

Input:  $5 \times 4$

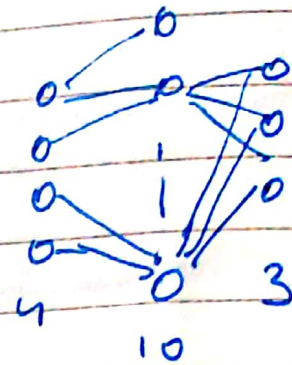
$W1: 4 \times 10$

$W2: 10 \times 3$

scores:  $5 \times 3$

$b1: 10$

$b2: 3$



$$Z = X \cdot W1 + b1$$

$$Z = \text{Relu}(Z)$$

$$\text{scores} = Z \cdot W2 + b2$$

$$L = \frac{e^{\text{scores}_i y_i}}{\sum_j e^{\text{scores}_j a_j}}$$

### Derivatives

$$\frac{\partial L}{\partial W2} = \text{same as previous pdf}$$

$$\frac{\partial L}{\partial b2} = \left[ \frac{e^{a_n}}{e^{a_1} + e^{a_2} + \dots} \right] \cdot 1 \quad \left[ \frac{\partial \text{scores}}{\partial b2} = 1 \right]$$

(Sum over scores)

scores =  $5 \times 3$      $b2 = 3$

[axis = 1]

$$\frac{\partial L}{\partial W1} = \frac{\partial L}{\partial \text{scores}} \cdot \frac{\partial \text{scores}}{\partial Z} \cdot \frac{\partial \text{relu}_Z}{\partial Z} \cdot \frac{\partial Z}{\partial W1}$$

common even for  $b1$

$$\frac{\partial L}{\partial b1} = \frac{\partial L}{\partial \text{scores}} \cdot \frac{\partial \text{scores}}{\partial Z} \cdot \frac{\partial \text{relu}_Z}{\partial Z} \cdot \frac{\partial Z}{\partial b1}$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \text{scores}} \cdot \frac{\partial \text{scores}}{\partial w_2} = \frac{(\text{scores} - y_i)}{(\text{scores} - y_i) \cdot Z}$$

$y_i$  represents a matrix of size and shape scores. shape, with 1 at positions of the correct class.

$$\frac{\partial L}{\partial w_1} = (\text{scores} - y_i) \cdot w_2 * (Z \neq 0) \times X$$

$$\frac{\partial L}{\partial b_1} = (\text{scores} - y_i) \cdot w_2 * (Z \neq 0) \times 1$$

(sum over axis=1)