Problem Set - 2

1) $a_1^{[1]} \rightarrow$ 128 vector

$$a_B^{[1]} = \begin{bmatrix} a^{[1]} \\ 1 \end{bmatrix}$$

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

$a_1^{[1]} \rightarrow$ Activated values of activation layer 1 (hidden layer 1)

$a_B^{[1]} \rightarrow$ input to dense layer 2, with bias included (129-vector)

L

↓

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

↓

$z^{[2]}$

↓

$a^{[1]}$

↓ (Sigmoid / Relu / Tanh)

$z^{[1]}$

↓

W

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

output of layer 1 (activation) is input to layer 2 with (dense) bias added. They are the same.

2) $z \rightarrow$ 128 vector, $a = ReLU(z)$, $\nabla_z(a)$

$$\nabla_z(a) = 1. * (z > 0)$$

3) $\underbrace{x_B}_{785} = \begin{bmatrix} x \\ 1 \end{bmatrix} \rightarrow \underbrace{z^{[1]}}_{128} = \underbrace{W^{[1]}}_{128 \times 785} \underbrace{x_B^T}_{785} \rightarrow \underbrace{a^{[1]}}_{128} = ReLU\left(\underbrace{z^{[1]}}_{128}\right)$

$L = \sum_{k=0}^{9} -y_k \log(\hat{y}_k) \leftarrow \underbrace{a^{[2]}}_{10} = softmax\left(\underbrace{z^{[2]}}_{10}\right) \leftarrow \underbrace{z^{[2]}}_{10} = \underbrace{W^{[2]}}_{10 \times 129} \underbrace{a_B^{[1]}}_{129} \leftarrow \underbrace{a_B^{[1]}}_{129} = \begin{bmatrix} a^{[1]} \\ 1 \end{bmatrix}$ ↓

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

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

5) $\nabla_{a_B^{[1]}}(z^{[2]}) = W^{[2]} \times \nabla_{a_B^{[1]}}(W a_B^{[1]}) =$ Same vector $\vee \cdot a_B^{[1]T}$