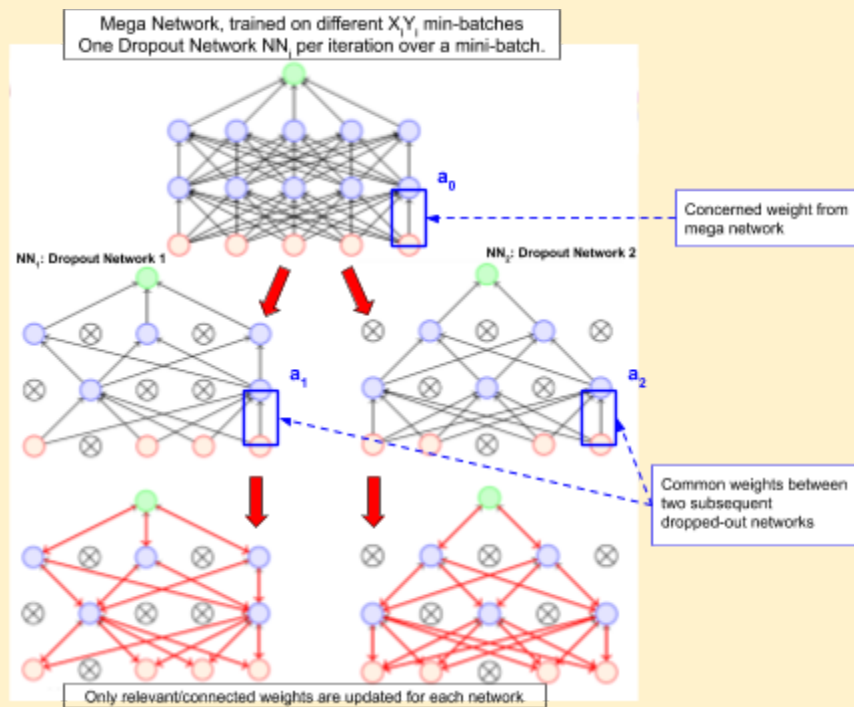


### How does weight sharing help?

How are weights shared between subsequent dropped-out networks

1. Let's take another look at the dropout training procedure



2. From the above, we can see that there are some common weights between the two dropped out networks.
  - a. Let  $a_0$  be the highlighted weight in the Mega Network
  - b. In the first iteration, for  $NN_1$ , the weight gets updated  $a_1 = a_0 + \eta \nabla a_0$
  - c. In the second iteration, for  $NN_2$ , the weight gets updated  $a_2 = a_1 + \eta \nabla a_0$
  - d. We can see how the weights are not freshly calculated for each iteration, and instead all **common weights are shared between the dropped-out networks.**
3. Consider a dataset of 1-million training samples. With a possible total of  $2^N$  dropped out networks, even for a small value of  $n=100$ , it is impossible to cover all possible networks even through multiple epochs over our dataset.  $2^N \gg 1$ -million. (recall that each-mini batch corresponds to 1 dropped-out network)
4. So, even with a small mini-batch size, most individual dropped-out networks will get trained only a small number of times. How do we ensure the efficient training of the entire set of dropped-out networks?
5. Now, instead of looking at it in terms of dropped-out networks getting trained, let's look at in terms of how many times each weight gets trained.
6. In short, each weight gets updated a suitable number of times. Here is the intuition behind it:
  - a. Each neuron will be present in half of all possible networks (Use truth table to verify)
  - b. The weights are shared across all networks
  - c. Hence each weight will get updated frequently during training.
7. Thus, we **do not need to train the network ( $2^N \times k$ ) times**, it is **sufficient if it is trained  $k$  times**, where  $k$  is the number of training steps/iterations whereby a set of weights are updated.
8. Now, the next step is figuring out how to combine the outputs of all the networks in the ensemble.