Learning representations by back-propagating errors

William Schultz

September 8, 2022

This paper [1] describes a new learning procedure, *back-propagation*, for networks of neurone-like units (i.e. neural networks).

Layered neural networks are defined as follows. They consist of a layer of input units at the bottom, any number of intermediate layers, and a layer of output units at the top. Connections within a layer or from higher to lower layers are forbidden (i.e. they are "feed-forward" networks). The total input x_j to unit j is a linear function of the outputs y_i of the units connected to j from the previous layer, and of the weights w_{ji} on the connections coming into j:

$$x_j = \sum_i y_i w_{ij} \tag{1}$$

A unit's final output is a real value, y_j, which is a non-linear function of the total input computed from the above equation

$$y_j = \frac{1}{1 + e^{-x_j}} \tag{2}$$

The overall goal is to find a set of weights so as to minimize error with respect to a set of given, finite set of samples (input-output cases). The total error E, is defined as

$$E = \frac{1}{2} \sum_x \sum_j (y_{j,c} - d_{j,c})^2 \tag{3}$$

where j ranges over the set of output values and c ranges over each input-output sample, y is the state of an output unit, and d is the desired state (as given by the sample).

Then, to minimize E by gradient descent, we must compute the partial derivative of E with respect to each weight in the network. This is simply the sum of the partial derivatives for each of the input-output cases.

Computing the gradient Forward pass: Add note.

The backward pass starts by computing $\partial E/\partial y$ for each of the output units. Differentiating Equation 3 gives us

$$\partial E/\partial = y_j - d_j$$

and then we can apply the chain rule to compute $\partial E/\partial x_j$

$$\partial E/\partial x_j = \partial E/\partial y_j \cdot dy_j/dx_j$$

TODO.

References

[1] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.