Description for the general public

The project is devoted to statistical learning for high dimensional data. High dimensional data refers to the cases when: the sample size n is huge, the number of model parameters p is higher than n or both n and p are large. Nowadays such data are often encountered in biology or genetics. It is very challenging problem and usually classical statistical tools are not adequate to these kinds of problems. In the case of the large sample size, the classical statistical tools usually fail due to computational complexity. When the number of parameters is larger than the sample size, the classical statistical procedures lead to ill-posed problems. Over the last years statistical and machine learning communities have been paying special attention to statistical learning for high dimensional data. Remarkable progress in this field has been achieved. However, there are still areas where progress is required. This project aims at addressing some of them. More precisely, we will focus on the following main groups of problems:

- Analysis of stochastic approximation methods for nonsmooth and nonconvex problems,
- Langevin Monte Carlo algorithms for efficient high dimensional Bayesian inference,
- Structure learning for graphical models.

In the project we plan to construct algorithms that would be able to solve problems from fields mentioned above. Moreover, we will also investigate rigorously the properties of these procedures, for instance their rates of convergence or statistical properties.

The fields of possible applications of obtained results are very diverse, for instance text categorization, multimedia (object recognition, face verification), biology (gene prediction, protein prediction), chemical analysis (drug discovery) or medical diagnosis.