

# Deconstructing Lottery Tickets: Zeros, Signs, and the Supermask

# Zeros and Sparsity

The paper investigates the role of zero weights in the lottery ticket networks. It finds that many zero weights are not just "dead" neurons but are, in fact, crucial for the network's training and performance. These zero weights contribute to the overall sparsity of the network.

# Weight Significance

The authors explore the idea that the signs (positive or negative) of weights play a meaningful role in the structure of lottery ticket networks. They found that zeroing out a weight and flipping its sign can have different effects on the network's performance.

# The Supermask

The concept of the "supermask" is introduced in this paper. The supermask is a binary mask applied to the weights of a network that determines which weights should be used (value of 1) and which should be pruned (value of 0). The supermask is learned through a structured pruning process, and it is the key to identifying winning lottery ticket subnetworks.

# Structured Pruning

The paper presents a structured pruning approach, where weights are pruned based on their importance to the network's performance. It's a more systematic way of finding the winning lottery tickets and is contrasted with unstructured pruning.

# Generalization vs. Performance

The paper discusses the trade-off between generalization and performance when using lottery ticket networks. It suggests that structured pruning can potentially lead to subnetworks that generalize better compared to random pruning methods.

# Training Speed

By analyzing the supermask, the paper proposes that it may be possible to train subnetworks faster, as a significant portion of the network can be frozen during training. This has implications for the efficiency of training deep neural networks.

# Robustness to Initialization

The paper investigates whether the existence of winning lottery tickets is robust to different weight initializations and training conditions. It suggests that structured pruning is more robust than unstructured pruning in this context.