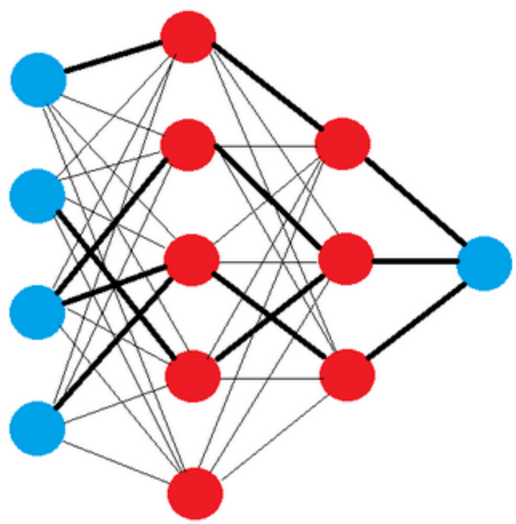
THE LOTTERY TICKET HYPOTHESIS:
FINDING SPARSE, TRAINABLE NEURAL NETWORKS

Group 2: Isha Dighe, Sneha Liza George, Kong Lin, Megan Malisani, Sujeeth Shetty

What is Pruning: Pruning (removing weights) is a valuable tactic for reducing computation time while maintaining the high accuracy of a more complex model. Typically, a complex model is trained, then its least important connections are pruned. Neural networks with a fewer parameters take less time in performance and have a decreased requirement of storage.



Problem with Pruning: Researchers found that is it difficult to achieve accuracy comparable with the complex model when training a neural network from the start using most pruned models. However, there appears to be a small subset of pruned models where training neural networks that achieve comparable accuracy is possible. Such models are called winning tickets.

Goal of Research: The authors present a method for identifying winning tickets. Furthermore, the authors explored why training a network from scratch with the properties of a pruned networks yields a less accurate performance. Overall, the wish to improve training performance, design better networks, and improve their theoretical understanding of neural networks.

Main Hypothesis: Authors hypothesize that gradient decent seeks out a subset of well-initialized weights and mainly trains these connections. If you double the total number of weights in a model, on average, you will also double the number of weights that are well-initialized. As a result, complex models achieve higher accuracy due to higher likelihood of having enough well-initialized weights to achieve a good result.

Method for identifying winning ticket models: The paper presents two pruning strategies to find winning tickets. The winning ticket represents the weight initializations of the pruned subnetwork that yields the highest accuracy. These steps represent 1 iteration of pruning. This process is generally performed iteratively until a satisfactory result is achieved. The authors found that iterative pruning successfully results in smaller pruned subnetworks. While this method is useful for computation time, it still requires a complex network to be trained first.

1. Randomly initialize a neural network
2. train a complex network
3. prune weights with the smallest magnitudes
4. Reset all remaining connections to the initialization values that were assigned prior to training the model

One-shot pruning: This strategy makes it possible to identify winning tickets without the cost of repeated training that accompanies iterative pruning. However, iterative pruned winning tickets learned faster and reached a higher accuracy. Therefore, the authors chose to focus on **Iterative Pruning** strategy as their goal is to identify the smallest possible winning tickets.

Winning tickets in Fully Connected Networks: The lottery ticket hypothesis was applied to Lenet architecture on MNIST dataset. Here authors used simple layer-wise pruning heuristic where they removed a percentage of weights with the lowest magnitudes within each layer. The found that in iterative pruning the winning tickets learnt faster than the original network. Since the goal was to identify the smallest possible winning tickets, Iterative Pruning was deemed best, even though it was computationally expensive compared to one-shot pruning.

Winning tickets in Convolutional Networks: The authors applied the lottery ticket hypothesis on CIFAR10, a more complex learning problem with large networks and convolutional layers.

They found that the gap between the test and training accuracy was smaller for the winning tickets and displayed that they generalized better. They also implemented dropout, which improves accuracy as it randomly disables a sample of the subnetwork on each training iteration. The learning rates improved suggesting that the iterative pruning strategy complements dropout.

The importance of winning ticket structure: The initialization that gives rise to a winning ticket is arranged in a particular sparse architecture. Since the winning tickets are uncovered through huge amount of training data, the authors suggest that the structure of their winning tickets encodes and inductive bias customized to the learning task at hand.

Authors find that original initialization values for winning tickets are key to success: When a winning ticket is reinitialized with random weight values, the model will train more slowly and achieves lower test accuracy. Authors suggest this is because the initialization values of winning weights may exist in areas of the loss landscape that have a viable path to the optimal point.

The improved generalization of winning tickets: The group found winning tickets with an even better accuracy than original network comparing to its training accuracy.

Limitations and Future Work: Although the iterative pruning strategy on classification task on small datasets displays the learning capacity of deep neural networks, it is computationally expensive and is inefficient in examining large datasets. In the future, authors hope to study other pruning methods such as structured and non-magnitude methods that could allow for an efficient(smaller pruned subnetworks), high accuracy performance.