ID de Contribution: **2** Type: **Non spécifié**

Analysis of Gradient Descent on Wide Two-Layer Neural Networks

vendredi 5 février 2021 11:20 (40 minutes)

Artificial neural networks are a class of “prediction” functions parameterized by a large number of parameters – called weights – that are used in various machine learning tasks (classification, regression, etc). Given a learning task, the weights are adjusted via a gradient-based algorithm so that the corresponding predictor achieves a good performance on a given training set. In this talk, we propose an analysis of gradient descent on wide two-layer ReLU neural networks for supervised machine learning tasks, that leads to sharp characterizations of the learned predictor. The main idea is to study the dynamics when the width of the hidden layer goes to infinity, which is a Wasserstein gradient flow. While this dynamics evolves on a non-convex landscape, we show that its limit is a global minimizer if initialized properly. We also study the “implicit bias” of this algorithm when the objective is the unregularized logistic loss: among the many global minimizers, we show that it selects a specific one which is a max-margin classifier in a certain functional space. We finally discuss what these results tell us about the generalization performance and the adaptivity to low dimensional structures of neural networks. This is based on joint work with Francis Bach.

Orateur: CHIZAT, Lenaïc (LMO)