Tight performant guarantees in learning through the localization method (Quentin Paris)

The project aims at studying state of the art methods to derive sharp performance guarantees for empirical risk minimizers in the context of strongly convex loss functions. The project aims at first studying general tools from the theory of empirical processes such as: the symmetrization and contraction principles, the chaining method and concentration inequalities. In a second part, the student will build upon the first part to understand the localization method in which rates of convergence are understood as fixed points of localized Rademacher complexities.

A theoretical study of gradient boosting (Quentin Paris)

In this project, students will study the recent and profound work of Biau and Cadre providing theoretical guarantees on general gradient boosting algorithms.

Bioinformatics: Learning generative probabilistic models for data identification of mass spectrometry data (Atilla Kertesz-Farkas)

The aim in this project is to learn to match a set of experimental data to their correct annotation. The matching will be implemented with Energy based models such as Restricted Boltzmann Machines, Bayesian networks. These models can then yield a probabilistic model which can be used to assess the probability of a correct matching. Algorithms and tests will be implemented and executed on Theano and GPU. Student in this project will acquire hands-on experience with generative models and basic deep learning techniques (Boltzmann machines, etc).

Methods for clustering and community detection in networks (Boris Mirkin)

Trend detection in CS or NLP research papers (Ekaterina Chernyak)

Topics suggested by Vladimir Spokoiny, who is the Head of our Statistical Learning Theory Master’s programme.

A few topics from Sergei Obiedkov:

- Probably Approximately Correct Learning of Description Logic Based Ontologies
- Automated Conjecture-Making for Automated Theorem Provers
- Discovery of Functional Dependencies in Databases
- Probabilistic Computation of Association Rules