

THE MATHEMATICS OF DEEP LEARNING*

VISHAL P. KASLIWAL¹

¹*Wave Computing*
42 W. Campbell Ave, # 301
Campbell, CA 95008, USA

(Received December 28, 2017; Revised December 28, 2017; Accepted December 28, 2017)

ABSTRACT

Deep Learning is a branch of Machine Learning in which ‘deep’ neural networks are used for various purposes.

Keywords: neural networks, training

1. PRELIMINARIES

2. NOTATION

We shall use \mathbf{a} to denote vectors and \mathbf{A} to denote tensors.

Tensors are associated with layers. The $p \times q$ tensor \mathbf{A} when associated with layer l is denoted $\mathbf{A}_{p \times q}^{[l]}$. The transpose of this tensor is denoted by $(\mathbf{A}_{p \times q}^{[l]})^\top$. If the tensor is an input tensor to a layer, it may additionally have a minibatch number, m , and instance within the minibatch, i , associated with it. Thus $(\mathbf{A}_{p \times q}^{[l]\{m\}(i)})^\top$ is the transpose of the i -th training example from the m -th minibatch in the l -th layer from the input tensor \mathbf{A} which is p rows high by q columns long.

The $n^{[l]}$ operator returns the number of nodes, i.e. the width, of the l -th layer. Traditionally, the depth of the neural network is denoted by L , and hence we must have $0 \leq l \leq L$. The 0-th layer is a pure input layer i.e. it has no transformation associated with it.

3. ACTIVATION FUNCTIONS

3.1. *Sigmoid Activation Function*

The sigmoid activation function takes the form

$$g(z) = \sigma(z) = \frac{1}{1 + e^{-z}}. \quad (1)$$

The derivative of the sigmoid activation function is

$$\frac{d\sigma}{dz} = \sigma(1 - \sigma). \quad (2)$$

3.1.1. *Derivation*

By definition

$$\sigma(z) = \frac{1}{1 + e^{-z}} = (1 + e^{-z})^{-1},$$

and so

$$\frac{d\sigma(z)}{dz} = -(1 + e^{-z})^{-2} \frac{d(1 + e^{-z})}{dz}.$$

But this is just

$$-(1 + e^{-z})^{-2}(-e^{-z}) = \sigma^2 e^{-z}.$$

Now $e^{-z} = \frac{1}{\sigma} - 1 = \frac{1-\sigma}{\sigma}$ and so

$$\frac{d\sigma(z)}{dz} = \sigma^2 e^{-z} = \sigma^2 \frac{1-\sigma}{\sigma} = \sigma(1-\sigma),$$

Q.E.D.

4. OUTPUT LAYERS
5. FEED-FORWARD LAYERS
6. CONVOLUTIONAL LAYERS
7. POOLING LAYERS

APPENDIX

REFERENCES