LATTICE 2024

LIVERPOOL

Contribution ID: **104**   Type: **Talk**

Random Matrix Theory for Stochastic Gradient Descent

Wednesday, 31 July 2024 11:15 (20 minutes)

Investigating the dynamics of learning in machine learning algorithms is of paramount importance for understanding how and why an approach may be successful. The tools of physics and statistics provide a robust setting for such investigations. Here we apply concepts from random matrix theory to describe stochastic weight matrix dynamics, using the framework of Dyson Brownian motion. We derive the linear scaling rule between the step size of the optimisation and the batch size, and identify universal and non-universal aspects of the weight matrix dynamics. We test our hypothesis in the (near-)solvable case of the Gaussian Restricted Boltzmann Machine, and explicitly identify the Wigner surmise and semi-circle, and the linear scaling rule.

Primary authors: LUCINI, Biagio (Swansea University); PARK, Chanju (Swansea University); AARTS, Gert (Swansea University)

Presenter: PARK, Chanju (Swansea University)

Session Classification: Algorithms and artificial intelligence

Track Classification: Algorithms and Artificial Intelligence