1

$J = J(g_1, \dots g_k)$, $g_j = g_j(\theta_1, \dots \theta_p)$ $\forall j \in \{1, \dots k\}$

$$\boxed{\frac{\partial J}{\partial \theta_i} = \sum_{j=1}^{k} \frac{\partial J}{\partial g_j} \frac{\partial g_j}{\partial \theta_i}}$$

$x = (x_1 \dots x_d)^T$

$z = (z_1 \dots z_m)^T$

$a \in (a_1 \dots a_m)^T$

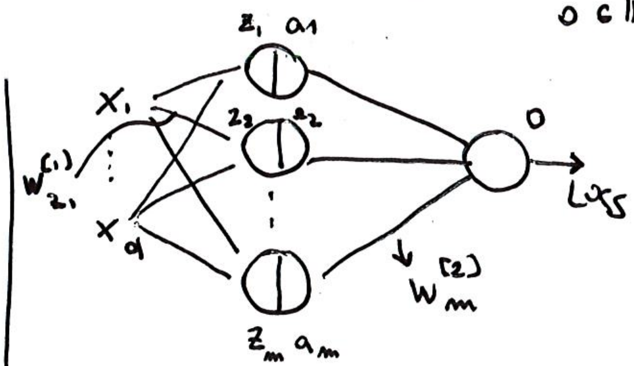
$b^{[1]} = (b_1^{[1]} \dots b_m^{[1]})$

$b^{[2]} \in \mathbb{R}$

$o \in \mathbb{R}$

$W^{[2]} \in \mathbb{R}^{1 \times m}$

$= (W_1^{[2]} \dots W_m^{[2]})^T$



$z = W^{[1]} x + b^{[1]}$

$m\times1 = m\times d \quad d\times1 \quad m\times1$

$a = RELU(z)$

$h_\theta(x)$

$o = W^{[2]} a + b^{[2]}$

$J = \frac{1}{2}(y - o)^2$

$\left(o = \sum_{i=1}^{m} W_i^{[2]} a_i + b^{[2]}\right)$

$$\boxed{\frac{\partial J}{\partial W^{[2]}}} \qquad \frac{\partial J}{\partial W_i^{[2]}} = \frac{\partial J}{\partial o} \frac{\partial o}{\partial W_i^{[2]}} = (o - y) \cdot a_i$$

$$\boxed{\frac{\partial J}{\partial W^{[2]}} = (o - y) \cdot a^T}$$

$1 \times m$

$$\boxed{\frac{\partial J}{\partial b^{[2]}}} \qquad \frac{\partial J}{\partial b^{[2]}} = \frac{\partial J}{\partial o} \cdot \frac{\partial o}{\partial b^{[2]}} = (o - y) \in \mathbb{R}$$

②

$$\boxed{\frac{\partial J}{\partial W^{[1]}}} \qquad \frac{\partial J}{\partial W_{ij}^{[1]}} = \frac{\partial J}{\partial z_i} \frac{\partial z_i}{\partial W_{ij}^{[1]}} = \frac{\partial J}{\partial z_i} \cdot x_j$$

$m \times d$

$$z_i = \sum_{k=1}^{d} W_{ik}^{[1]} x_k + b_i^{[1]}$$

Vect.

$$\left[\frac{\partial J}{\partial W^{[1]}}\right] = \frac{\partial J}{\partial W_{ij}^{[1]}} \qquad \boxed{\underset{m \times d}{\frac{\partial J}{\partial W^{[1]}}} = \underset{m \times 1}{\frac{\partial J}{\partial z}} \cdot \underset{1 \times d}{x^T}}$$

$$\boxed{\frac{\partial J}{\partial z}} \qquad \frac{\partial J}{\partial z_i} = \frac{\partial J}{\partial a_i} \frac{\partial a_i}{\partial z_i} = \frac{\partial J}{\partial a_i} \cdot \mathrm{RELU}'(z_i)$$

