which have become ubiquitous nowadays.

From a practical point of view, a crucial question is to provide more insight in how to enhance CV to drastically reduce the traditional computational burden while achieving a reliable statistical performance. This is a particular instance of the trade-off between statistical performance and computational resources exploitation.

**The goals of the post-doc are:**

- a- deriving a new general strategy to approximate the CV estimators in order to get fast-to-compute quantities,
- b- applying this new strategy to various examples in density estimation (Gaussian mixtures, ...), in model-based clustering, variable selection with Lasso and Ridge procedures, as well as in non-parametric regression where reproducing kernels will be used,
- c- providing a non-asymptotic quantification of what has been lost by the approximation,
- d- identifying guidelines on how to choose the splitting parameter of the CV estimator,
- e- deriving model selection results in terms of oracle-type inequalities.

**Key words**: machine learning

This post-doc position will be held with the MODAL Inria projet-team at the Inria center of Lille.

This team is already experienced in machine/statistical learning.

The post-doc will have access to any usual hardware and software resources.

---

**Assignment**

**Assignments**

The post-doc student will have to publish the results in journal and conferences papers, give talks in France and abroad. The post-doc student will have the opportunity to attend some conferences

---

**General Information**

- **Theme/Domain**: Optimization, machine learning and statistical methods
  Statistics (Big data) (BAP E)
- **Town/city**: Villeneuve d'Ascq
- **Inria Center**: CRI Lille - Nord Europe
- **Starting date**: 11/1/18
- **Duration of contract**: 12 months
- **Deadline to apply**: 3/29/18

---

**Contacts**

- **Inria Team**: MODAL
- **Recruiter**: Celisse Alain / alain.celisse@inria.fr

---

**Conditions for application**

**Instructions to apply:**

Candidates will be treated firstly with a complete file : CV + letter of motivation + list of publications + 2 representative publications + one or more letters of recommendation + prospects for professional integration after the post-doc.

**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.

---

**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.
Main activities

Main activities:
1- write papers
2- implement new algorithms
3- read journal papers and books

Skills

The post-doc student should have a high level of English (spoken and written).

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Sports facilities
- Flexible working hours

More information about Lille:
http://www.lille3000.eu/portail/
http://www.lillemetropole.fr/mel.html

Remuneration

Remunerating

The gross monthly salary is 2653€