

WILL ARRIVE SOON

ABSTRACT. Neural networks with a large number of parameters admit a mean-field description, which has recently served as a theoretical explanation for the favorable training properties of “overparameterized” models. In this regime, gradient descent obeys a deterministic partial differential equation (PDE) that converges to a globally optimal solution for networks with a single hidden layer under appropriate assumptions. In this work, we propose a non-local mass transport dynamics that leads to a modified PDE with the same minimizer. We implement this non-local dynamics as a stochastic neuronal birth/death process and we prove that it accelerates the rate of convergence in the mean-field limit. We subsequently realize this PDE with two classes of numerical schemes that converge to the mean-field equation, each of which can easily be implemented for neural networks with finite numbers of parameters. We illustrate our algorithms with two models to provide intuition for the mechanism through which convergence is accelerated.

Joint work with G. Rotskoff, S. Jelassi and E. Vanden-Eijnden