AML5202 | Deep Learning | Problem Set-2

1. Suppose $\mathbf{a}^{[1]}$ is a 128-vector. Let $\mathbf{a}_B^{[1]} = \begin{bmatrix} \mathbf{a}^{[1]} \\ 1 \end{bmatrix}$. Calculate the gradient $\nabla_{\mathbf{a}^{[1]}} \left(\mathbf{a}_B^{[1]} \right)$.

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

3. Consider the following forward propagation for a sample 784-vector $\mathbf{x}$ with one-hot encoded correct label 10-vector $\mathbf{y}$ (shapes shown below):

$$\boxed{\underbrace{\mathbf{x}_B}_{785} = \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}} \rightarrow \boxed{\underbrace{\mathbf{z}^{[1]}}_{128} = \underbrace{\mathbf{W}^{[1]}}_{?\times ?} \underbrace{\mathbf{x}_B}_{785}} \rightarrow \boxed{\underbrace{\mathbf{a}^{[1]}}_{?} = \text{ReLU}\left(\underbrace{\mathbf{z}^{[1]}}_{128} \right)} \rightarrow \boxed{\underbrace{\mathbf{a}_B^{[1]}}_{?} = \begin{bmatrix} \mathbf{a}^{[1]} \\ 1 \end{bmatrix}} \rightarrow \boxed{\underbrace{\mathbf{z}^{[2]}}_{10} = \underbrace{\mathbf{W}^{[2]}}_{?\times ?} \underbrace{\mathbf{a}_B^{[1]}}_{?}} \rightarrow \boxed{\underbrace{\mathbf{a}^{[2]}}_{10} = \text{softmax}\left(\underbrace{\mathbf{z}^{[2]}}_{10} \right)} \rightarrow \boxed{L = \sum_{k=0}^{9} -y_k \log\left(\hat{y}_k\right)},$$

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

4. Fill in the missing entries in $\begin{cases} \nabla_{\mathbf{a}^{[1]}}(L) & = \nabla_{?}\left(\mathbf{a}_B^{[1]} \right) \times \nabla_{\mathbf{a}_B^{[1]}}\left(\mathbf{z}^{[2]} \right) \times \nabla_{?}(L), \\ \nabla_{?}(L) & = \nabla_{\mathbf{z}^{[1]}}\left(\mathbf{a}^{[1]} \right) \times \nabla_{\mathbf{a}^{[1]}}(L). \end{cases}$

5. Calculate the gradients $\nabla_{\mathbf{a}_B^{[1]}}\left(\mathbf{z}^{[2]} \right)$ and $\nabla_{\mathbf{z}^{[1]}}\left(\mathbf{a}^{[1]} \right)$.