Laboratoire Alexander Grothendieck
Équipe de Recherche Labellisée 9216 CNRS-IHÉS

Journée Statistique / Apprentissage à Paris Saclay
Paris Saclay Statistics / Machine Learning Workshop

Organisée par Pierre Alquier (ENSAE) et Alexandre Gramfort (INRIA)

Centre de conférences Marilyn et James Simons

Mercredi 30 janvier 2019

Résumés

Robert Gower (Telecom ParisTech) : *Stochastic Gradient Descent with Arbitrary Sampling: General Analysis and Improved Rates*

I will propose a general yet simple theorem describing the convergence of SGD under the arbitrary sampling paradigm. The analysis relies on the recently introduced notion of expected smoothness and does not rely on a uniform bound of the variance of the stochastic gradients. By specializing our theorem to different mini-batching strategies, such as sampling with replacement and independent sampling, we derive exact expressions for the stepsize as a function of the mini-batch size. With this we can also determine the mini-batch size that optimizes the total complexity, and show explicitly that as the variance of the stochastic gradient evaluated at the minimum grows, so does the optimal mini-batch size. For zero variance, the optimal mini-batch size is one. Moreover, we prove insightful stepsize-switching rules which describe when one should switch from a constant to a decreasing stepsize regime.

Flora Jay (CNRS, LRI) : *Machine Learning and Deep Learning for Population Genetics*

Résumé en attente

Edouard Oyallon (CentraleSupélec / INRIA) **:** *Greedy learning meets large scale*

State-of-the-art neural networks are typically trained with gradients obtained via the back propagation algorithm. However, this can be seen as a black-box method. Indeed, the intermediate objectives of each layers are not explicit. We propose to consider layerwise strategies by decoupling the training objective. First, we show that training each layer sequentially leads to surprisingly good performances on large datasets. In particular, one can train a sequence of 1-hidden layer and obtain competitive performance whereas those neural networks are easier to understand. Second, we observe that the optimization can be done online: this leads to a parallel training of each layer, and we study an asynchronous extension. In particular, this permits to address the issue of update locking, and potentially of forward locking. A large set of analysis is performed on the large scale ImageNet dataset.

Thanh Mai Pham Ngoc (LMO – Université Paris Sud) **:** *Adaptive Estimation of Nonparametric Geometric Graphs*

We study the recovery of graphons when they are convolution kernels on compact (symmetric) metric spaces. This case is of particular interest since it covers the situation where the probability of an edge depends only on some unknown nonparametric function of the distance between latent points, referred to as Nonparametric Geometric Graphs (NGG). In this setting, almost minimax adaptive estimation of NGG is possible using a spectral procedure combined with a Goldenshluger-Lepski adaptation method. The latent spaces covered by our framework encompasses (among others) compact symmetric spaces of rank one, namely real spheres and projective spaces. For these latter, explicit computations of the eigenbasis and of the model complexity can be achieved, leading to quantitative non-asymptotic results. The time complexity of our method scales cubicly in the size of the graph and exponentially in the regularity of the graphon. Hence, our procedure is algorithmically and theoretically efficient to estimate smooth NGG. As a by product, we show a non-asymptotic concentration result on the spectrum of integral operators defined by symmetric kernels (not necessarily positive). This is a joint work with Yohann de Castro and Claire Lacour.

Erwan Scornet (CMAP, Ecole polytechnique) **:** *The Art of Data Science via Mondrian Forests*

The recent and ongoing digital world expansion now allows anyone to have access to a tremendous amount of information. However collecting data is not an end in itself and thus techniques must be designed to gain in-depth knowledge from these large data bases.

This has led to a growing interest for statistics, as a tool to find patterns in complex data structures, and particularly for turnkey algorithms which do not require specific skills from the user.

Such algorithms are quite often designed based on a hunch without any theoretical guarantee. Indeed, the overlay of several simple steps (as in random forests or neural networks) makes the analysis more arduous. Nonetheless, the theory is vital to give assurance on how algorithms operate thus preventing their outputs to be misunderstood.

In this talk, we analyze a stylized version of random forest called Mondrian Forests and prove that it reaches minimax rates of consistency for Lipschitz and twice differentiable regression functions. This is the first result showing the optimality of a particular random forest algorithm in arbitrary dimension. We will also elaborate on the importance of the aggregation in the forest.