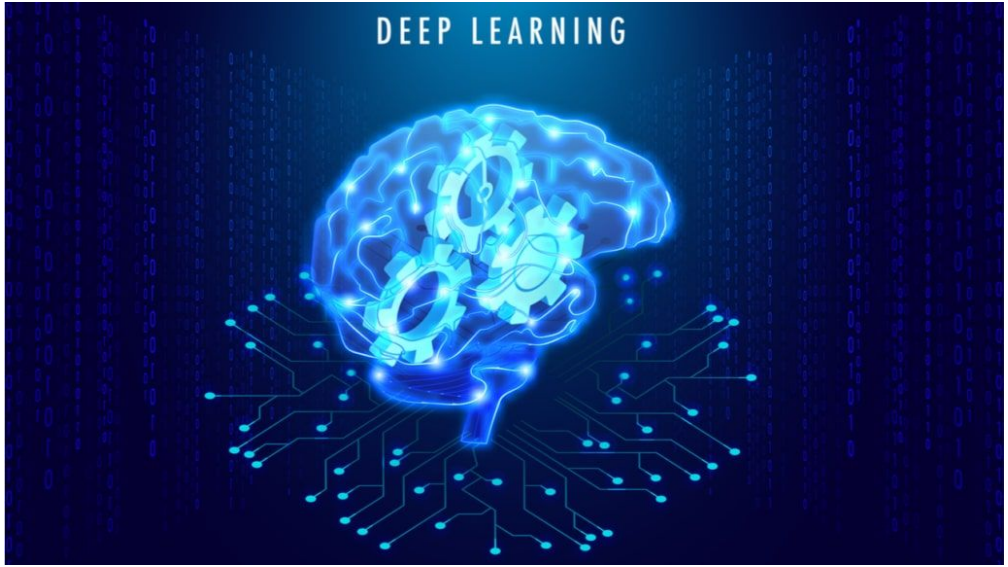
Efficient Training of Deep Learning Models

By Sai Dheeraj Gummadi

Scope of Presentation

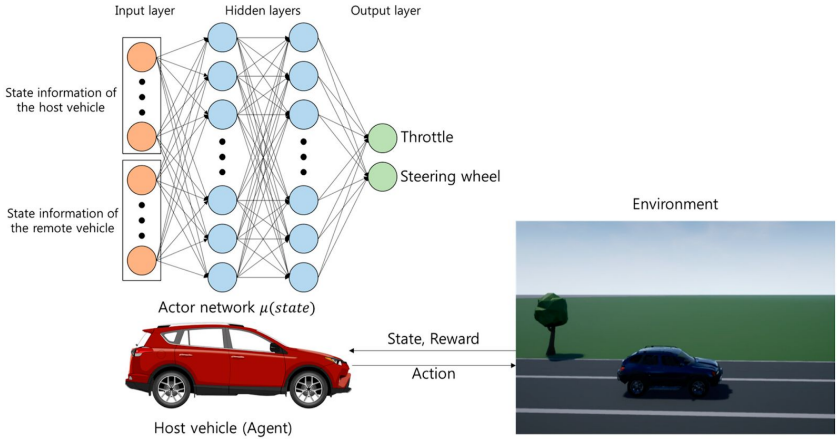
- Introduction to Deep learning
- Motivate the rationale and use of pruning techniques in the context of Deep Learning model training.
- Explain the “Lottery Ticket Hypothesis” and the associated “Lottery Ticket Conjecture”.
- Summarize the identification of so-called “winning lottery tickets” in fully connected and convolutional Deep Learning models.

Introduction to Deep Learning



- Deep learning is a method in artificial intelligence (AI) that teaches computers to process data in a way that is inspired by the human brain.