$$= \frac{\partial J}{\partial a_i} 1\{z_i \geq 0\}$$

$$\boxed{\underset{m \times 1}{\frac{\partial J}{\partial z}} = \underset{m \times 1}{\frac{\partial J}{\partial a}} \odot \mathrm{RELU}'\{z\}}$$

$$o = \sum_{i=1}^{m} W_i^{2} a_i + b^{(2)}$$

$$\boxed{\frac{\partial J}{\partial a}} \qquad \frac{\partial J}{\partial a_i} = \frac{\partial J}{\partial o} \cdot \frac{\partial o}{\partial a_i} = (o - y) \cdot W_i^{(2)}$$

$$W^{(2)} \in \mathbb{R}^{1 \times m}$$

$$\boxed{\underset{m \times 1}{\frac{\partial J}{\partial a}} = W^{[2]^T} \cdot (o - y)}$$

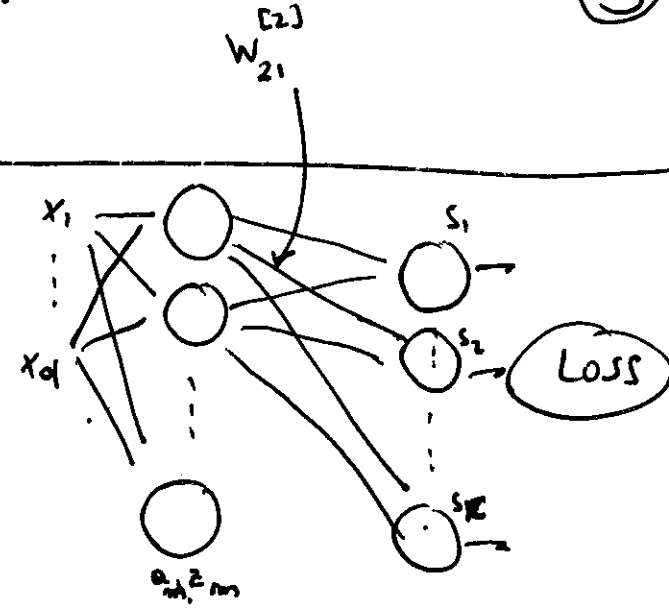
$$\frac{\partial J}{\partial b_i^{[1]}} = \frac{\partial J}{\partial z_i} \cdot \frac{\partial z_i}{\partial b_i^{[1]}} = \frac{\partial J}{\partial z_i} \rightarrow \frac{\partial J}{\partial b^{[1]}} = \frac{\partial J}{\partial z}$$

Show algorithm ③ page 17. ③

$$\underset{m\times1}{z} = \underset{m\times d}{W^{(1)}} \underset{d\times1}{x} + \underset{m\times1}{b^{(1)}}$$

$$a = \text{ReLU}(z)$$

$$\underset{c\times1}{s} = \underset{c\times m}{W^{(2)}} \underset{m\times1}{a} + \underset{c\times1}{b^{(2)}}$$



$$J = -\log\left(\frac{e^{s_y}}{\sum_{j=1}^{c} e^{s_j}}\right) = -\log(p_y)$$

p_k

$$= -\sum_{k=1}^{c} 1\{y=k\} \log\left(\frac{e^{s_k}}{\sum_{j=1}^{c} e^{s_j}}\right)$$

$$\boxed{S_i = \sum_{k=1}^{m} W^{(2)}_{ik} a_k + b^{(2)}_i}$$

$$\boxed{\underset{c\times m}{\frac{\partial J}{\partial W^{(2)}}}} = \frac{\partial J}{\partial W^{(2)}_{ij}} = \frac{\partial J}{\partial S_i} \cdot \frac{\partial S_i}{\partial W^{(2)}_{ij}}$$

$$= \frac{\partial J}{\partial s_i} \cdot a_j$$

$$\boxed{\text{Vect} = \underset{c\times1}{\frac{\partial J}{\partial s}} \cdot \underset{1\times m}{a^T}}$$

$$\frac{\partial J}{\partial s_k} = \begin{cases} p_k & k \neq y \\ p_k - 1 & k = y \end{cases}$$ check Jupyter notebook.

