Biological motivation

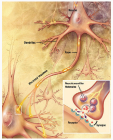

Image source: en.wikipedia.org

- A Human brain contains approximately 86 billion neurons.
- 10^{14} to 10^{15} synapses connect these neurons.
- Neurons receive inputs from dendrites.
- and can produce output signals along its axon.
- Axons connect neurons, modeled by weighting inputs wx.
- Neuron inputs can be inhibitive (negative weight) or
- excitatory (positive weight).
- If enough inputs excite a neuron, it fires.
- The activation function aims to mimic this behavior.
- Even though neural networks started out as biologically motivated,
- engineering efforts have since diverged from biology.

The gradient of a dense layer

The chain rule tells us the gradients for the dense layer [Nie15]

$$\delta \mathbf{W} = [f'(\bar{\mathbf{h}}) \odot \triangle]\mathbf{x}^T, \qquad \delta \mathbf{b} = f'(\bar{\mathbf{h}}) \odot \triangle, \tag{5}$$

$$\delta \mathbf{x} = \mathbf{W}^T[f'(\bar{\mathbf{h}}) \odot \triangle], \tag{6}$$

where $\odot$ is the element-wise product. δ denotes the cost function gradient for the value following it [Gre+16].
With $\delta \mathbf{W} \in \mathbb{R}^{m,n}$, $\delta \mathbf{x} \in \mathbb{R}^n$ and $\delta \mathbf{b} \in \mathbb{R}^m$. **Modern libraries will take care of these computations for you!**

On the board, derive: Recall the chain rule $(g(h(x)))' = g'(h(x)) \cdot h'(x)$. For the activation function, we have,

$$\mathbf{h} = f(\bar{\mathbf{h}}) \tag{7}$$

$$\Rightarrow \delta \bar{\mathbf{h}} = f'(\bar{\mathbf{h}}) \odot \triangle \tag{8}$$

For the weight matrix,

$$\bar{\mathbf{h}} = \mathbf{W}\mathbf{x} + \mathbf{b} \tag{9}$$

$$\Rightarrow \delta \mathbf{W} = \delta \bar{\mathbf{h}} \mathbf{x}^T = [f'(\bar{\mathbf{h}}) \odot \triangle]\mathbf{x}^T \tag{10}$$

For the bias,

$$\bar{\mathbf{h}} = \mathbf{W}\mathbf{x} + \mathbf{b} \tag{11}$$

$$\Rightarrow \delta \mathbf{b} = 1 \odot \delta \bar{\mathbf{h}} = [f'(\bar{\mathbf{h}}) \odot \triangle] \tag{12}$$

Gradients and cross-entropy

If a sigmoidal activation function produced $\mathbf{o}$ the gradients can be computed using [Nie15; Bis06]

$$\frac{\partial C_{ce}}{\partial \mathbf{h}} = \sigma(\mathbf{o}) - \mathbf{y} = \triangle_{ce} \tag{19}$$

Following [Nie15], substitute $\sigma(\mathbf{o})$ into eq 18.