

SINGLE PARAMETER DERIVATIVES.

DERIVATIVES OF THE LAST LAYER:

$$\frac{\partial cost_s}{\partial w_{ik}^L} = \frac{\partial cost(a_i^L, y_i)}{\partial a_i^L} \cdot \frac{\partial \sigma(z_i^L)}{\partial z_i^L} \cdot a_k^{L-1}$$

$$\frac{\partial cost_s}{\partial b_i^L} = \frac{\partial cost(a_i^L, y_i)}{\partial a_i^L} \cdot \frac{\partial \sigma(z_i^L)}{\partial z_i^L}$$

$$\frac{\partial cost_s}{\partial a_k^{L-1}} = \sum_{j=0}^{n_L-1} \frac{\partial cost(a_j^L, y_j)}{\partial a_j^L} \cdot \frac{\partial \sigma(z_j^L)}{\partial z_j^L} \cdot w_{jk}^L$$

GENERALIZATION FOR ALL LAYERS:

$$\frac{\partial cost_s}{\partial z_i^L} = \frac{\partial cost}{\partial a_i^L} \cdot \frac{\partial a_i^L}{\partial z_i^L} = \frac{\partial cost}{\partial a_i^L} \cdot \frac{\partial \sigma(z_i^L)}{\partial z_i^L}$$

$$\frac{\partial cost_s}{\partial w_{ik}^L} = \frac{\partial cost}{\partial a_i^L} \cdot a_k^{L-1}$$

$$\frac{\partial cost_s}{\partial b_i^L} = \frac{\partial cost}{\partial a_i^L}$$

$$\frac{\partial cost_s}{\partial a_k^{L-1}} = \sum_{j=0}^{n_L-1} \frac{\partial cost}{\partial a_j^L} \cdot w_{jk}^L$$

DERIVATIVES FOR ACTIVATION LAYERS.

$$\frac{\partial cost_s}{\partial z^L} = \begin{bmatrix} \frac{\partial cost_s}{\partial z_0^L} \\ \frac{\partial cost_s}{\partial z_1^L} \\ \vdots \\ \frac{\partial cost_s}{\partial z_j^L} \end{bmatrix} = \begin{bmatrix} \frac{\partial cost}{\partial a_0^L} \cdot \frac{\partial \sigma(z_0^L)}{\partial z_0^L} \\ \frac{\partial cost}{\partial a_1^L} \cdot \frac{\partial \sigma(z_1^L)}{\partial z_1^L} \\ \vdots \\ \frac{\partial cost}{\partial a_j^L} \cdot \frac{\partial \sigma(z_j^L)}{\partial z_j^L} \end{bmatrix} = \begin{bmatrix} \frac{\partial cost}{\partial a_0^L} \\ \frac{\partial cost}{\partial a_1^L} \\ \vdots \\ \frac{\partial cost}{\partial a_j^L} \end{bmatrix} \odot \begin{bmatrix} \frac{\partial \sigma(z_0^L)}{\partial z_0^L} \\ \frac{\partial \sigma(z_1^L)}{\partial z_1^L} \\ \vdots \\ \frac{\partial \sigma(z_j^L)}{\partial z_j^L} \end{bmatrix}$$

$$\frac{\partial cost_s}{\partial z^L} = \frac{\partial cost}{\partial a^L} \odot \frac{\partial \sigma(z^L)}{\partial z^L}$$

DERIVATIVES FOR DENSE LAYERS.

WEIGHT MATRIX:

$$\frac{\partial cost_s}{\partial w^L} = \begin{bmatrix} \frac{\partial cost_s}{\partial w_{00}^L} & \frac{\partial cost_s}{\partial w_{01}^L} & \dots & \frac{\partial cost_s}{\partial w_{0k}^L} \\ \frac{\partial cost_s}{\partial w_{10}^L} & \frac{\partial cost_s}{\partial w_{11}^L} & \dots & \frac{\partial cost_s}{\partial w_{1k}^L} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial cost_s}{\partial w_{j0}^L} & \frac{\partial cost_s}{\partial w_{j1}^L} & \dots & \frac{\partial cost_s}{\partial w_{jk}^L} \end{bmatrix}$$