Use of pruning techniques

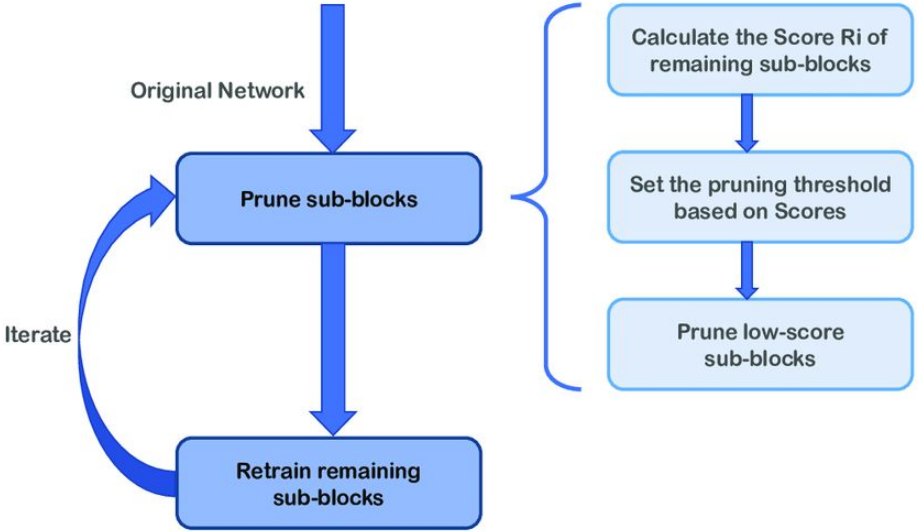


- Deep learning models are often massive, with millions or even billions of parameters.
- This leads to high computational costs and memory demands, especially for deployment on resource-constrained devices.
- Pruning techniques aim to reduce model size without sacrificing accuracy.

The Lottery Ticket Hypothesis

- The Lottery Ticket Hypothesis (LTH) proposes that randomly initialized dense neural networks contain sparse subnetworks (winning tickets) that can achieve comparable or even better performance when trained in isolation.
- These winning tickets are initialized with a special configuration that makes them particularly effective for training.

Finding Winning Tickets



The LTH is often demonstrated empirically through iterative magnitude-based pruning:

- Train the original dense network.
- Prune a percentage of the weights with the smallest magnitudes.
- Reinitialize the remaining weights to their initial values.
- Retrain the pruned network from scratch.

The Lottery Ticket Conjecture

The Lottery Ticket Conjecture extends the LTH, suggesting that every sufficiently over-parameterized dense network contains a subnetwork that can achieve impressive accuracy without any training.

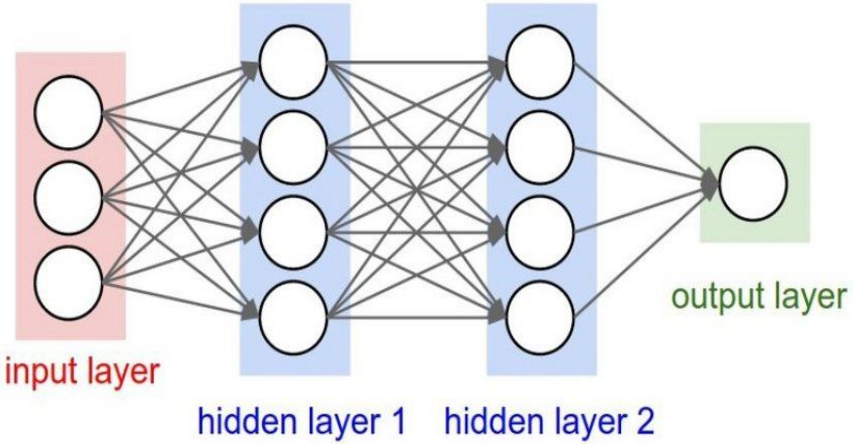
This implies that the winning ticket already exists at initialization, and its initial weights are already good enough for the task.

Implications of the LTH

Training from Scratch: The LTH suggests that we can potentially skip training a massive dense network and instead focus on training a small, pruned subnetwork from scratch, potentially saving a lot of computational resources.

Understanding Initialization: The LTH highlights the importance of initialization in deep learning. Winning tickets show that the right combination of initial weights can significantly affect a network's training efficiency and final performance.

