ADAM GARDNER, Artinus Consulting Inc.
Decoding Neural Scaling Laws

For a large variety of models and datasets, neural network performance has been empirically observed to scale as a power-law with model size and dataset size. We will explore the origins of these scaling laws and their relationship to geometric proprieties such as the dimension of the data manifold and symmetries shared by the model and dataset. While the takeaway from these scaling laws for many prominent artificial intelligence labs is to improve performance by increasing model and dataset sizes, we propose an alternative perspective - a deeper mathematical understanding of these scaling laws will help researchers discover more efficient neural network architectures. We conclude with some potential future directions for this line of research.