Visualizing the PHATE of Neural Networks

Scott Gigante[1], Adam S. Charles[2], Smita Krishnaswamy[1] and Gal Mishne[3]

[1]Yale University, USA, [2]Princeton University, USA, [3]University of California, San Diego, USA.

Contact: gmishne@ucsd.edu

scottgigante/M-PHATE

arXiv:1908.02831

NEURAL INFORMATION PROCESSING SYSTEMS

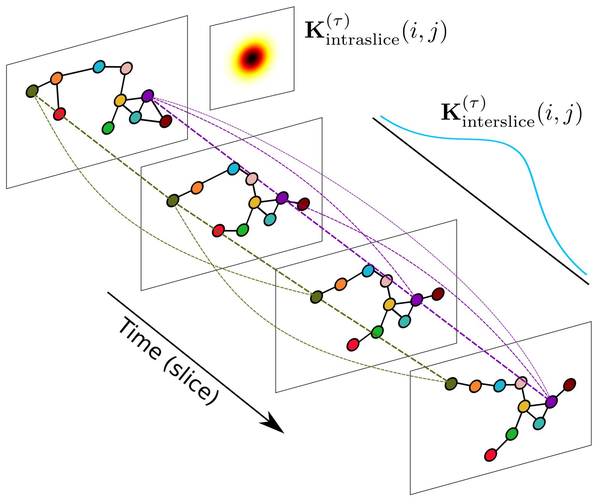
@scottgigante

Introduction

- Understanding why and how certain neural networks outperform others is a challenging and important direction in deep learning
- We introduce Multislice PHATE (M-PHATE), an algorithm designed to visualize how a neural network evolves throughout training
- M-PHATE captures both the dynamics and community structure of the hidden units without the need to access validation data

Multislice graph construction

Multislice graph represents similarities between hidden units over time



$$\mathbf{K}^{(\tau)}_{\text{intraslice}}(i,j) = \exp\left(-\|\mathbf{T}(\tau,i) - \mathbf{T}(\tau,j)\|_2^\alpha / \sigma^\alpha_{(\tau,i)}\right)$$

$$\mathbf{K}^{(i)}_{\text{interslice}}(\tau,\upsilon) = \exp\left(-\|\mathbf{T}(\tau,i) - \mathbf{T}(\upsilon,i)\|_2^2 / \epsilon^2\right)$$

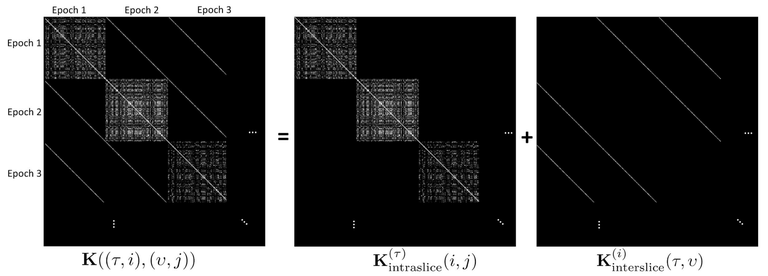
$\mathbf{T}(\tau,i)$ Activations of hidden unit i at epoch τ over a representative sample of training set

$\sigma_{(\tau,i)}$ Distance of $\mathbf{T}(\tau,i)$ to its kth nearest neighbor in epoch τ

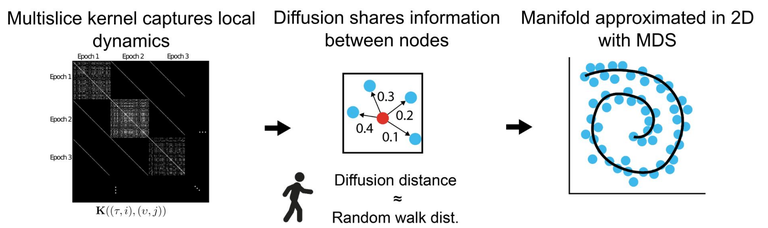
ϵ Mean distance of $\mathbf{T}(\tau,i)$ to its kth nearest neighbor in from hidden unit i

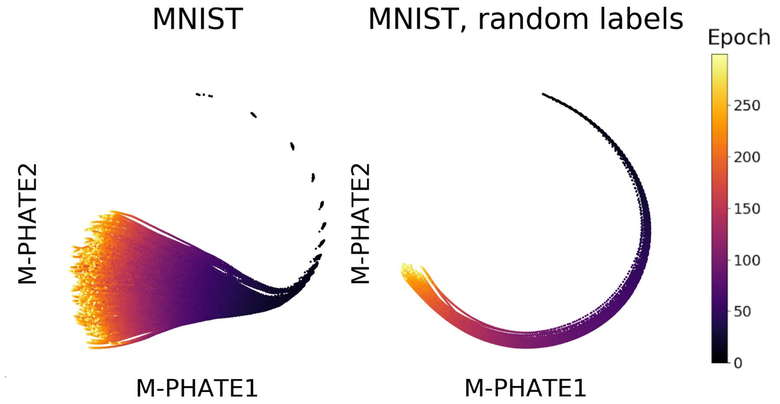
Low-dimensional embedding

Multislice kernel represents each hidden unit at each epoch as a single data point

