



Contribution ID: 13

Type: **not specified**

Renormalizing the optimal hyperparameters of a neural network

Tuesday 14 December 2021 18:30 (30 minutes)

Hyperparameter tuning in deep learning is an expensive process, prohibitively so for neural networks (NNs) with billions of parameters that often can only be trained once. We show that, in the recently discovered Maximal Update Parametrization (μP), many optimal hyperparameters remain stable even as model size changes. Using this insight, for example, we are able to re-tune the 6.7-billion-parameter model of GPT-3 and obtain performance comparable to the 13-billionparameter model of GPT-3, effectively doubling the model size. In this context, there is a rich analogy we can make to Wilsonian effective field theory. For example, if “coupling constants” in physics correspond to “optimal hyperparameters” in deep learning and “cutoff scale” corresponds to “model size”, then we can say “ μP ” is a renormalizable theory of neural networks.” We explore this analogy further in the talk and leave open the question whether methods from effective field theory itself can make advances in tuning hyperparameters.

Primary author: YANG, Greg (Microsoft Research)

Presenter: YANG, Greg (Microsoft Research)