Towards a Phenomenological Understanding of Neural Networks

Samuel Tovey[1], Sven Krippendorf[2], Konstantin Nikolaou[1], Christian Holm[1]

[1]) University of Stuttgart, Institute for Computational Physics, Allmandring 3, 70569, Stuttgart, Germany
stovey@icp.uni-stuttgart.de
[2]) Universitäts-Sternwarte and Arnold Sommerfeld Center for Theoretical Physics, Faculty of Physics, LMU, Munich, Germany

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. In this work, we introduce two such variables, the entropy and the trace of the empirical neural tangent kernel (NTK) [2] 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 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 with some additional results on optimizing network training. To this end, random network distillation (RND) [1] is used as a means of selecting training data which is then compared with random selection of data. It is shown that not only does RND select data-sets capable of outperforming random selection, but that the collective variables associated with the RND data-sets are larger than those of the randomly selected sets. The results of this investigation provide a stable ground from which the selection of data for NN training can be driven by this phenomenological framework. In the direction of network training, an adaptive optimizer is introduced which takes into consideration the learning dynamics of the underlying network to improve model training.

References

[1] Burda, Y. and Edwards, H. and Storkey, A. and Klimov, O. 2018. Exploration by Random Network Distillation. *arXiv* 1810.12894.

[2] Jacot, A. and Gabriel, F. and Hongler, C. 2018. Neural Tangent Kernel: Convergence and Generalization in Neural Networks. *arXiv* 1806.07572.