4

$$\boxed{\frac{\partial J}{\partial b^{(2)}}} \qquad \frac{\partial J}{\partial b_i^{(2)}} = \frac{\partial J}{\partial s_i} \cdot \frac{\partial s_i}{\partial b_i^{(2)}}$$

$$= \frac{\partial J}{\partial s_i}$$

$$\boxed{\frac{\partial J}{\partial b^{(2)}} = \frac{\partial J}{\partial s}}$$

$$S_i = \sum_{k=1}^{m} W_{ik}^{(2)} a_k + b_i^{(2)}$$

$$\boxed{\frac{\partial J}{\partial a}} = \qquad \frac{\partial J}{\partial a_i} = \sum_{e=1}^{c} \frac{\partial J}{\partial s_e} \cdot \frac{\partial s_e}{\partial a_i} =$$

$$= \sum_{e=1}^{c} \frac{\partial J}{\partial s_e} \cdot W_{ei}^{(2)}$$

i-th row of

$$= W_i^{(2)T} \frac{\partial J}{\partial s}$$

$1 \times c \qquad c \times 1$

$$\boxed{\underset{m\times 1}{\frac{\partial J}{\partial a}} = \underset{m \times c}{W^{(2)T}} \cdot \underset{c \times 1}{\frac{\partial J}{\partial s}}}$$

Alg 3

Compute $z \in \mathbb{R}^m$, $a \in \mathbb{R}^m$ $s \in \mathbb{R}^c$

$$\delta^{(2)} = \frac{\partial J}{\partial s} \in \mathbb{R}^c$$

$$\delta^{(1)} = \Big(\underbrace{W^{(2)T} \cdot \frac{\partial J}{\partial s}}_{\frac{\partial J}{\partial a}}\Big) \odot \text{relu}'(z) \quad \in \mathbb{R}^{m \times 1}$$

$$\begin{cases} \frac{\partial J}{\partial W^{(2)}} = \delta^{(2)} a^T \\ \frac{\partial J}{\partial b^{(2)}} = \delta^{(2)} \end{cases}$$

$$\frac{\partial J}{\partial W^{(1)}} = \delta^{(1)} x^T \qquad \frac{\partial J}{\partial b^{(1)}} = \delta^{(1)}$$

Vectorization over training examples

$$z^{[1](1)} = W^{[1]} x^{(1)} + b^{[1]}$$
$$z^{[1](2)} = W^{[1]} x^{(2)} + b^{[1]}$$
$$\vdots$$
$$z^{[1](m)} = W^{[1]} x^{(m)} + b^{[1]}$$

$$\boxed{\underset{n \times m}{Z = \begin{bmatrix} | & & | \\ z^{[1](1)} & \cdots & z^{[1](m)} \\ | & & | \end{bmatrix}} = \underset{n \times d}{W^{[1]}} \underset{d \times m}{X} + \underset{n \times 1}{b^{[1]}} \cdot \underset{1 \times m}{1_m^T}, \quad X = \begin{bmatrix} | & | & | \\ x^{(1)} & x^{(2)} & x^{(3)} \\ | & | & | \end{bmatrix}}$$

$$A \in \mathbb{R}^{n \times m}$$

$$S \in \mathbb{R}^{c \times m}$$

$$\boxed{\underset{c \times m}{S} = \underset{c \times n}{W^{[2]}} \cdot \underset{n \times m}{A} + \underset{c \times 1}{b^{[2]}} \underset{1 \times m}{1_m^T}}$$

$$\boxed{J = -\frac{1}{m} \sum_{i=1}^{m} \log\left(p_{y^{(i)}}\right) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{c} 1\{y^{(i)} = k\} \log\left(p_k\right)}$$

