Mathematics of a Deep Neural Network

Gabe Millikan

October 26, 2021

Note

This is intended to act as a reference to use while writing code; not as a comprehensive guide to how neural networks work. The goal is only to describe the mathematical operations required to implement stochastic gradient descent in CUDA.

Variable	**Value**
S_n	Height of the nth layer
L	Number of layers in the network. Does not include input vector
S_L	Number of neurons in the last layer. Equal to number of outputs
S_1	Number of neurons in the first layer
S_0	Number of inputs to the network. Just a convention; there is no "0th layer", and the inputs are not part of the network shape
C	Cost function. Mean squared error is used exclusively in this paper
σ_n	Activation function of the nth layer (any differentiable function - not just sigmoid)
w_{nij}	jth weight of the ith neuron in the nth layer. j corresponds to a neuron in the previous layer, or in the input vector if $i = 1$
b_{ni}	The bias of ith neuron in the nth layer
o_{ni}	Output of the ith neuron in the nth layer, prior to activation
a_{ni}	Output of the ith neuron in the nth layer, after activation
a_{0i}	The ith input value. Just a convention; like S_0
T_i	The ith target value. In a perfectly trained network, this is equal to a_{Li}

$$o_{ni} = \left[\sum_{j=1}^{S_{n-1}} w_{nij} \cdot a_{(n-1)j}\right] + b_{ni}$$

$$a_{ni} = \sigma(o_{ni})$$

$$C = \frac{1}{S_L} \sum_{i=1}^{S_L} (a_{Li} - T_i)^2$$

$$\frac{\partial C}{\partial a_{Li}} = 2 \cdot \frac{1}{S_L} \cdot (a_{Li} - T_i)$$

$$\frac{\partial C}{\partial o_{Li}} = \frac{\partial C}{\partial a_{Li}} \cdot \frac{\partial a_{Li}}{\partial o_{Li}} = \frac{\partial C}{\partial a_{Li}} \cdot \sigma'(o_{Li})$$

$$\frac{\partial C}{\partial a_{ni}} = \sum_{j=1}^{S_{n+1}} \frac{\partial C}{\partial o_{(n+1)j}} \cdot w_{(n+1)ji}$$

$$\frac{\partial C}{\partial o_{ni}} = \frac{\partial C}{\partial a_{ni}} \cdot \frac{\partial a_{ni}}{\partial o_{ni}} = \frac{\partial C}{\partial a_{ni}} \cdot \sigma'(o_{ni})$$

$$\frac{\partial C}{\partial w_{nij}} = \frac{\partial C}{\partial o_{ni}} \cdot \frac{\partial o_{ni}}{\partial w_{nij}} = \frac{\partial C}{\partial o_{ni}} \cdot a_{(n-1)j}$$

$$\frac{\partial C}{\partial b_{ni}} = \frac{\partial C}{\partial o_{ni}} \cdot \frac{\partial o_{ni}}{\partial b_{ni}} = \frac{\partial C}{\partial o_{ni}} \cdot 1 = \frac{\partial C}{\partial o_{ni}}$$