Dropout Layer

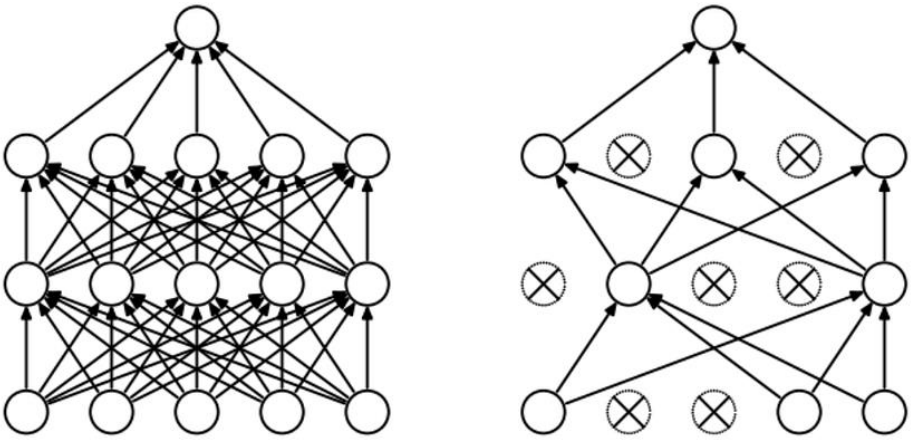
Dropout refers to ignoring units (i.e. neurons) during the training phase of certain set of neurons which is chosen at random. By "ignoring", it means, these units are not considered during a particular forward or backward pass.

At each training stage, individual nodes are either dropped out of the net with probability 1-p or kept with probability p, so that a reduced network is left; incoming and outgoing edges to a dropped-out node are also removed.

When you have training data, if you try to train your model too much, it might overfit, and when you get the actual test data for making predictions, it will not probably perform well. Dropout regularization is one technique used to tackle overfitting problems in deep learning.

A fully connected layer occupies most of the parameters, and hence, neurons develop co-dependency amongst each other during training which curbs the individual power of each neuron leading to over-fitting of training data.

In machine learning, regularization is way to prevent over-fitting. Dropout is an approach to regularization in neural networks which helps reducing interdependent learning amongst the neurons.



(a) Standard Neural Net (b) After applying dropout.

Why will dropout help with overfitting?

- It can't rely on one input as it might be randomly dropped out.
- Neurons will not learn redundant details of inputs

Dropout is implemented per layer in a neural network. It works with the vast majority of layers, including dense, fully connected, convolutional, and recurrent layers such as the long short-term memory network layer. Dropout can occur on any or all of the network's hidden layers as well as the visible or input layer. It is not used on the output layer.

- Dropout forces a neural network to learn more robust features that are useful in conjunction with many different random subsets of the other neurons.
- Dropout roughly doubles the number of iterations required to converge. However, training time for each epoch is less.
- With H hidden units, each of which can be dropped, we have 2^H possible models.