Proposition for a Master's thesis

Variants of CMA-ES for Hyperparameter Optimization

Laboratory:	Inria and CMAP - RandOpt team
Supervisors:	Anne Auger, Dimo Brockhoff, Nikolaus Hansen
Contact	anne.auger@inria.fr
Location:	CMAP - Ecole Polytechnique
	Route de Saclay, 91128 Palaiseau cedex

Context and objective of the Master's thesis:

In machine learning, hyperparameter optimization consists in optimizing the hyperparameters of a learning algorithm. For those problems, it is typically impossible to compute any derivative (also, often some parameters are continuous while others are categorical or discrete) and they are just handled with black-box optimization algorithms.

While (surprisingly) hyperparameter optimization is often done via grid search or random search (particularly in the case of deep neural networks), a class of methods also used for hyperparameter tuning is Bayesian optimization methods (that build a model of the objective function to be optimized) [1]. Recently however, it was proposed to use the CMA-ES algorithm [4] for tuning hyperparameters of deep neural networks [2]. In addition, Google communicated lately that they commonly use Bayesian optimization algorithms or CMA-ES for hyperparameters optimization [3].

In the first part of this Master's thesis, we want to understand why random search is still so popular when it comes to hyperparameter optimization in complex machine learning algorithms and on which type of problems it makes sense to use more advanced algorithms and how much we can typically gain.

In a second part we want to design and test novel algorithms that combine Bayesian optimization and CMA-ES or more generally that use the building of a model of the objective function to speed up the optimization process of CMA-ES. We will lean on previous work building quadratic meta-models of the objective function combined with CMA-ES [5].

The questions investigated in this Master's thesis are fully open and in case of success will lead to a scientific publication. Additionally, this Master's thesis can be followed by a PhD thesis.

Practical aspects:

The internship will take place within the applied math laboratory of Ecole Polytechnique (CMAP) within the Inria RandOpt team. The intern will receive a "gratification" (ca. 650 euros per month) and will benefit of all the facilities of the Ecole Polytechnique campus.

References:

[1] E. Brochu, V. M. Cora, and N. De Freitas. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv e-print*, arXiv:1012.2599, 2010.

[2] Ilya Loshchilov and Frank Hutter CMA-ES for Hyperparameter Optimization of Deep Neural Networks https://arxiv.org/pdf/1604.07269.pdf.

[3] Daniel Golovin, Benjamin Solnik, Subhodeep Moitra, Greg Kochanski, John Elliot Karro, D. Sculley Google Vizier: A Service for Black-Box Optimization http://www.kdd.org/kdd2017/papers/view/google-vizier-a-service-for-black-box-optimization

[4] N. Hansen and A. Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[5] Zyed Bouzarkouna, Anne Auger, Didier Yu Ding. Investigating the Local-Meta-Model CMA-ES for Large Population Sizes.