Forward Propagation and Back Propagation

Vectorisation form:

Forward propagation	**Back Propagation**
$\mathbf{z}^{[L]} = \mathbf{w}^{[L]} \cdot \mathbf{a}^{[L-1]} + b^{[L]}$ $\mathbf{a}^{[L]} = f^{[L]}(\mathbf{z}^{[L]})$	$d\mathbf{z}^{[L]} = d\mathbf{a}^{[L]} * f^{[L]'}(\mathbf{z}^{[L]})$ $d\mathbf{w}^{[L]} = \frac{1}{m} d\mathbf{z}^{[L]} \cdot \mathbf{a}^{[L-1]}$ $db^{[L]} = \frac{1}{m} \sum_d^D d\mathbf{z}^{[L]}$ where D denotes the number of features. $d\mathbf{a}^{[L-1]} = \mathbf{w}^{[L]T} \cdot d\mathbf{z}^{[L]}$

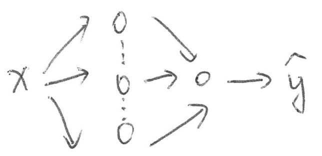
Notice,

$\mathbf{a}^0 = \mathbf{x}$

$f^{[L]}$ is the activation function of layer L.

$f^{[L]'}$ is the gradient of the activation function of layer L.

Ex



$$* \quad \sigma(x) = \frac{1}{1+e^{-x}}$$

$$\sigma'(x) = \sigma(x)(1-\sigma(x))$$

$$x \rightarrow z^{[1]} = W^{[1]}x + b^{[1]} \rightarrow a^{[1]} = \sigma(z^{[1]}) \rightarrow z^{[2]} = W^{[2]}a^{[1]} + b^{[2]} \rightarrow a^{[2]} = \sigma(z^{[2]}) \rightarrow \mathcal{L}(a^{[2]}, y)$$

$$\mathcal{L}(\hat{y}, y) = \mathcal{L}(a^{[2]}, y) = -(y \log a^{[2]} + (1-y)\log(1-a^{[2]}))$$

$$\frac{\partial \mathcal{L}}{\partial a^{[2]}} = -\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}$$

$$dz^{[2]} = \frac{\partial \mathcal{L}}{\partial z^{[2]}} = \frac{\partial \mathcal{L}}{\partial a^{[2]}} \frac{\partial a^{[2]}}{\partial z^{[2]}} = \left(-\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}\right) \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}} = \left(-\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}\right) \cdot a^{[2]}(1-a^{[2]}) = a^{[2]} - y$$

$$dz^{[1]} = \frac{\partial \mathcal{L}}{\partial z^{[1]}} \cancel{= \frac{\partial \mathcal{L}}{\partial a^{[1]}}} = \underbrace{\boxed{\frac{\partial \mathcal{L}}{\partial a^{[2]}} \frac{\partial a^{[2]}}{\partial z^{[2]}}}}_{dz^{[2]}} \frac{\partial z^{[2]}}{\partial a^{[1]}} \frac{\partial a^{[1]}}{\partial z^{[1]}} = dz^{[2]} \cdot W^{[2]} \cdot a^{[1]}(1-a^{[1]})$$

$$dW^{[1]} = dz^{[1]} \cdot \frac{\partial z^{[1]}}{\partial W^{[1]}} = dz^{[1]} \cdot x$$

$$db^{[1]} = dz^{[1]} \cdot \frac{\partial z^{[1]}}{\partial b^{[1]}} = dz^{[1]} \cdot 1 = dz^{[1]}$$

$$dW^{[2]} = \frac{\partial \mathcal{L}}{\partial W^{[2]}} = \frac{\partial \mathcal{L}}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial W^{[2]}} = dz^{[2]} \cdot a^{[1]} = (a^{[2]} - y)a^{[1]}$$

$$db^{[2]} = \frac{\partial \mathcal{L}}{\partial b^{[2]}} = \frac{\partial \mathcal{L}}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial b^{[2]}} = dz^{[2]} \cdot 1 = a^{[2]} - y$$