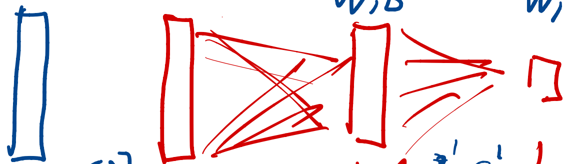

$\hat{y} = a^{[2]}$

$\mathcal{L}(a^{[2]}, y)$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

$$\tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

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

$$g'^{[2]} = \sigma'(z) = \underbrace{\sigma(z)}^{a^{[2]}}\left[\underbrace{1 - \sigma(z)}^{1 - a^{[2]}}\right]$$

$$g'^{[1]} = \tanh'(z) = \underbrace{1 - \tanh^2(z)}_{1 - a^{[1]2}}$$

$$\frac{d}{dz}\left[\frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}\right] = \frac{(e^{z} + e^{-z})(e^{z} + e^{-z}) - (e^{z} - e^{-z})(e^{z} - e^{-z})}{(e^{z} + e^{-z})^2}$$

$$\frac{(e^{z} + e^{-z})^2}{(e^{z} + e^{-z})^2} - \left[\frac{(e^{z} - e^{-z})}{(e^{z} + e^{-z})}\right]^2 = 1 - \tanh^2(z)$$

$$\frac{\partial L}{\partial a^{[2]}} = \frac{\partial}{\partial a^{[2]}} - \left[y \log(a^{[2]}) + (1-y)\log(1-a^{[2]}) \right] = \boxed{\frac{a^{[2]}-y}{a^{[2]}(1-a^{[2]})}}$$

$$\frac{\partial L}{\partial z^{[2]}} = \underbrace{\frac{\partial L}{\partial a^{[2]}}}_{\checkmark} \cdot \underbrace{\frac{\partial a^{[2]}}{\partial z^{[2]}}}_{g'^{[2]}} = \frac{a^{[2]}-y}{\cancel{a^{[2]}(1-a^{[2]})}} \cdot \cancel{a^{[2]}(1-a^{[2]})}$$

$$= a^{[2]} - y$$

$$\boxed{A^{[2]} - Y}$$

$$z^{[2]} = W^{[2]} a^{[1]} + b^{[2]} \qquad a^{[2]} = g^{[2]}\left(z^{[2]}\right)$$

$$\frac{\partial L}{\partial w^{2}} = \underbrace{\frac{\partial L}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}}}_{dz_2} \cdot \frac{\partial z^{[2]}}{\partial W^{[2]}} = dz_2 \cdot a^{[1]} \Rightarrow \boxed{\frac{1}{m} dz_2 A^{[1]T}}$$

$$\frac{\partial z^{[2]}}{\partial w^{[2]}} = \frac{\partial}{\partial w^{[2]}}\left[w^{[2]} a^{[1]} + b^{[2]} \right] = a^{[1]}$$

$$\frac{\partial L}{\partial b^{[2]}} = dz_2 \cdot \frac{\partial z^{[2]}}{\partial b^{[2]}} = dz_2 \Rightarrow \boxed{np.sum(dz_2, axis=1)}$$

$$\frac{\partial z^{[2]}}{\partial b^{[2]}} = 1$$

$$\frac{\partial L}{\partial a^{[1]}} = \underbrace{\frac{\partial L}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}}}_{dz_2} \cdot \frac{\partial z^{[2]}}{\partial a^{[1]}} = dz_2 \, W^{[2]} \qquad \boxed{W^{T[2]} dz_2}$$

$$\frac{\partial z^{[2]}}{\partial a^{[1]}} = \frac{\partial}{\partial a^{[1]}} \left[W^{[2]} a^{[1]} + b^{[2]} \right] = W^{[2]}$$

$$\frac{\partial L}{\partial z^{[1]}} = \underbrace{\frac{\partial L}{a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}} \cdot \frac{\partial z^{[2]}}{\partial a^{[1]}}} \cdot \underbrace{\frac{\partial a^{[1]}}{\partial z^{[1]}}}_{g'^{[1]}} = dz_2 \, W^{[2]} \left(1 - a^{[1]^2}\right)$$

$$\underbrace{\boxed{W^{T[2]} dz_2 \odot \left(1 - A^{[1]^2}\right)}}_{dz_1}$$

$$z^{[1]} = W^{[1]} a^{[0]} + b^{[1]} \qquad a^{[1]} = g^{[1]}(z^{[1]})$$

$$\frac{\partial L}{\partial W^{[1]}} = \underbrace{\frac{\partial L}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}} \cdot \frac{\partial z^{[2]}}{\partial a^{[1]}} \cdot \frac{\partial a^{[1]}}{\partial z^{[1]}}}_{dz_1} \cdot \frac{\partial z^{[1]}}{\partial W^{[1]}} = dz_1 \, a^{[0]}$$

$$\Rightarrow \boxed{\frac{1}{m} dz_1 A^{[0]T}}$$

$$\frac{\partial z^{[1]}}{\partial W^{[1]}} = \frac{\partial}{\partial W^{[1]}} \left[W^{[1]} a^{[0]} + b^{[1]} \right] = a^{[0]}$$

$$\frac{\partial L}{\partial b^{[1]}} = dz_1 \frac{\partial z^{[1]}}{\partial b^{[1]}} = dz_1 \quad \Rightarrow \boxed{\frac{1}{m} np.sum(dz_1 \ axis=1)}$$

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