A time scale separation problem for label noise stochastic gradient descent in overparametrised statistical models

Julien Reygner

March 22, 2023

We consider a supervised learning model, such as a neural network, in the so-called overparametrised regime, that is to say with a high-dimensional manifold of parameters which perfectly interpolate the data. In such a case, the details of the optimisation algorithm employed to train the model have a dramatic influence on the estimated value of the parameter. This phenomenon is referred to as the implicit bias of the learning algorithm. Its description is an important step toward the explicability of statistical learning.

In this context, we consider a continuous-time approximation of the stochastic gradient descent with label noise, for a quadratic architecture reminiscent of two-layer neural networks. Numerical experiments show that this dynamics exhibits three phases, which take place at different time scales. A full understanding and characterisation of these three phases is an exciting challenge. In this talk we shall focus on the longest one, which is directly related with the outcome of the training algorithm. Our main result provides a characterisation of the estimated parameter as the solution to a weighted LASSO problem, which thereby makes the bias explicit.

This is a joint work with Loucas Pillaud-Vivien (NYU) and Nicolas Flammarion (EPFL).