Research Proposal for CS4900: UGRC-I
Feature Learning in Neural Networks

Balakrishnan A
CS20B012
Guide: Prof. Harish Guruprasad Ramaswamy

October 2022

Background

Artificial neural networks (ANNs) came into existence from the inspiration of simulating learning like an animal brain. They are widely used for problems that have lots of data available and typically outperform kernel machines such as SVMs in tasks of computer vision and natural language processing. Yet, so much of the theory behind how ANNs train remains largely unclear, mostly due to the immense number of parameters an ANN learns. From the works of Jacot et al., Atanasov et al., Domingos and several others, approximations and equivalences have been developed between ANNs (with imposed limitations on hyperparameters such as infinite width) and kernel machines [1, 3], allowing ANNs to be approximately interpreted as kernel machines that learn a data-dependent kernel, rather than choosing a fixed kernel beforehand the way typical kernel machines do. The neural tangent kernel (NTK) [5] was introduced by Jacot et al. and is used extensively to try and interpret ANNs as approximate kernel machines.

It is debated as to whether a neural network learns symmetry present in the dataset. The NTK approximation suggesting that the ANN relies only on the gradient of a well-defined function like regular kernel machines [3] and the fact that ANNs can learn randomly-labelled datasets [6] suggest that there is no pattern being picked up, whereas divergence from NTK representation [7] and usability of hierarchical learning ability in transfer learning [2, 4] support the argument that ANNs do learn intermediate features or symmetry in the data.

Goals for the course

There is a clear distinction in beliefs on feature learning of ANNs. We will take up an unbiased position by taking neither belief for granted, and try to verify both sides of the argument.

1. We shall identify and analyse situations where ANNs beat kernel machines with a visible/significant improvement, by attempting to build simple low-dimensional datasets and understand why (if at all) neural networks are better. This will help us verify whether or not ANNs can be modelled as kernel machines.

2. The other argument is less formal, due to the difficulty in verifying whether the ANN has actually utilised symmetry present in the data. We shall, however, investigate the ability of ANNs to pick up symmetry in the data by building datasets where symmetry can be clearly exploited and analysing the performance of ANNs over kernel machines.

Tentative timeline

- Study of relevant theory/literature - Winter break
- Goal 1: Building & analysing situations where ANNs beat kernel machines - Jan & Feb 2023
- Goal 2: Analysing ability of ANNs to exploit symmetry in data - March & April 2023
- Report & Presentation preparation - A continuous process, with special focus in the last 2 weeks

Approval of guide

References

[1] Alexander Atanasov, Blake Bordelon, and Cengiz Pehlevan. Neural networks as kernel learners: The silent alignment effect. *arXiv preprint arXiv:2111.00034*, 2021.

[2] Minshuo Chen, Yu Bai, Jason D Lee, Tuo Zhao, Huan Wang, Caiming Xiong, and Richard Socher. Towards understanding hierarchical learning: Benefits of neural representations. *Advances in Neural Information Processing Systems*, 33:22134–22145, 2020.

[3] Pedro Domingos. Every model learned by gradient descent is approximately a kernel machine. *arXiv preprint arXiv:2012.00152*, 2020.

[4] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[5] Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. *Advances in neural information processing systems*, 31, 2018.

[6] Hartmut Maennel, Ibrahim M Alabdulmohsin, Ilya O Tolstikhin, Robert Baldock, Olivier Bousquet, Sylvain Gelly, and Daniel Keysers. What do neural networks learn when trained with random labels? *Advances in Neural Information Processing Systems*, 33:19693–19704, 2020.

[7] Mariia Seleznova and Gitta Kutyniok. Analyzing finite neural networks: Can we trust neural tangent kernel theory? In *Mathematical and Scientific Machine Learning*, pages 868–895. PMLR, 2022.