

Towards a phenomenological understanding of neural networks

A theory of neural networks (NNs) built upon collective variables would provide scientists with the tools to better understand the learning process at every stage. I argue that a fruitful path for this endeavour of understanding non-linear neural network dynamics is to consider the analogy with physical systems. As an example, I demonstrate that the dynamics of neural networks trained with gradient descent and the dynamics of scalar fields in a flat, vacuum energy dominated Universe are structurally profoundly related. This duality provides the framework for synergies between these systems, to understand and explain neural network dynamics and new ways of simulating and describing early Universe models. In an attempt to capture the dynamics in the non-linear regime effectively, I then introduce two such variables, the entropy and the trace of the empirical neural tangent kernel (NTK) built on the training data passed to the model. We empirically analyze the NN performance in the context of these variables and find that there exists a correlation between the starting entropy, the trace of the NTK, and the generalization of the model computed after training is complete. This framework is then applied to the problem of optimal data selection for the training of NNs.

Presenter: KRIPPENDORF, Sven (LMU)