1.4 Deriving Backprop

For too long, I did not know where the backprop equations came from. Let's derive them today. (Don't memorize, understand!)

1.4.1 Forward Pass

First note the equations for a forward pass at a given layer l:

$$z^{[l]} = w^{[l]}a^{[l-1]} + b^{[l]} \tag{7}$$

$$a^{[l]} = g(z^{[l]}) \tag{8}$$

where g is your activation function. Note the partial derivatives:

$$\frac{\partial z^{[l]}}{\partial w^{[l]}} = a^{[l-1]},\ \frac{\partial z^{[l]}}{\partial a^{[l-1]}} = w^{[l]},\ \text{and}\ \frac{\partial a^{[l]}}{\partial z^{[l]}} = g^{[l]\prime}(z^{[l]})\ .$$

1.4.2 Backward Pass

In order to calculate dw and db, the gradient of the network, we need to first find da and dz, the gradients of the activations and pre-activations. Remember that everything is with respect to the loss function L.

$$dz^{[l]} = \frac{\partial L}{\partial z^{[l]}} = \frac{\partial L}{\partial a^{[l]}} \cdot \frac{\partial a^{[l]}}{\partial z^{[l]}} = da^{[l]} \cdot g'(z^{[l]}) \tag{9}$$

$$dw^{[l]} = \frac{\partial L}{\partial w^{[l]}} = \frac{\partial L}{\partial z^{[l]}} \cdot \frac{\partial z^{[l]}}{\partial w^{[l]}} = dz^{[l]} \cdot a^{[l-1]} \tag{10}$$

$$db^{[l]} = \frac{\partial L}{\partial b^{[l]}} = \frac{\partial L}{\partial z^{[l]}} \cdot \frac{\partial z^{[l]}}{\partial b^{[l]}} = dz^{[l]} \tag{11}$$

$$da^{[l-1]} = \frac{\partial L}{\partial a^{[l-1]}} = \frac{\partial L}{\partial z^{[l]}} \cdot \frac{\partial z^{[l]}}{\partial a^{[l-1]}} = dz^{[l]} \cdot w^{[l]} \tag{12}$$

To get $da^{[l]}$ for the first time (in the last layer of the network), directly compute it from the loss function. For instance, for cross-entropy loss $L = -(y\log(a) + (1-y)\log(1-a))$,

$$da^{[l]} = \frac{\partial L}{\partial a^{[l]}} = -\frac{y}{a} + \frac{1-y}{1-a}\ .$$