Winning Tickets in Fully Connected Networks



- Winning tickets often display local connectivity patterns, especially in the first hidden layer.
- The pruned weights tend to be less correlated, implying the network benefits from a simpler structure.
- Winning tickets can be identified and trained from scratch, showcasing their potential for efficient model training

Winning Tickets in Convolutional Networks

Winning tickets tend to recover the key features of Convolutional Neural Networks (CNNs), including local connectivity and weight sharing.

Pruning often leads to localized masks, especially in the early layers, resembling the patterns found in CNNs.

The process of discovering these winning tickets is more nuanced, often requiring a careful selection of hyperparameters and a specific approach to pruning.

Open Questions

Universality: Do winning tickets exist for all types of networks and all tasks?

Finding Winners: How can we efficiently identify winning tickets, especially in larger networks?

Real-world Applicability: How can we effectively leverage winning tickets in real-world scenarios for efficient deployment of deep learning models?

Research Directions

Develop robust and efficient methods for finding winning tickets, especially in larger networks.

Explore the universality of winning tickets across different network architectures and tasks.

Investigate the potential of winning tickets for transfer learning and other applications.

Conclusion

The Lottery Ticket Hypothesis and Conjecture offer a new perspective on deep learning, highlighting the importance of initialization and the potential for efficient model training.

While much research is still needed, the LTH has the potential to revolutionize our understanding of how to design and train deep learning models.

References

1. Yeom, Seul-Ki, et al. "Pruning by explaining: A novel criterion for deep neural network pruning." *Pattern Recognition* 115 (2021): 107899.
2. Zhu, Michael, and Suyog Gupta. "To prune, or not to prune: exploring the efficacy of pruning for model compression." *arXiv preprint arXiv:1710.01878* (2017).
3. Frankle, Jonathan, and Michael Carbin. "The lottery ticket hypothesis: Finding sparse, trainable neural networks." *arXiv preprint arXiv:1803.03635* (2018).
4. Liu, Zhuang, et al. "Rethinking the value of network pruning." *arXiv preprint arXiv:1810.05270* (2018).
5. Ramanujan, Vivek, et al. "What's hidden in a randomly weighted neural network?." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
6. Pellegrini, Franco, and Giulio Biroli. "Sifting out the features by pruning: Are convolutional networks the winning lottery ticket of fully connected ones?." *arXiv preprint arXiv:2104.13343* (2021).
7. Raju, Ravi S., Kyle Daruwalla, and Mikko Lipasti. "Accelerating deep learning with dynamic data pruning." *arXiv preprint arXiv:2111.12621* (2021).
8. Liu, Bohan, et al. "A survey of lottery ticket hypothesis." *arXiv preprint arXiv:2403.04861* (2024).
9. da Cunha, Arthur, Emanuele Natale, and Laurent Viennot. "Proving the strong lottery ticket hypothesis for convolutional neural networks." *ICLR 2022-10th International Conference on Learning Representations*. 2022.
10. Anwar, Sajid, Kyuyeon Hwang, and Wonyong Sung. "Structured pruning of deep convolutional neural networks." *ACM Journal on Emerging Technologies in Computing Systems (JETC)* 13.3 (2017): 1-18.
11. Li, Lianqiang, Jie Zhu, and Ming-Ting Sun. "Deep learning based method for pruning deep neural networks." *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2019.

References

1. Van Soelen, Ryan, and John W. Sheppard. "Using winning lottery tickets in transfer learning for convolutional neural networks." *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019.
2. Malach, Eran, et al. "Proving the lottery ticket hypothesis: Pruning is all you need." *International Conference on Machine Learning*. PMLR, 2020.
3. Vadera, Sunil, and Salem Ameen. "Methods for pruning deep neural networks." *IEEE Access* 10 (2022): 63280-63300.
4. Roy, Sourjya, et al. "Pruning filters while training for efficiently optimizing deep learning networks." *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020.
5. da Cunha, Arthur, Emanuele Natale, and Laurent Viennot. "Proving the strong lottery ticket hypothesis for convolutional neural networks." *ICLR 2022-10th International Conference on Learning Representations*. 2022.

Thank You