

Convergence of Stochastic Gradient Descent for analytic target functions

Sebastian Kassing, MSc,

Westfälische Wilhelms-Universität Münster, Institut für mathematische Stochastik, Germany

In this talk we discuss almost sure convergence of Stochastic Gradient Descent in discrete and continuous time for a given twice continuously-differentiable target function F . In a first step we give assumptions on the step-sizes and perturbation size to ensure convergence of the target value F and gradient $f=DF$ assuming that f is locally Hölder-continuous. This result entails convergence of the iterates itself in the case where F does not possess a continuum of critical points.

In a general non-convex setting with F possibly containing a rich set of critical points, convergence of the process itself is sometimes taken for granted, but actually is a non-trivial issue as there are solutions to the gradient flow ODE for smooth target functions that stay in a compact set but do not converge. Using the Lojasiewicz-inequality we give sharp bounds on the step-sizes and the size of the perturbation in order to guarantee convergence of the SGD scheme for analytic target functions. Also, we derive the convergence rate of the function value under the assumptions that F satisfies a particular Lojasiewicz-inequality with exponent in $[1/2, 1)$. Finally, we compare the discrete and continuous time results and discuss optimality of the assumptions. This is joint work with Steffen Dereich (WWU Münster).

Related publication: <https://arxiv.org/abs/2102.09385>, <https://arxiv.org/abs/2106.03510>