Deep Learning

6.2 Rectifiers and Dropout

Dr. Konda Reddy Mopuri
kmopuri@iittp.ac.in
Dept. of CSE, IIT Tirupati

Rectifiers

1. Use of ReLU improved the training greatly compared to tanh

Rectifiers

1. Use of ReLU improved the training greatly compared to tanh
2. Since the derivative of ReLU does not vanish for the positive activations (also the encodings become sparse)

Variants of ReLU

1. Leaky ReLU

Variants of ReLU

1. Leaky ReLU
2. PReLU (optimized parameter)

Variants of ReLU

1. Leaky ReLU
2. PReLU (optimized parameter)
3. PReLU (randomly parameterized)

Variants of ReLU

1. Leaky ReLU
2. PReLU (optimized parameter)
3. PReLU (randomly parameterized)
4. ELU (Exponential Linear Unit)

Dropout

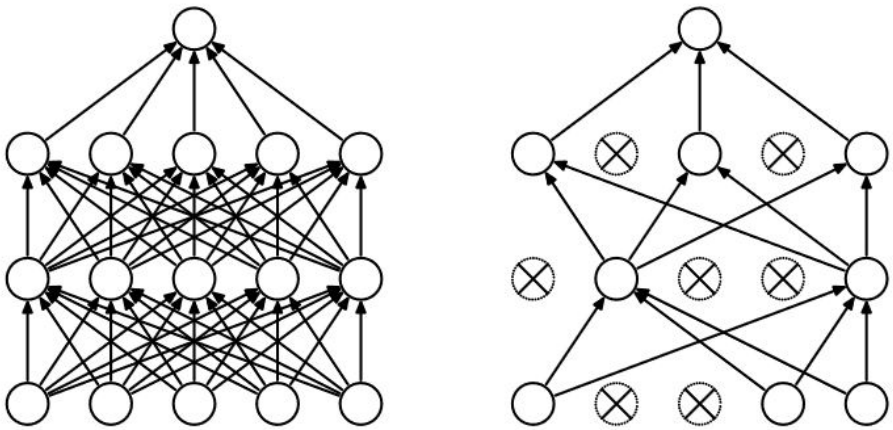
1. Key ideas and contributions in DL have been to engineer architectures for making them easier to train

Dropout

1. Key ideas and contributions in DL have been to engineer architectures for making them easier to train
2. Dropout is one such ('deep') regularization technique (Srivastava et al. 2014)

Dropout

1. During the forward pass, some of the units are randomly 'zeroed' out (activations are removed)



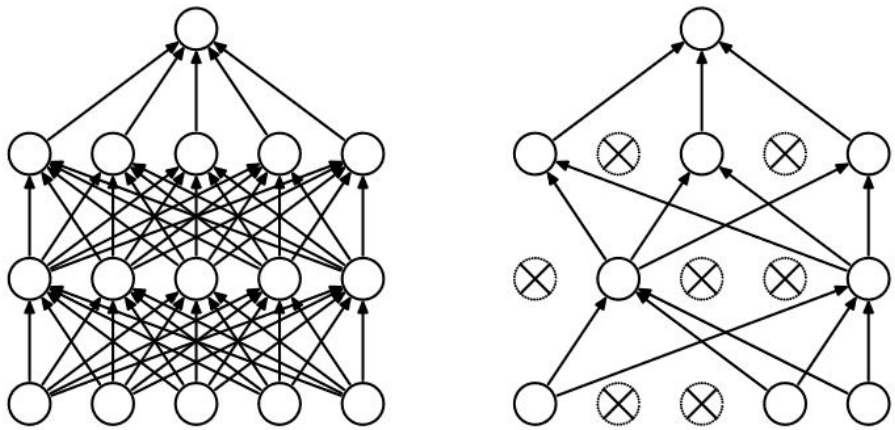
(a) Standard Neural Net (b) After applying dropout.

Figure 1: Dropout Neural Net Model. **Left**: A standard neural net with 2 hidden layers. **Right**: An example of a thinned net produced by applying dropout to the network on the left. Crossed units have been dropped.

Figure from Srivastava et al. 2014

Dropout

1. During the forward pass, some of the units are randomly 'zeroed' out (activations are removed)
2. Dropped units are randomly selected in each layer independent of others



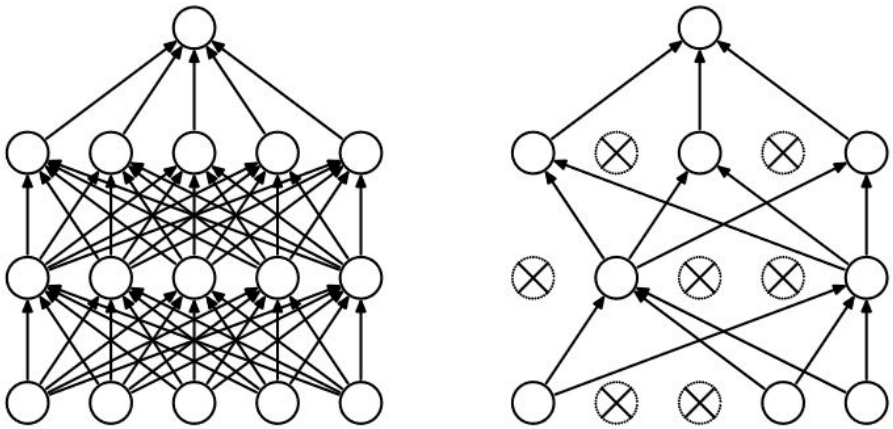
(a) Standard Neural Net (b) After applying dropout.

Figure 1: Dropout Neural Net Model. **Left**: A standard neural net with 2 hidden layers. **Right**: An example of a thinned net produced by applying dropout to the network on the left. Crossed units have been dropped.

Figure from Srivastava et al. 2014

Dropout

1. During the forward pass, some of the units are randomly 'zeroed' out (activations are removed)
2. Dropped units are randomly selected in each layer independent of others
3. Backpropagation happens through the remaining activations



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

Figure 1: Dropout Neural Net Model. **Left**: A standard neural net with 2 hidden layers. **Right**: An example of a thinned net produced by applying dropout to the network on the left. Crossed units have been dropped.

Figure from Srivastava et al. 2014

Dropout

1. Improves independence between the units

Dropout

1. Improves independence between the units
2. Distributes the representation among all the units

Dropout

1. Improves independence between the units
2. Distributes the representation among all the units
3. Avoids co-adaptation of the units in the architecture

Dropout

1. We will decide on which units/layers to use dropout, and with what probability p units are dropped.

Dropout

1. We will decide on which units/layers to use dropout, and with what probability p units are dropped.
2. For each sample, as many Bernoulli variables as units are sampled independently for dropping the units.

Dropout

1. The standard variant uses the 'inverted dropout'. It multiplies activations by $\frac{1}{(1-p)}$ during train and keeps the network untouched during test.