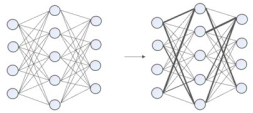
Heavy-tailed update distributions arise from information-driven self-organization in nonequilibrium learning

Xin-Ya Zhang[1,2], Chao Tang[1]

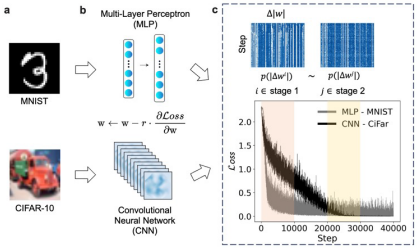
1. Center for Interdisciplinary Studies and Department of Physics, School of Science, Westlake University, Hangzhou 310030, People's Republic of China
2. Institute of Natural Sciences, Westlake Institute for Advanced Study, Hangzhou 310024, People's Republic of China

INTRODUCTION

What is happening during AI learning? Although the learning process is often treated as an opaque box, a central question is whether it exhibits shared patterns beyond specific models and datasets.



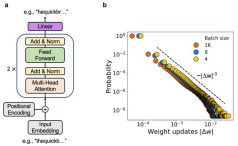
METHODS

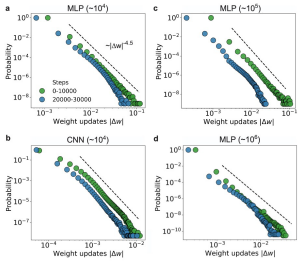


Record the full parameter update in the training process. Neural networks are typically initialized randomly and optimized using algorithms such as stochastic gradient descent (SGD) and Adam, which guide them toward local minima. To investigate the underlying learning dynamics, we analysed the full parameter updates across diverse neural network architectures and training stages.

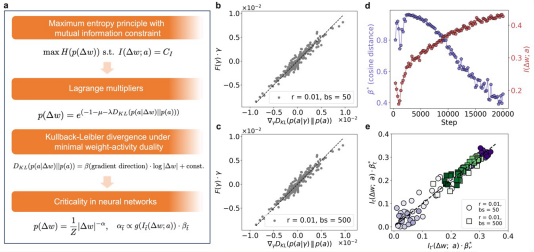
RESULTS

Heavy-tailed update distributions are observed across training stages, architectures, datasets, and a wide range of hyperparameter settings, including optimizers, learning rates, and batch sizes (even full-batch gradient descent).

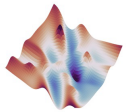




Heavy-tailed update distributions arise from maximum entropy principle under the mutual information constraint.



Furthermore, we find that the loss landscape exhibits multi-scale ruggedness in the absence of mini-batch noise. We also observe a power-law distribution in the intervals between large updates, indicating an intermittent learning process.



CONCLUSIONS

We identify consistent signatures of criticality during neural network training and provide theoretical evidence that such scaling behaviour arises naturally from information-driven self-organization. This points to learning as a nonequilibrium process governed by the fundamental trade-off between randomness and relevance, highlighting its dynamic nature and offering insights into the interpretability of AI systems.

Reference: Xin-Ya Zhang, Chao Tang. Heavy heavy-tailed update distributions arise from information-driven self-organization in nonequilibrium learning. PNAS, 2025.