$$\underset{K_1\times 1}{\vec{z}_1} = \underset{K_1\times N}{W_1}\,\underset{N\times 1}{\vec{x}} + \underset{K_1\times 1}{\vec{b}_1}$$

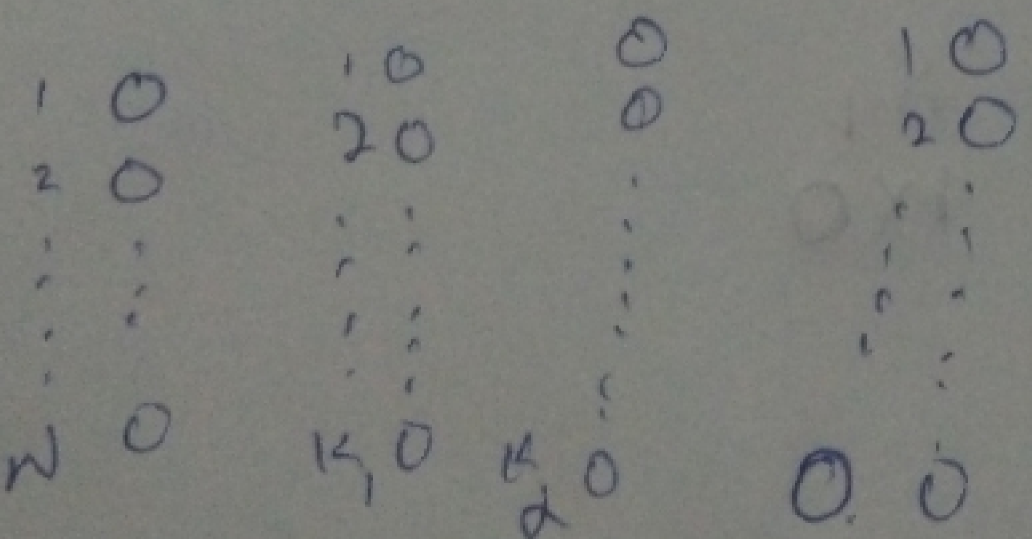
$$\underset{K_1\times 1}{\vec{h}_1} = \sigma(\vec{z}_1)$$

$$\underset{K_2\times 1}{\vec{z}_2} = \underset{K_2\times K_1}{W_2}\,\underset{K_1\times 1}{\vec{h}_1} + \underset{K_2\times 1}{\vec{b}_2}$$

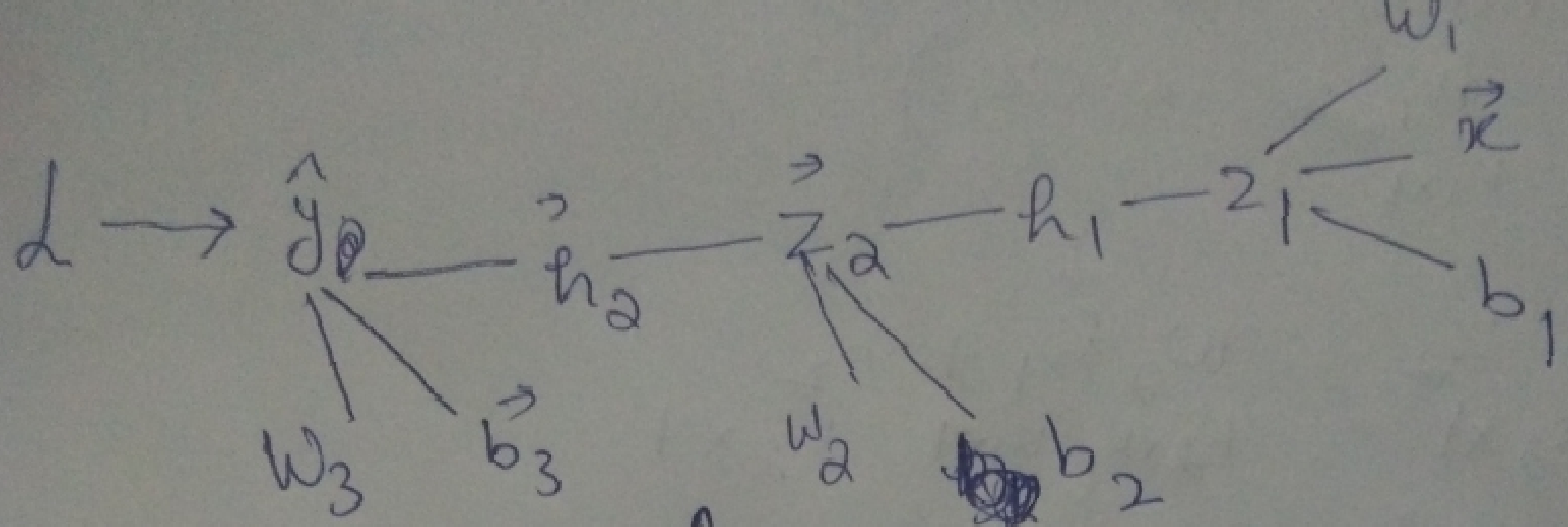
$$\underset{K_2\times 1}{\vec{h}_2} = \sigma(\vec{z}_2)$$

$$\underset{O\times 1}{\hat{y}} = \underset{O\times K_2}{W_3}\,\underset{K_2\times 1}{\vec{h}_2} + \underset{O\times 1}{b_3}$$

