Project Synopsis for 02456 Deep Learning
Investigating Parameter Sharing in Neural Networks (P7)

Authors: Justine Fastner, Stefanie Lanz, Niclas Pokel, Maximilian Riedl (Group 40)
Supervisor: Laurits Fredsgaard Larsen (laula@dtu.dk)

11-11-2024

Motivation

BatchEnsemble has demonstrated itself as an effective method for training multiple neural networks in parallel, consistently achieving high levels of accuracy and predictive performance. However, as the number of networks grows, the associated computational costs increase significantly, making this approach resource-intensive [1]. To address these limitations, we aim to investigate methods for reducing the parameter count within neural networks. By minimizing the number of parameters, we seek to enable more computationally efficient training processes while preserving the high model performance and reliability that BatchEnsemble provides.

Background

Ensemble methods have demonstrated remarkable effectiveness in improving the robustness and accuracy of machine learning models. By combining the predictions of multiple models, ensembles can often outperform individual models. However, this diversity of models leads to a significant increase in the number of parameters, resulting in higher computational costs. BatchEnsemble offers an elegant solution to this problem. Instead of associating each ensemble member with a fully trainable matrix of its own, BatchEnsemble shares a common base matrix across all members. Each ensemble member then computes its prediction by multiplying this base matrix with a specific, low-rank perturbation vector. This creates a family of ensemble weights that are closely related. This approach allows for a direct computation from an input data point to an output value, eliminating the need for sequential calculations for each individual ensemble member. Instead, all predictions can be computed in parallel, leading to significant efficiency gains [1]. In this project, we aim to adapt this strategy to train single neural networks. We will investigate how parameters can be shared within a network to improve both performance and efficiency. In particular, we will explore the possibility of sharing parameters across different layers and groups within a layer.

Milestones

Milestone 1.1 - Explore Convolutional Layers: Implement convolutional and batch normalization layers in a model for the CIFAR-10 dataset.

Milestone 1.2 - Compare with IVON Optimizer: Compare different optimizers for training deep neural networks, particularly focusing on their impact on convergence speed and performance [2].

Optional Milestones

Milestone 2.1 - Parameter Sharing Within Single Networks: Explore techniques for sharing parameters within a single neural network to improve efficiency and performance.

Milestone 2.2 - Investigate Bayesian Neural Networks: Investigate the application of Bayesian neural networks to improve uncertainty quantification and model robustness.

References

[1] Yeming Wen, Dustin Tran, and Jimmy Ba, "Batchensemble: an alternative approach to efficient ensemble and lifelong learning," *arXiv preprint arXiv:2002.06715*, 2020.

[2] Yuesong Shen, Nico Daheim, Bai Cong, Peter Nickl, Gian Maria Marconi, Clement Bazan, Rio Yokota, Iryna Gurevych, Daniel Cremers, Mohammad Emtiyaz Khan, et al., "Variational learning is effective for large deep networks," *arXiv preprint arXiv:2402.17641*, 2024.