Assignment Problem set 2

1) → suppose $a^{[1]}$ is a 128-vector. let $a_B^{[1]}$

$a_B^{[1]} = \begin{bmatrix} a^{[1]} \\ 1 \end{bmatrix}$. calculate the gradient $\nabla_{a^{[1]}}(a_B^{[1]})$

$$\nabla_{\underset{\text{input}}{a^{[1]}}} \overset{\text{output}}{(a_B^{[1]})}$$

Shape of gradient : input shape x output shape

$= (128 \times 1) \times (129 \times 1)$

$= (128 \times 129)$

$$\begin{array}{cc} & \leftarrow 129 \rightarrow \\ 128 & \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & \cdot \\ 0 & 0 & 1 & \cdots & \cdot \\ & & & & \\ & & \cdots & 1 & 0 \end{bmatrix} \end{array} \leftarrow \text{gradient } \nabla_{a^{[1]}}(a_B^{[1]})$$

2. Suppose z is a 128-vector. If $a = \text{ReLU}(z)$, calculate the gradient $\nabla_z(a)$.

$$\nabla_{\underset{\text{input}}{z}} \overset{\text{output}}{(a)}$$

shape of gradient:-

$=$ input shape x output shape

$= (128 \times 1) \times (128 \times 1)$

$= 128 \times 128$.

The values of gradient will be 1 or 0 depending on whether values of z are +ve or −ve

$z_1 \to +ve$ $z_{128} \leftarrow +ve$ valu

$z_2 \to -ve$

Eg. $z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{128} \end{bmatrix}$ $a = \begin{bmatrix} z_1 \\ 0 \\ \vdots \\ z_{128} \end{bmatrix}$

$$\nabla_a z = \begin{array}{c c} & \xleftarrow{\quad 128 \quad} \\ 128 & \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ & & \ddots & \\ 0 & \cdots & & 1 \end{bmatrix} \end{array}$$

3) consider the following forward propagation for a sample 784-vector x with one hot encoded correct label 10-vector y (shapes shown below).

$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{784} \end{bmatrix}$ $y = \begin{array}{c c} & \leftarrow 10 \rightarrow \\ 10 & \begin{bmatrix} \\ \\ \\ \end{bmatrix} \end{array}$ $\hat{y} =$

$\hat{y} = a^{[2]}$, predicted probability vector. What are missing shapes above?

$\to \underbrace{x_B}_{785} = \begin{bmatrix} x \\ 1 \end{bmatrix} \to \underset{128}{z^{[1]}} = \underset{128 \times 785}{w^{[1]}} \; \underset{785}{x_B} \to$

$\underbrace{a^{[1]}}_{128} = ReLU\left(\underbrace{z^{[1]}}_{128}\right) \to \underbrace{a_B^{[1]}}_{129} = \underset{129}{\begin{bmatrix} a^{[1]} \\ 1 \end{bmatrix}}$

$\rightarrow \underbrace{z^{[2]}}_{10} = \underset{10 \times 129}{W^{[2]}} \; \underset{129 \times 1}{a_B^{[1]}} \rightarrow \underbrace{a^{[2]}}_{10} = \text{softmax}(\underbrace{z^{[2]}}_{10})$

$\rightarrow L = \sum_{k=0}^{9} -y_k \log(\hat{y}_k)$

4) Fill in the missing entries

$$\nabla_{a^{[1]}} L = \nabla_{a^{[1]}}(a_B^{[1]}) \times \nabla_{a_B^{[1]}}(z^{[2]}) \times \nabla_{z^{[2]}}(L)$$

$$\nabla_{z^{[1]}} L = \nabla_{z^{[1]}}(a^{[1]}) \times \nabla_{a^{[1]}}(L)$$

5) calculate the gradients $\nabla_{a_B^{[1]}}(z^{[2]})$ and $\nabla_{z^{[1]}}(a^{[1]})$

output shape

$\rightarrow \nabla_{a_B^{[1]}}(z^{[2]})$ $\Rightarrow$ shape of gradient

input shape

= input shape x output shape

= 129 x 10

$$\therefore z^{[2]} = W^{[2]} a_B^{[1]} =$$

$$\underset{10}{\uparrow} \qquad \underset{10 \times 129}{\uparrow} \qquad \underset{129 \times 1}{\uparrow}$$

let

$$z^{[2]} = \begin{array}{c} \uparrow \\ 10 \\ \downarrow \end{array} \begin{bmatrix} w_{1,1} a_1 + w_{12} a_2 + \cdots + w_{1,129} a_{129} \\ \vdots \\ w_{10,1} a_1 + \cdots + w_{10,129} a_{129} \end{bmatrix}$$

$$\nabla_{a_B^{[1]}} \left(z^{[2]}\right) = \begin{bmatrix} w_{11} & 0 & 0 & \cdots & 0 \\ 0 & w_{22} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & & w_{10,129} \end{bmatrix}$$

$\downarrow$

$(a_1 \;\; a_2 \;\; a_3 \cdots a_{129})$

output shape

$\rightarrow \nabla_{z^{[1]}} \left(a^{[1]}\right) \Rightarrow$ shape of gradient =

$\uparrow$ input shape

= input shape $\times$ output shape

$= (128 \times 1) \times (128 \times 1)$

$= (128 \times 128)$

in ReLU if $x > 0 \rightarrow x$ or else 0.

$\downarrow$

so it is same answer as question (2).