I. BACKWARD PROPAGATION

$$C = (a^{(L)} - y)^2. \tag{1}$$

$$a^{(L)} = \sigma(z^{(L)}). \tag{2}$$

$$z^{(L)} = w^{(L)}a^{(L-1)} + b^{(L)}. \tag{3}$$

$$\frac{\partial C}{\partial a^{(L)}} = 2(a^{(L)} - y). \tag{4}$$

$$\frac{\partial a^{(L)}}{\partial z^{(L)}} = \sigma'(z^{(L)}). \tag{5}$$

$$\frac{\partial z^{(L)}}{\partial w^{(L)}} = a^{(L-1)} \tag{6}$$

$$\frac{\partial C}{\partial w^{(L)}} = \frac{\partial z^{(L)}}{\partial w^{(L)}}\frac{\partial a^{(L)}}{\partial z^{(L)}}\frac{\partial C}{\partial a^{(L)}}. \tag{7}$$

$$= a^{(L-1)}\sigma'(z^{(L)})2(a^{(L)} - y). \tag{8}$$

$$\frac{\partial C}{\partial b^{(L)}} = 1 \cdot \sigma'(z^{(L)})2(a^{(L)} - y). \tag{9}$$

$$\tag{10}$$

In the next-to-last layer,

$$\frac{\partial C}{\partial w^{(L-1)}} = \frac{\partial z^{(L-1)}}{\partial w^{(L-1)}}\frac{\partial a^{(L-1)}}{\partial z^{(L-1)}}\frac{\partial z^{(L)}}{\partial a^{(L-1)}}\frac{\partial a^{(L)}}{\partial z^{(L)}}\frac{\partial C}{\partial a^{(L)}}. \tag{11}$$

$$\tag{12}$$

In a general case,

$$z_j^{(L)} = \sum_k w_{jk}^{(L)} a_k^{(L-1)} + b_j^{(L)} \tag{13}$$

$$\nabla C = \begin{bmatrix} \frac{\partial C}{\partial w^{(1)}} & \frac{\partial C}{\partial b^{(1)}} & \frac{\partial C}{\partial w^{(2)}} & \frac{\partial C}{\partial b^{(2)}} \end{bmatrix}^T \tag{14}$$

Start with the forward propagation equations

$$\mathbf{z}^{(1)} = W^{(1)}\mathbf{x} + \mathbf{b}^{(1)}.\ (16,) \tag{15}$$

$$\mathbf{a}^{(1)} = \sigma^{(1)}(\mathbf{z^{(1)}}).\ (16,) \tag{16}$$

$$\mathbf{z}^{(2)} = W^{(2)}\mathbf{a^{(1)}} + \mathbf{b}^{(2)}.\ (10,) \tag{17}$$

$$\mathbf{a}^{(2)} = \sigma^{(2)}(\mathbf{z^{(2)}}).\ (10,) \tag{18}$$

$$C = |\mathbf{a}^{(2)} - \mathbf{y}|^2 = (a_0^{(2)} - y_0)^2 + (a_1^{(2)} - y_1)^2 + \cdots + (a_9^{(2)} - y_9)^2. \tag{19}$$

$$\begin{bmatrix} z_0^{(2)} \\ z_1^{(2)} \\ \vdots \\ z_9^{(2)} \end{bmatrix} = \begin{bmatrix} w_{0,0} & w_{0,1} & \cdots & w_{0,16} \\ w_{1,0} & w_{1,1} & \cdots & w_{1,16} \\ \vdots & \vdots & \ldots & \vdots \\ w_{9,0} & w_{9,1} & \cdots & w_{9,16} \end{bmatrix} \begin{bmatrix} a_0^{(1)} \\ a_1^{(1)} \\ \vdots \\ a_{16}^{(1)} \end{bmatrix} = \begin{bmatrix} w_{0,0}a_0^{(1)} + w_{0,1}a_1^{(1)} + \cdots + w_{0,16}a_{16}^{(1)} \\ w_{1,0}a_0^{(1)} + w_{1,1}a_1^{(1)} + \cdots + w_{1,16}a_{16}^{(1)} \\ \vdots \\ w_{9,0}a_0^{(1)} + w_{9,1}a_1^{(1)} + \cdots + w_{9,16}a_{16}^{(1)} \end{bmatrix} \tag{20}$$

Want to find

$$\frac{\partial C}{\partial w_{jk}^{(2)}} = \frac{\partial z_j^{(2)}}{\partial w_{jk}^{(2)}} \frac{\partial a_j^{(2)}}{\partial z_j^{(2)}} \frac{\partial C}{\partial a_j^{(2)}}. \tag{21}$$

$$= a_k^{(1)} \sigma^{(2)\prime}(z_j^{(2)}) 2(a_j^{(2)} - y_j) \tag{22}$$

$$\frac{\partial C}{\partial b_j^{(2)}} = \frac{\partial z_j^{(2)}}{\partial b_j^{(2)}} \frac{\partial a_j^{(2)}}{\partial z_j^{(2)}} \frac{\partial C}{\partial a_j^{(2)}}. \tag{23}$$

$$= \sigma^{(2)\prime}(z_j^{(2)}) 2(a_j^{(2)} - y_j) \tag{24}$$

$$\frac{\partial C}{\partial w_{jk}^{(1)}} = \frac{\partial z_j^{(1)}}{\partial w_{jk}^{(1)}} \frac{\partial a_j^{(1)}}{\partial z_j^{(1)}} \frac{\partial C}{\partial a_j^{(1)}} \tag{25}$$

$$= a_k^{(0)} \sigma^{(1)\prime}(z_j^{(1)}) \frac{\partial C}{\partial a_j^{(1)}} \tag{26}$$

$$\tag{27}$$

$$\frac{\partial C}{\partial a_j^{(1)}} = \frac{\partial z_m^{(2)}}{\partial a_j^{(1)}} \frac{\partial a_m^{(2)}}{\partial z_m^{(2)}} \frac{\partial C}{\partial a_m^{(2)}} \tag{28}$$

$$= w_{m,j}^{(2)} \sigma^{(2)\prime}(z_m^{(2)}) 2(a_m^{(2)} - y_m) \tag{29}$$