George Philipp  
(Georg P. Schoenherr)  
Predicting the performance of neural networks with neural nonlinearity analysis  
Tuesday, April 9, 2019 – 1:00 p.m. – GHC 8102

When building neural architectures, an large number of design decisions need to be made. How many layers should there be? How wide should layers be? What activation functions and normalization methods should be used? How should layers be connected? Because of the size of the architecture design space, we cannot rely exclusively on black-box search algorithms for finding architectures that perform well, especially when the task or dataset is novel and we cannot infuse the search with strong priors from related tasks. Instead, we want to develop an understanding of which architectures are likely to work well on a given task, and why they would work, without having to first train a large number of them.

In this thesis, we introduce the concept of the `degree of nonlinearity’ of a neural network and a scalar metric we term the `nonlinearity coefficient’ (NLC) for measuring the concept of nonlinearity. Via extensive empirical study, we show that the NLC, computed in the network's randomly initialized state before training, is a powerful predictor of test error after training and that attaining a right-sized NLC is essential for attaining an optimal test error. The NLC exceeds previously proposed performance metrics either in terms of predictive power, well-definedness, reliability / robustness, conceptual / theoretical grounding, computability, simplicity, or all of the above. We argue extensively that nonlinearity is the best concept for capturing model complexity in the field of neural networks.

In addition to the NLC itself, we present / propose a broad range of empirical and theoretical analysis that shows how to use the NLC to improve network performance, that explains the reasons behind the predictive power of the NLC and shows how specific design choices influence an architecture's ultimately failure or success. We draw on concepts from random matrix theory, mean field theory, kernel methods and stochastic processes, among others. Ultimately, we propose to establish neural nonlinearity analysis as a design philosophy that we believe can be of great help to architecture designers.

Thesis Committee:  
Jaime G. Carbonell, Chair  
Ruslan Salakhutdinov  
Barnabas Poczos  
Sergey Ioffe, Google  

Thesis Summary: https://www.dropbox.com/sh/k3dof3gb1wxphao/AAAlxHgziOD-rIOLqzP--veta?dl=0