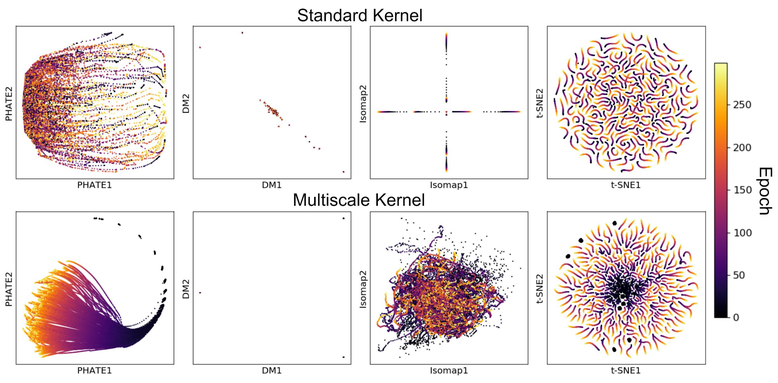


PHATE[3] embeds diffusion distances in low dimensions



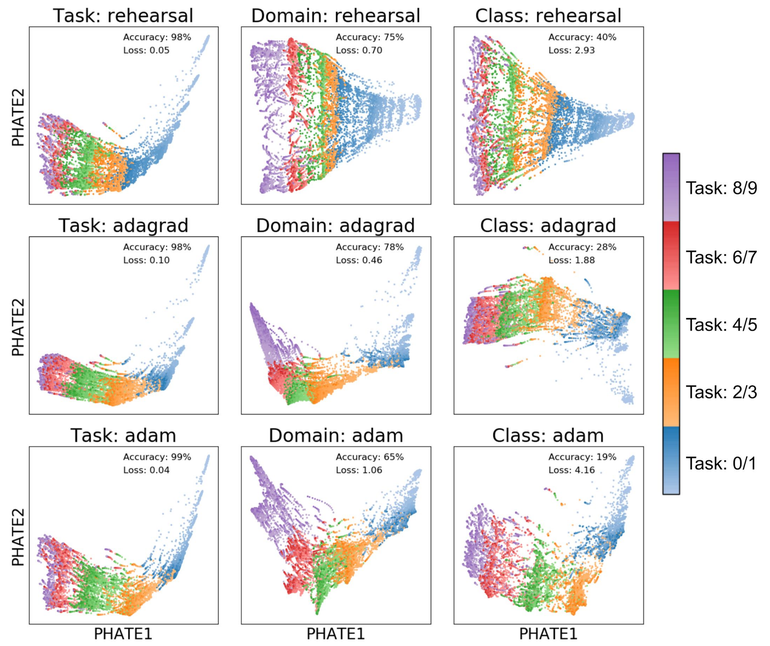


M-PHATE provides unique insight into network's evolution

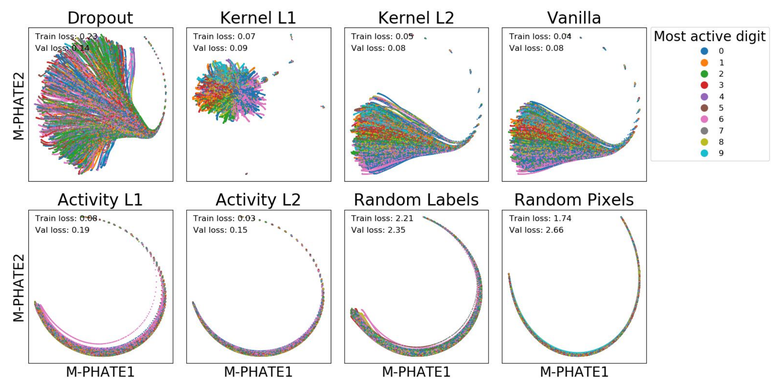


Applications

Continual learning: performance of task-switching networks[2] trained on MNIST is predicted by retention of structure in M-PHATE



Generalization: discrepancy between training and validation loss in classifiers corresponds to complexity of M-PHATE visualization



References

1. Gigante, Charles, Krishnaswamy and Mishne. *Visualizing the PHATE of Neural Networks.* NeurIPS 2019.
2. Hsu, Liu and Kira. *Re-evaluating Continual Learning Scenarios: A Categorization and Case for Strong Baselines.* Continual learning workshop, NeurIPS 2018. arXiv:1810.12488.
3. Moon, van Dijk, Wang, Gigante, et al. *Visualizing structure and transitions in high-dimensional biological data.* Nature Biotechnology, doi:10.1038/s41587-019-0336-3.