$$\frac{\partial cost_s}{\partial w^L} = \begin{bmatrix} \frac{\partial cost}{\partial a_0^L} \cdot a_0^{L-1} & \frac{\partial cost}{\partial a_0^L} \cdot a_1^{L-1} & \dots & \frac{\partial cost}{\partial a_0^L} \cdot a_k^{L-1} \\ \frac{\partial cost}{\partial a_1^L} \cdot a_0^{L-1} & \frac{\partial cost}{\partial a_1^L} \cdot a_1^{L-1} & \dots & \frac{\partial cost}{\partial a_1^L} \cdot a_k^{L-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial cost}{\partial a_j^L} \cdot a_0^{L-1} & \frac{\partial cost}{\partial a_j^L} \cdot a_1^{L-1} & \dots & \frac{\partial cost}{\partial a_j^L} \cdot a_k^{L-1} \end{bmatrix}$$

$$\frac{\partial cost_s}{\partial w^L} = \begin{bmatrix} \frac{\partial cost}{\partial a_0^L} \\ \frac{\partial cost}{\partial a_1^L} \\ \vdots \\ \frac{\partial cost}{\partial a_j^L} \end{bmatrix} [a_0^{L-1} \quad a_1^{L-1} \quad \dots \quad a_k^{L-1}]$$

$$\frac{\partial cost_s}{\partial w^L} = \frac{\partial cost}{\partial a^L} \cdot (a^{L-1})^T$$

BIAS VECTOR:

$$\frac{\partial cost_s}{\partial b^L} = \begin{bmatrix} \frac{\partial cost}{\partial b_0^L} \\ \frac{\partial cost}{\partial b_1^L} \\ \vdots \\ \frac{\partial cost}{\partial b_j^L} \end{bmatrix} = \begin{bmatrix} \frac{\partial cost}{\partial a_0^L} \\ \frac{\partial cost}{\partial a_1^L} \\ \vdots \\ \frac{\partial cost}{\partial a_j^L} \end{bmatrix}$$

$$\frac{\partial cost_s}{\partial b^L} = \frac{\partial cost}{\partial a^L}$$

COST FOR PREVIOUS LAYER:

$$\frac{\partial cost_s}{\partial a^{L-1}} = \begin{bmatrix} \frac{\partial cost}{\partial a_0^{L-1}} \\ \frac{\partial cost}{\partial a_1^{L-1}} \\ \vdots \\ \frac{\partial cost}{\partial a_j^{L-1}} \end{bmatrix} = \begin{bmatrix} \sum_{j=0}^{n_L-1} \frac{\partial cost}{\partial a_j^L} \cdot w_{j0}^L \\ \sum_{j=0}^{n_L-1} \frac{\partial cost}{\partial a_j^L} \cdot w_{j1}^L \\ \vdots \\ \sum_{j=0}^{n_L-1} \frac{\partial cost}{\partial a_j^L} \cdot w_{jk}^L \end{bmatrix}$$

$$\frac{\partial cost_s}{\partial a^{L-1}} = \begin{bmatrix} \frac{\partial cost}{\partial a_0^L} \cdot w_{00}^L + \frac{\partial cost}{\partial a_1^L} \cdot w_{10}^L \dots + \frac{\partial cost}{\partial a_j^L} \cdot w_{j0}^L \\ \frac{\partial cost}{\partial a_0^L} \cdot w_{01}^L + \frac{\partial cost}{\partial a_1^L} \cdot w_{11}^L \dots + \frac{\partial cost}{\partial a_j^L} \cdot w_{j1}^L \\ \vdots \\ \frac{\partial cost}{\partial a_0^L} \cdot w_{0k}^L + \frac{\partial cost}{\partial a_1^L} \cdot w_{1k}^L \dots + \frac{\partial cost}{\partial a_j^L} \cdot w_{jk}^L \end{bmatrix}$$

$$\frac{\partial cost_s}{\partial a^{L-1}} = \begin{bmatrix} w_{00}^L & w_{10}^L & \dots & w_{j0}^L \\ w_{01}^L & w_{11}^L & \dots & w_{j1}^L \\ \vdots & \vdots & \ddots & \vdots \\ w_{0k}^L & w_{1k}^L & \dots & w_{jk}^L \end{bmatrix} \begin{bmatrix} \frac{\partial cost}{\partial a_0^L} \\ \frac{\partial cost}{\partial a_1^L} \\ \vdots \\ \frac{\partial cost}{\partial a_j^L} \end{bmatrix}$$

$$\frac{\partial cost_s}{\partial a^{L-1}} = (w^L)^T \cdot \frac{\partial cost}{\partial a^L}$$