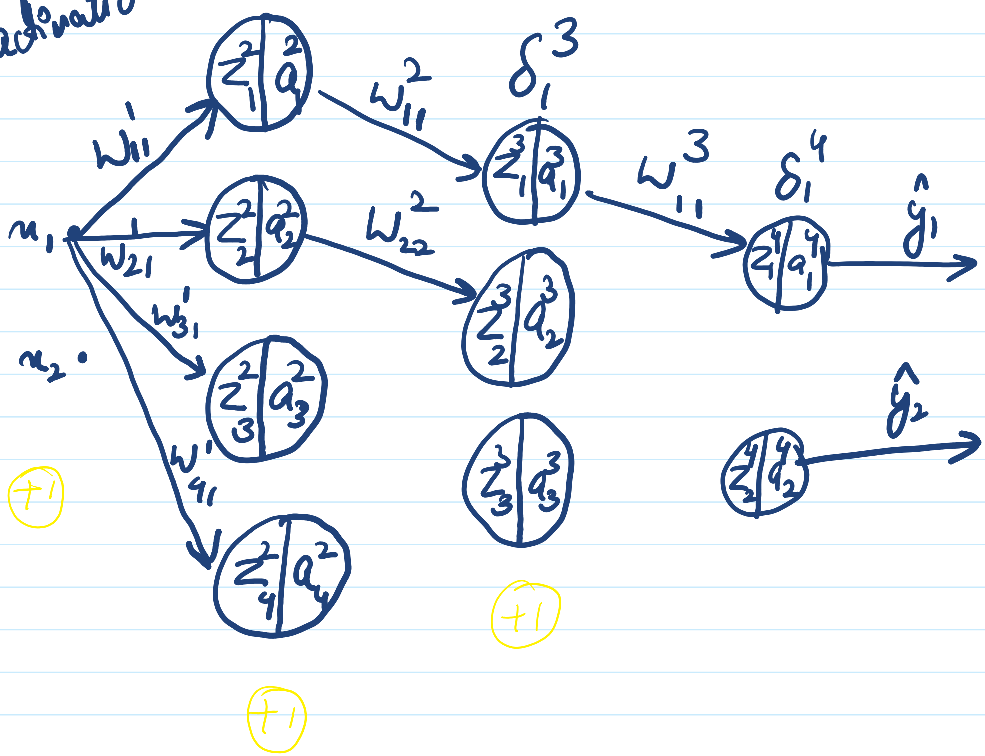
11_Backpropagation_in_neural_networks

23 August 2024 10:40

w_{ij}^{l}, b_i^{l} $\quad z_i^{l} / a_i^{l}$ $\quad \delta_i^{l} \rightarrow$ error

before activation $\quad$ after activation



Input Layer $\quad$ $l=1$ $\quad$ $l=2$ $\quad$ $l=3$ $\quad$ $l=4$

$$L(\theta) = \frac{1}{2} \sum_{i=1}^{2} (\hat{y}_i - y_i)^2$$

$L(\theta) \rightarrow w, b$

$y_i \rightarrow$ GT value

2 1 $\quad$ 1

$$z_i^2 = \sum_{j=1} w_{ij} x_j + b_i' \ , \quad i = 1, 2, 3, 4$$

$$a_i^2 = f(z_i^2) \ , \quad i = 1, 2, 3, 4.$$

$$z_i^3 = \sum_{j=1}^{4} w_{ij}^2 \, a_j^2 + b_i^2 \ , \quad i = 1, 2, 3$$

$$a_i^3 = f(z_i^3)$$

$$z_i^4 = \sum_{j=1}^{3} w_{ij}^3 \cdot a_j^3 + b_i^3 \ , \quad i = 1, 2$$

$$a_i^4 = f(z_i^4) = \hat{y}_i \ , \quad i = 1, 2$$

→ Randomly initialized weights and biases.

→ One forward propagation

→ Calculate error terms and propagate the error backwards.

"Chain Rule"

$$\frac{\partial L}{\partial w_{ij}^3} = \underbrace{\frac{\partial L}{\partial z_i^4}}_{\delta_i^4} \cdot \underbrace{\frac{\partial z_i^4}{\partial w_{ij}^3}}_{a_j^3}$$

$$\frac{\partial L}{\partial w_{ij}^2} = \underbrace{\frac{\partial L}{\partial z_i^3}}_{\delta_i^3} \cdot \underbrace{\frac{\partial z_i^3}{\partial w_{ij}^2}}_{a_j^2}$$

$$\frac{\partial L}{\partial w_{ij}^1} = \underbrace{\frac{\partial L}{\partial z_i^2}}_{\delta_i^2} \cdot \underbrace{\frac{\partial z_i^2}{\partial w_{ij}^1}}_{x_j}$$

Calculate δ_i^4 first and using that solve for δ_i^3, δ_i^2, δ_i^1 in a recursive manner.

"Delta Learning Rule"

$$\delta_i^4 = \frac{\partial L}{\partial z_i^4} \qquad \left\{ L = \frac{1}{2}\sum_{i=1}^{2}(\hat{y}_i - y_i)^2 \right.$$

$$= (\hat{y}_i - y_i) \cdot \frac{\partial \hat{y}_i}{\partial z_i^4} \quad f'(z_i^4)$$

$$\delta_i^3 = \frac{\partial L}{\partial z_i^3} = \frac{\partial L}{\partial a_i^3} \cdot \frac{\partial a_i^3}{\partial z_i^3} \quad f'(z_i^3)$$

$$\frac{\partial L}{\partial a_i^2} = \sum_{j=1}^{2} \underbrace{\frac{\partial L}{\partial z_j^4}}_{\delta_j^4} \cdot \underbrace{\frac{\partial z_j^4}{\partial a_i^3}}_{w_{ji}^3}$$

$$\Rightarrow \delta_i^3 = \left(\sum_{j=1}^{2} \delta_j^4 \cdot w_{ji}^3\right) \cdot f'(z_i^3)$$

$$\Rightarrow \delta_i^2 = \left(\sum \delta_j^3 \cdot w_{ji}^2\right) \cdot f'(z_i^2)$$

In general $\delta_i^l = \sum_{j=1}^{s_{l+1}}$

$s_{l+1} \rightarrow$ # of neurons in $(l+1)^{th}$ layer

$$\delta_i^l = \left(\sum_{j=1}^{s_{l+1}} \delta_j^{l+1} \cdot w_{ji}^l\right) \cdot f'(z_i^l)$$