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



Computational chart



$$\frac{\partial \mathcal{L}}{\partial \vec{b}_3} = \frac{\partial \mathcal{L}}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial \vec{b}_3}$$

$$\frac{\partial \mathcal{L}}{\partial w_3} = \frac{\partial \mathcal{L}}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w_3}$$

$$\frac{\partial \mathcal{L}}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}} \left(\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \right) = \frac{1}{N} \sum_{i=1}^{N} (-2)(y_i - \hat{y}_i)$$

$$\frac{1}{N} \sum_{i=1}^{N} (-2)(y_i - \hat{y}_i) = \underset{1 \times O}{\delta_1}$$

$$\frac{\partial \hat{y}}{\partial \vec{b}_3} = I_O$$

$$\frac{\partial \mathcal{L}}{\partial \vec{b}_3} = \underset{O \times 1}{\delta_1^T}$$ $\because$ because $\vec{b}_3$ is column vector.

$$\frac{\partial \mathcal{L}}{\partial W_3} = \delta_1^T \vec{h}_2^T$$ Using identity (5) in Notes.

$$\underset{O\times K_2}{} \quad \underset{O\times 1}{} \; \underset{1\times K_2}{}$$

$$\frac{\partial \mathcal{L}}{\partial \vec{b}_2} = \frac{\partial \mathcal{L}}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial \vec{h}_2} \times \frac{\partial \vec{h}_2}{\partial \vec{z}_2} \times \frac{\partial \vec{z}_2}{\partial \vec{b}_2}$$

$$\frac{\partial \hat{y}}{\partial \vec{h}_2} = \frac{\partial}{\partial \vec{h}_2}(W_3 \vec{h}_2 + b_3) = \underset{O\times K_2}{W_3}$$

$$\frac{\partial \vec{h}_2}{\partial \vec{z}_2} = \underset{(K_2)}{\sigma'(\vec{z}_2)} \quad \Rightarrow \sigma'(\vec{z}_2) = \sigma(\vec{z}_2)(1-\sigma(\vec{z}_2))$$

$$\frac{\partial \vec{z}_2}{\partial b_2} = \underset{K_2}{I}$$

$$\cancel{\frac{\partial \mathcal{L}}{\partial b_2} \; \underset{1\times O}{\delta_1} \; \underset{O\times K_2}{W_3} \; \underset{K_2}{\sigma'(\vec{z}_2)} = \delta}$$

$$\underset{1\times O}{\delta_1} \; \underset{O\times K_2}{W_3} \; \underset{K_2}{\sigma'(\vec{z}_2)} = \underset{1\times K_2}{\delta_2}$$

$$\frac{\partial \mathcal{L}}{\partial \vec{b}_2} = (\delta_2 \times I)^T = \underset{K_2\times 1}{\delta_2^T}$$ $\because$ because $\vec{b}_2$ is column vector.

(4)

$$\frac{\partial L}{\partial w_2} = \underbrace{\frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial \vec{h}_2} \times \frac{\partial \vec{h}_2}{\partial \vec{z}_2}}_{\delta_2} \times \frac{\partial z_2}{\partial w_2}$$

$$\underset{K_2 \times K_1}{\frac{\partial L}{\partial w_2}} = \underset{K_2 \times 1}{\delta_2^T} \; \underset{1 \times K_1}{\vec{h}_1^T}$$

Using Identity (5) in Notes.

$$\frac{\partial L}{\vec{b}_3} = \underbrace{\underbrace{\frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial \vec{h}_2} \times \frac{\partial \vec{h}_2}{\partial \vec{z}_2}}_{\delta_2} \times \frac{\partial \vec{z}_2}{\partial \vec{h}_1} \times \frac{\partial h_1}{\partial z_1}}_{\delta_3} \times \frac{\partial z_1}{\partial b_1}$$

$$\frac{\partial \vec{z}_2}{\partial \vec{h}_1} = \underset{K_2 \times K_1}{w_2}$$

$$\frac{\partial \vec{h}_1}{\partial \vec{z}_1} = \underset{K_1}{\sigma'(z_1)} = \sigma(z_1)(1-\sigma(z_1))$$

$$\underset{1 \times K_1}{\delta_3} = \underset{1 \times K_2}{\delta_2} \; \underset{K_2 \times K_1}{w_2} \; \underset{K_1}{\sigma'(z_1)}$$

$$\frac{\partial \vec{z}_1}{\partial \vec{b}_1} = \underset{K_1}{I}$$

$$\frac{\partial \mathcal{L}}{\partial \vec{b}_1} = (\delta_3 I)^T = \underset{K_1 \times 1}{\delta_3^T} \quad \therefore b_1 \text{ is column Matrix}$$

$$\frac{\partial \mathcal{L}}{\partial w_3} = \underbrace{\frac{\partial \mathcal{L}}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial \vec{h}_2} \times \frac{\partial \vec{h}_2}{\partial \vec{z}_2} \times \frac{\partial \vec{z}_2}{\partial \vec{h}_1} \times \frac{\partial h_1}{\partial z_1}}_{\delta_3} \times \frac{\partial z_1}{\partial w_1}$$

$$\underset{K_1 \times N}{\frac{\partial \mathcal{L}}{\partial w_3}} = \underset{K_1 \times 1}{\delta_3^T} \; \underset{1 \times N}{x^T}$$