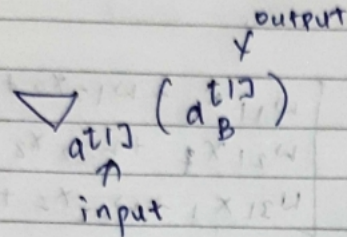


Assignment Problem set 2

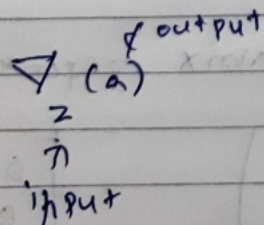
1) → suppose $a^{[1]}$ is a 128-vector. let $d_8^{[1]}$
 $a_B^{[1]} = \begin{bmatrix} a^{[1]} \\ 1 \end{bmatrix}$ calculate the gradient $\nabla_{a^{[1]}}(a_B^{[1]})$



Shape of gradient = input shape \times output shape
 $= (128 \times 1) \times (129 \times 1)$
 $= (128 \times 129)$

$$\begin{matrix} & \leftarrow 129 \rightarrow \\ \begin{matrix} \uparrow \\ 128 \\ \downarrow \end{matrix} & \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} & \leftarrow \text{gradient} \\ & \nabla_{a^{[1]}}(a_B^{[1]}) \end{matrix}$$

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



Shape of gradient:
 $= \text{input shape} \times \text{output shape}$
 $= (128 \times 1) \times (128 \times 1)$
 $= 128 \times 128$

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

$z_1 \rightarrow +ve$
 $z_2 \rightarrow -ve$
 $z_{128} \leftarrow +ve \text{ value}$

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

$\nabla_z a = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 1 \end{bmatrix}$
 (Dimensions: 128 rows, 128 columns)

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

$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{784} \end{bmatrix}$ $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{10} \end{bmatrix}$ $\hat{y} = x$

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

$\rightarrow \underbrace{x_B}_{785} = \begin{bmatrix} x \\ 1 \end{bmatrix} \rightarrow z^{[1]}_{128} = W^{[1]}_{128 \times 785} x_B \rightarrow$

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

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

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

4) Fill in the missing entries

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

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

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

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

✓ output shape

$$\rightarrow \nabla_{a_B^{[1]}}(z^{[2]}) \Rightarrow \text{shape of gradient} \\ = \text{input shape} \times \text{output shape} \\ = 129 \times 10$$

input shape

$$z^{[2]} = w^{[2]} a_B^{[1]} =$$

\uparrow \uparrow \uparrow
 10 10x129 129x1

$$\text{Let } \underbrace{z^{[2]}}_{10} = \left[\begin{array}{c} w_{11} a_{\phi 1} + w_{12} a_{\phi 2} + \dots + w_{1,129} a_{\phi 129} \\ \vdots \\ w_{10,1} a_{\phi 1} + \dots + w_{10,129} a_{\phi 129} \end{array} \right]$$

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

$(a_1 \ a_2 \ a_3 \dots a_{129})$

↖ output shape

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

↑ input shape

$$= \text{input shape} \times \text{output shape}$$

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

$$= (128 \times 128)$$

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

↓
so it is same answer as question (2).