$$\frac{\partial J}{\partial S_{ij}} = \begin{cases} \frac{p_j}{m} & j \neq y^{(i)} \\ \frac{(p_j - 1)}{m} & j = y^{(i)} \end{cases}$$

$$\underset{m \times d}{\frac{\partial J}{\partial W^{[1]}}} = \underset{m \times m}{\frac{\partial J}{\partial Z}} \cdot \underset{m \times d}{X^T}$$

$$\underset{m \times m}{\frac{\partial J}{\partial A}} = \underset{m \times c}{W^{[2]T}} \cdot \underset{c \times m}{\frac{\partial J}{\partial S}}$$

$$\underset{c \times m}{\frac{\partial J}{\partial W^{[2]}}} = \underset{c \times m}{\frac{\partial J}{\partial S}} \cdot \underset{m \times m}{A^T}$$

$$\underset{c \times 1}{\frac{\partial J}{\partial b^{[2]}}} = \underset{c \times m}{\frac{\partial J}{\partial S}} \cdot \underset{m \times 1}{1_m}$$

$$\underset{m \times 1}{\frac{\partial J}{\partial b^{[1]}}} = \underset{m \times m}{\frac{\partial J}{\partial Z}} \cdot \underset{m \times 1}{1_m}$$

Alg 3

Compute $Z \in \mathbb{R}^{m \times n}$

1. $A \in \mathbb{R}^{m \times m}$

$S \in \mathbb{R}^{c \times m}$

2. compute

$$\delta^{[2]} = \frac{\partial J}{\partial S}$$

prime

$$\underset{m \times n}{\delta^{[1]}} = \boxed{\frac{\partial J}{\partial Z} = \underset{m \times m}{\frac{\partial J}{\partial A}} \odot \underset{m \times m}{\mathrm{RELU}'(Z)}}$$

$$\frac{\partial J}{\partial W^{[2]}} = \delta^{[2]} A^T$$

$$\frac{\partial J}{\partial b^{[2]}} = \delta^{[2]}$$

$$\frac{\partial J}{\partial W^{[1]}} = \delta^{[1]} X^T$$

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

IN practice

$$\underset{m\times d}{X} = \begin{bmatrix} - x^{(1)} - \\ - x^{(m)} - \end{bmatrix} \qquad \underset{m\times m}{Z} = \begin{bmatrix} - z^{[1](1)} - \\ - z^{[1](m)} \end{bmatrix}$$

7

$$\underset{m\times m}{Z} = \underset{m\times d}{X}\ \underset{d\times m}{W^{[1]}} + \underset{m\times 1}{1_m^T} \cdot \underset{1\times m}{b^{[1]}}$$

$$\underset{m\times c}{S} = \underset{m\times m}{A}\ \underset{m\times c}{W^{[2]}} + \underset{m\times 1}{1_m^T} \cdot \underset{1\times c}{b^{[2]}}$$

$$\underset{d\times m}{\frac{\partial J}{\partial W^{[1]}}} = \underset{d\times m}{X^T} \cdot \underset{m\times m}{\frac{\partial J}{\partial Z}}$$

$$\frac{\partial J}{\partial Z} = \frac{\partial J}{\partial A} \odot \mathrm{REW}'(Z)$$

$\downarrow$

$$\underset{m\times m}{\frac{\partial J}{\partial A}} = \underset{m\times c}{\frac{\partial J}{\partial S}} \cdot \underset{c\times m}{W^{[2]T}}$$

$$\underset{1\times c}{\frac{\partial J}{\partial b^{[2]}}} = \underset{1\times m}{1_m^T} \cdot \underset{m\times c}{\frac{\partial J}{\partial S}}$$

$$\underset{m\times c}{\frac{\partial J}{\partial W^{[2]}}} = \underset{m\times m}{A^T} \cdot \underset{m\times c}{\frac{\partial J}{\partial S}}$$

$$\underset{1\times m}{\frac{\partial J}{\partial b^{[1]}}} = \underset{1\times m}{1_m} \underset{m\times m}{\frac{\partial J}{\partial Z}}$$