

Dropout:

- Neural Networks are likely to quickly Overfit a training dataset with few examples
- Large weights in neural networks are a sign of a more complex network that will overfit the training data



Solution



Ensemble of neural network with different model configuration



Is it an effective solution?



Require additional computation expense of training.



Solution

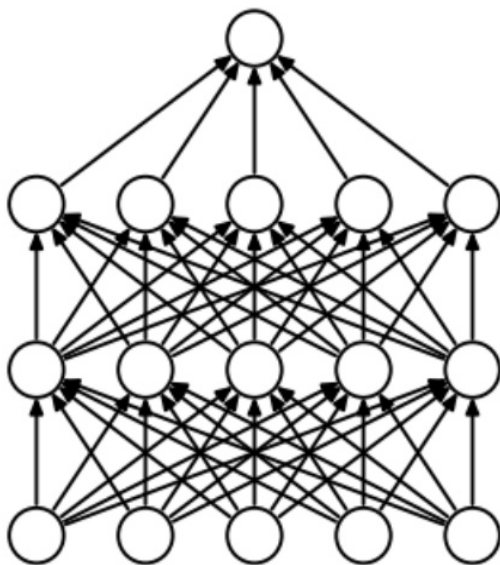


Drop Out

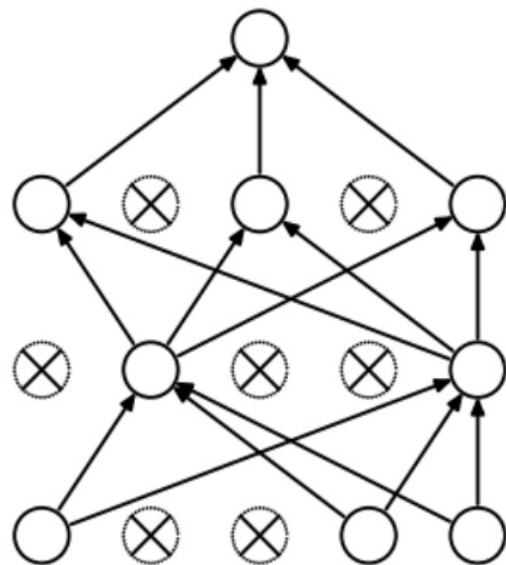
Randomly Drop Nodes

Dropout is a regularization method that approximates training a large number of neural networks with different architectures in Parallel .

- During training, some numbers of layers outputs are randomly Ignored.
- This has the effect of making the layer look like a layer with a different number of nodes and connectivity to the prior layer. In effect each update to the layer during training is performed with a different view of the configured layer.



(a) Standard Neural Net



(b) After applying dropout.

Some Observations:

1. Dropout forces a neural network to learn more robust features that are useful in conjunction with many different random subsets of the other neurons.
2. Dropout roughly doubles the number of iterations required to converge. However, training time for each epoch is less.

