Quiz 3: Neural Network Forward Propagation

Instructions

Select the best answer or provide the calculated result. Questions 1-4 cover core concepts, while question 5 involves calculation and broadcasting.

Questions

1. Dimensions Check

Consider a neural network with an input layer of size $n_x = 10$, a single hidden layer of size $n_h = 20$, and an output layer of size $n_y = 5$. What are the dimensions of the weight matrix $W^{[1]}$ connecting the input layer to the hidden layer? A) $(10, 20)$ B) $(20, 10)$ C) $(10, 5)$ D) $(20, 5)$

2. Scalar Calculation (Hidden Layer)

Given an input $\mathbf{x} = [\begin{matrix} 2 \\ -1 \end{matrix}]$, weights $W^{[1]} = [\begin{matrix} 1 & 0.5 \\ -2 & 1 \end{matrix}]$, and biases $b^{[1]} = [\begin{matrix} 0.1 \\ -0.2 \end{matrix}]$, calculate the pre-activation vector $\mathbf{z}^{[1]}$ for the hidden layer. A) $[\begin{matrix} 1.6 \\ -5.2 \end{matrix}]$ B) $[\begin{matrix} 2.1 \\ -4.7 \end{matrix}]$ C) $[\begin{matrix} 1.1 \\ -3.2 \end{matrix}]$ D) $[\begin{matrix} 2.5 \\ -5.0 \end{matrix}]$

3. Vector Form Equation

Which equation correctly represents the *activation* $\mathbf{a}^{[1]}$ of the hidden layer for a single input example $\mathbf{x}$, using activation function $g^{[1]}$? A) $\mathbf{a}^{[1]} = g^{[1]}(\mathbf{W}^{[1]}\mathbf{x})$ B) $\mathbf{a}^{[1]} = \mathbf{W}^{[1]}g^{[1]}(\mathbf{x}) + \mathbf{b}^{[1]}$ C) $\mathbf{a}^{[1]} = g^{[1]}(\mathbf{W}^{[1]}\mathbf{x} + \mathbf{b}^{[1]})$ D) $\mathbf{a}^{[1]} = g^{[1]}(\mathbf{x}\mathbf{W}^{[1]} + \mathbf{b}^{[1]})$

4. Batch Processing Dimensions

If you process a batch of $m = 32$ examples through the network described in Question 1 ($n_x = 10, n_h = 20, n_y = 5$), what will be the dimensions of the final output activation matrix $A^{[2]}$? A) $(5, 20)$ B) $(32, 5)$ C) $(5, 32)$ D) $(20, 32)$

5. Broadcasting Calculation (Hard)

Let $W^{[1]} = [\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}]$, input batch $X = [\begin{matrix} 2 & 3 \\ 4 & 5 \end{matrix}]$, and bias $b^{[1]} = [\begin{matrix} 10 \\ 20 \end{matrix}]$. Calculate the matrix $Z^{[1]} = W^{[1]}X + b^{[1]}$ after performing the matrix multiplication and broadcasting the bias. A) $[\begin{matrix} 12 & 13 \\ 24 & 25 \end{matrix}]$ B) $[\begin{matrix} 2 & 3 \\ 4 & 5 \end{matrix}]$ C) $[\begin{matrix} 12 & 23 \\ 14 & 25 \end{matrix}]$ D) $[\begin{matrix} 20 & 30 \\ 40 & 50 \end{matrix}]$

6. Two Hidden Layers Dimensions

Consider a neural network with architecture: Input ($n_x = 50$) -> Hidden 1 ($n_{h1} = 100$) -> Hidden 2 ($n_{h2} = 60$) -> Output ($n_y = 10$). If you process a batch of $m = 128$ examples, what are the dimensions of the activation matrix $A^{[2]}$ (the output of the second hidden layer)? A) $(100, 128)$ B) $(60, 100)$ C) $(60, 128)$ D) $(128, 60)$

Answers

1. B
2. A
3. C
4. C
5. A
6. C

Explanations

1. **B)** $(20, 10)$: The dimensions of a weight matrix $W^{[l]}$ connecting layer $l-1$ to layer l are $(n_{\text{neurons in layer } l} \times n_{\text{neurons in layer } l-1})$. Here, $l = 1$, layer l is the hidden layer ($n_h = 20$), and layer $l-1$ is the input layer ($n_x = 10$).
2. **A)** $\begin{bmatrix} 1.6 \\ -5.2 \end{bmatrix}$: Calculate $\mathbf{z}^{[1]} = \mathbf{W}^{[1]}\mathbf{x} + \mathbf{b}^{[1]}$. $z_1 = (1 \times 2 + 0.5 \times -1) + 0.1 = (2 - 0.5) + 0.1 = 1.5 + 0.1 = 1.6$. $z_2 = (-2 \times 2 + 1 \times -1) - 0.2 = (-4 - 1) - 0.2 = -5 - 0.2 = -5.2$.
3. **C)** $\mathbf{a}^{[1]} = g^{[1]}(\mathbf{W}^{[1]}\mathbf{x} + \mathbf{b}^{[1]})$: The activation is computed by first calculating the linear pre-activation part ($\mathbf{W}^{[1]}\mathbf{x} + \mathbf{b}^{[1]}$) and then applying the non-linear activation function $g^{[1]}$ to the result.
4. **C)** $(5, 32)$: The output activation matrix $A^{[2]}$ will have dimensions $(n_{\text{neurons in output layer}} \times m)$, which is $(n_y \times m) = (5 \times 32)$.
5. **A)** $\begin{bmatrix} 12 & 13 \\ 24 & 25 \end{bmatrix}$: First, compute $W^{[1]}X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$. Then, broadcast the bias $b^{[1]} = \begin{bmatrix} 10 \\ 20 \end{bmatrix}$ across the columns to match the shape of $W^{[1]}X$, resulting in $\begin{bmatrix} 10 & 10 \\ 20 & 20 \end{bmatrix}$. Finally, add them: $Z^{[1]} = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} + \begin{bmatrix} 10 & 10 \\ 20 & 20 \end{bmatrix} = \begin{bmatrix} 12 & 13 \\ 24 & 25 \end{bmatrix}$.
6. **C)** $(60, 128)$: For batch processing, the activation matrix $A^{[l]}$ of layer l has dimensions $(n_{\text{neurons in layer } l} \times m)$. The second hidden layer has $n_{h2} = 60$ neurons, and the batch size is $m = 128$. Therefore, the dimensions of $A^{[2]}$ are $(60, 128)$.