A general system of differential equations to model first order adaptive algorithms

André BELOTTO DA SILVA

Aix✳Marseille
université
Initiative d'excellence

Institut de Mathématiques de Marseille

Mathematics, Signal Processing and Learning

CIRM, Aix-Marseille Université

25 of January, 2021
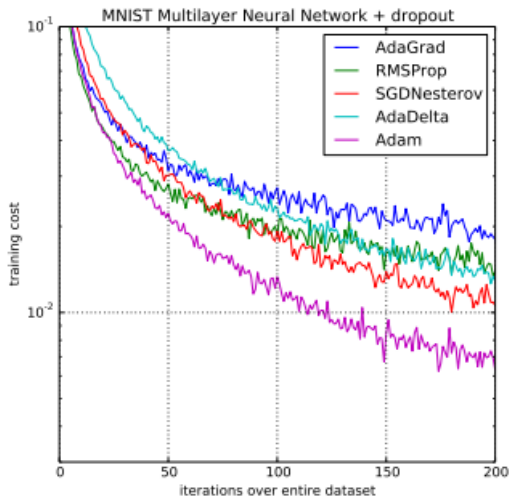
Optimization in deep learning

Figure: Training of multilayer neural networks on MNIST images. (Kingma et Ba, "ADAM: a method for stochastic optimization", 2014).

Project : Dynamical study of Adaptive methods, in collaboration with M. Gazeau, Borealis AI

Our intended goal :
- Help the practitioner to choose the initial conditions, the learning rate and the hyper-parameters
- Provide qualitative advice on when to use ADAM instead of SGD.

First step: Build a differential theory
- Provide a differential equation that describes the system;
- Study the convergence of the system.

Second step: Practical advice
- Find the appropriate intervals where to choose the hyper-parameters;
- Find specific situations where ADAM converges (or does not converge);
- Provide advice to improve the algorithm.

Bibliographic reference :
- A. Belotto da Silva et M. Gazeau, *A general system of differential equations to model first order adaptive algorithms*, Journal of Machine Learning Research, 21(129):1-42, 2020.

Thank you for your attention !