

Pierre Marion

Sorbonne Université, France

pierre.marion@sorbonne-universite.fr

We study the training dynamics of shallow neural networks, in a two-timescale regime in which the stepsizes for the inner layer are much smaller than those for the outer layer. In this regime, we prove global convergence of the gradient flow to the optimum of the non-convex optimization problem in a simple univariate setting. The number of neurons need not be asymptotically large for our result to hold, distinguishing our result from popular recent approaches such as the neural tangent kernel or mean-field regimes. Experimental illustration is provided, showing that the stochastic gradient descent behaves according to our description of the gradient flow and thus converges to the global minimum in the two-timescale regime, but can fail outside of this regime.

Joint work with Raphaël Berthier (EPFL, Switzerland).