The Backpropagation Algorithm

John Purcell

May 11, 2021

Abstract

Finding weight gradients via calculus.

1 Notation

First, let's describe our notation for our neural network nodes, weights, biases and other quantities.

The activation (output) of a neuron will be referred to via the lowercase letter a. We will refer to weights via lowercase w's. Similarly, biases will be referred to by lowercase b's.

Our neural network consists of a number of layers, each layer containing multiple neurons. We will number the layers relative to some arbitrary layer **L**. So, if we want to refer to the next layer after **L** (the layer that takes **L**'s output as its input), we refer to layer **L+1**. The previous layer supplies layer **L**'s input, and we will refer to that as layer **L-1**.

Quantities in the layer **L** will be denoted using an uppercase superscript.

The vector of all