Introduction

Artificial neural networks (ANNs) are a collection of connected computational nodes inspired by biological neural networks. Each connection can transmit a helpful signal to another computational node like synapses in a brain. ANNs demonstrated colossal advancements in the last decades. Thanks to these advancements, it is possible to solve complex problems in computer vision, speech recognition, and natural language processing within a reasonable amount of time and with satisfactory performance. These advancements were actualized through an old but robust algorithm called backpropagation (BP). BP is a training algorithm for ANNs based on repeatedly adjusting network weights to minimize the difference (loss) between the output of the network and the ground truth [1].

Although nowadays BP is the workhorse algorithm for training ANNs, it has some drawbacks, and it is not the only alternative. Recent studies offered different algorithms to train ANNs by addressing these drawbacks. These algorithms have other properties and principles than BP. Some of them are competitive with BP, or they even outperform the BP in terms of performance or convergence speed for specific problems.

This thesis investigates the learning structures through BP and one of the alternative algorithm called direct feedback alignment (DFA) on the particular problem. Unlike BP, the error is propagated through a fixed random matrix instead of the layers' weights in DFA. Then network learns how to make this feedback useful [2]. Due to this error propagation mechanism, DFA is considered more biologically plausible than BP, and it opens the gate of parallelism in the training phase of ANNs.

The problem at hand is known as the parity learning problem. Previous results showed that these parities are learnable by BP and lazy methods in a more simple setting, whereas it is only learnable by BP in a more complex setting [3]. That is why it is intriguing to test alternative algorithms on this problem to understand their learning dynamics and capabilities.

The experiment results might lead us to three possible outcomes. First, we might acquire a similar performance as BP. If it is the case, it would be beneficial to test DFA and BP on a more challenging problem for further studies. Second, there might be a gap between BP and DFA then it would be intriguing to understand where the difference is coming from and how we can close this gap. Third, the alternative algorithm might not even learn, and in this case, it is interesting to ask what makes a problem learnable by BP but not DFA. In all cases, results should help to understand the dynamics of learning of both methods.

For applying BP and DFA in a more realistic setting, experiments are performed on the MNIST dataset by imitating the parity learning problem. After putting DFA to this frame, the reason behind the results is interpreted, and possible improvements are motivated and implemented.

Chapter **??** constructs the theoretical bases of the algorithms that are used for the experiments. These bases are composed of simple definitions, mathematical foundations, and the drawbacks of the algorithms. They help to dig deeper into the learning structures of the training algorithms. It is expected to have more control over their learning behaviors by tweaking components of these foundations. Also, it is beneficial to have these theoretical bases for acquiring a better understanding of the further interventions. Moreover, these theoretical foundations are used to implement the algorithms from scratch to use in experiments.

Chapter **??** introduces the parity learning problem at hand. First, the formal definition of the problem is demonstrated then how the problem is imitated by using the MNIST dataset is explained in detail. This part is also highly correlated with the training phase of the algorithms. Later experiment results from [3] are replicated to have a concrete picture of previous studies.

Chapter **??** presents results of the experiments. This chapter is the main contribution of the study. The first experiment is testing DFA on the same problem and observing the difference with BP. As it is specified before, depending on the experiment outcome, different further experiments are performed. Such as closing the gap between BP and DFA, if any, trying harder problems if we acquire similar results. Explain why DFA cannot learn the problem if we get a similar behavior as lazy methods. In addition to the main experiment, we performed side experiments that helped us understand the algorithms' learning dynamics. For instance, using different random matrices for DFA, comparing their results, and observing hidden representation of BP and DFA to understand if the networks learn the digits individually to calculate the parities or memorizes the data without knowing the digits.

Chapter **??** wraps up the findings from experiments by summarizing the key findings. It creates a path for future studies that are not covered in this study.

References

[1] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning Representations by Back-propagating Errors". In: *Nature* 323.6088 (1986), pp. 533–536. URL: `http://www.nature.com/articles/323533a0`.

[2] Arild Nøkland. *Direct Feedback Alignment Provides Learning in Deep Neural Networks.* 2016. arXiv: `1609.01596 [stat.ML]`.

[3] Amit Daniely and Eran Malach. "Learning Parities with Neural Networks". In: *CoRR* abs/2002.07400 (2020). arXiv: `2002.07400`. URL: `https://arxiv.org/abs/2002.07400`.