Backpropagation

using different loss Function

$$\delta^L = (a^L - y)$$

$$\boxed{\delta^L = \underbrace{(a^L - y) \odot \sigma'(z^l)}}$$ ×

$$\delta^l = \left(w^{l+1} \delta^{l+1}\right) \odot \sigma'(z^l)$$

Same

$$\frac{\partial L}{\partial b^l} = \delta^l$$

$$\frac{\partial L}{\partial w^l} = a^{l-1} (\delta^l)^T$$

Assumption

MSE loss $= \frac{1}{2} \sum_i (y_i - a_i)^2$

$$\delta^L = \boxed{\frac{\partial L}{\partial z^L} = (a^L - y) \cdot \sigma'(z^l)}$$

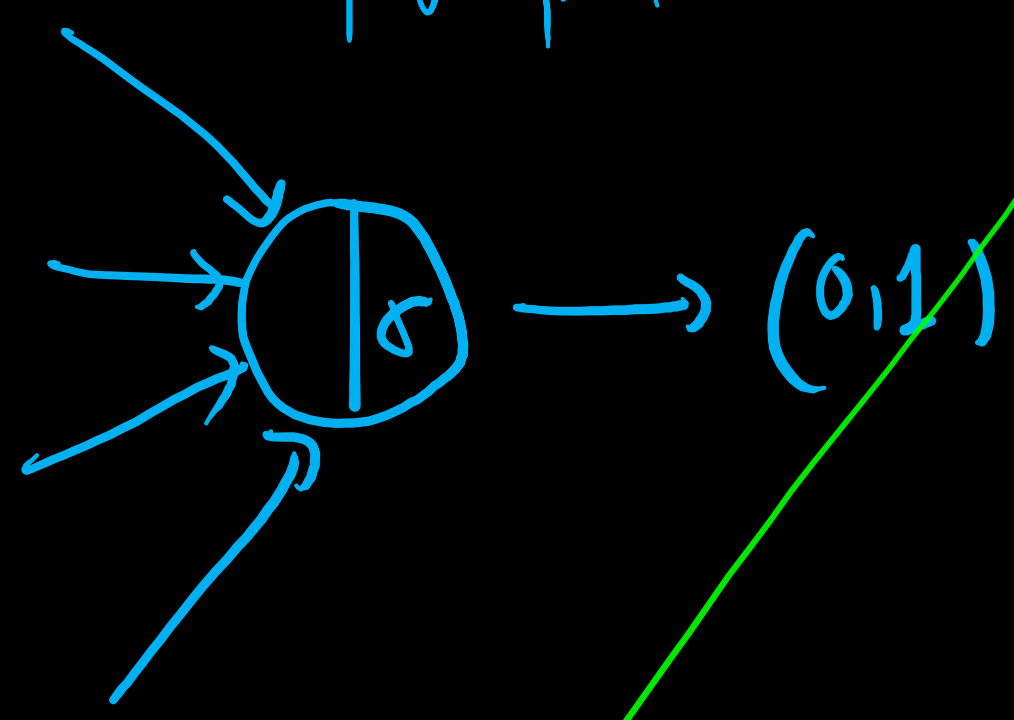
vectorization for one example

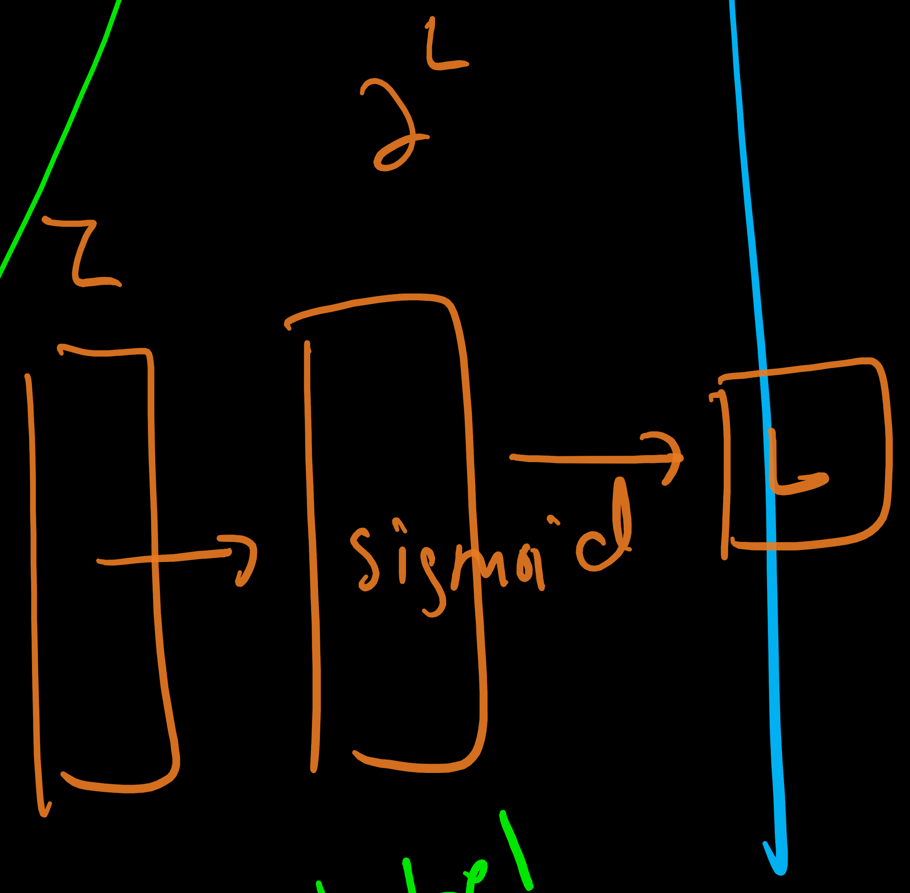
Categorical cross entropy

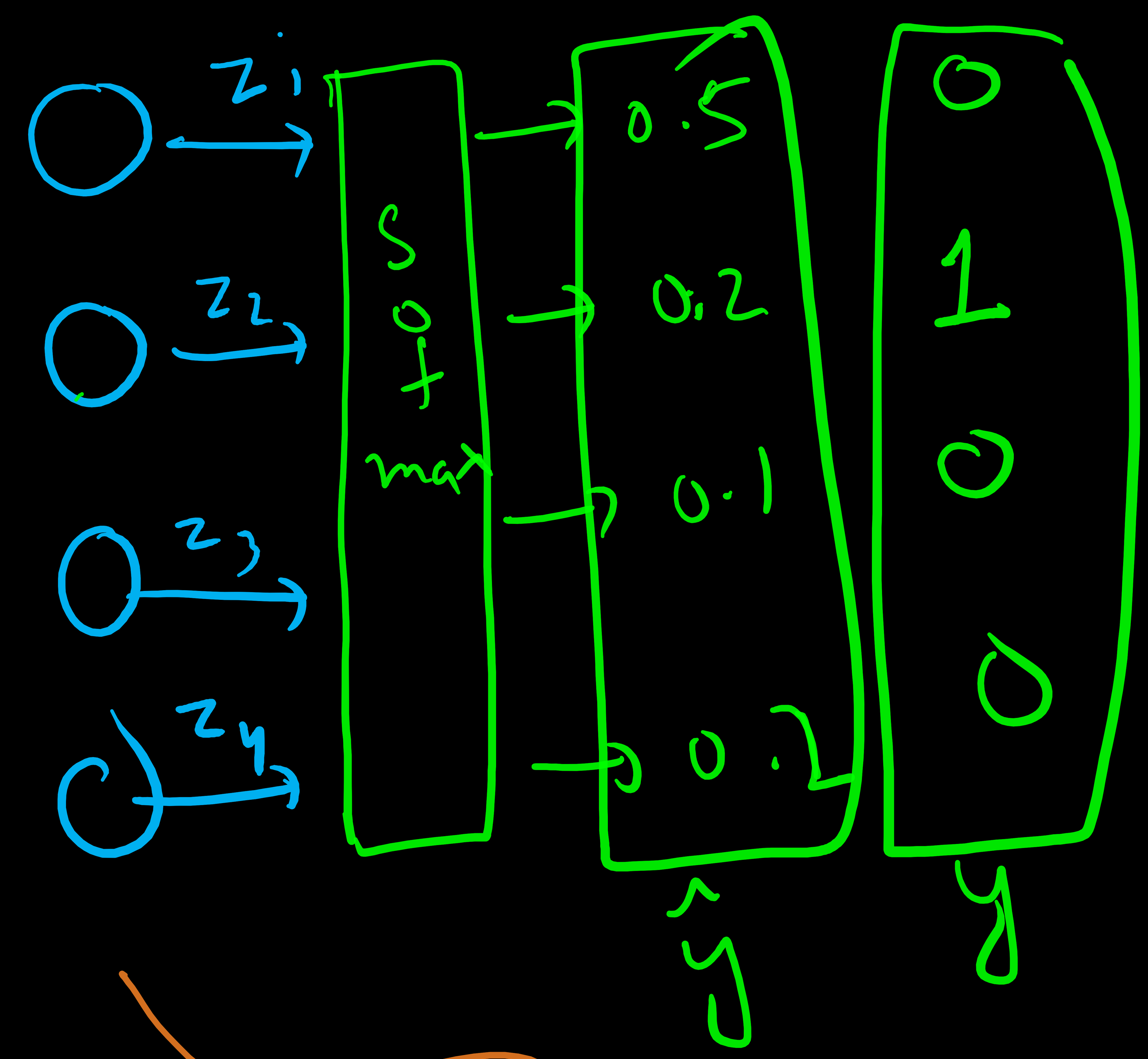
K categories

Binary classification

1 output





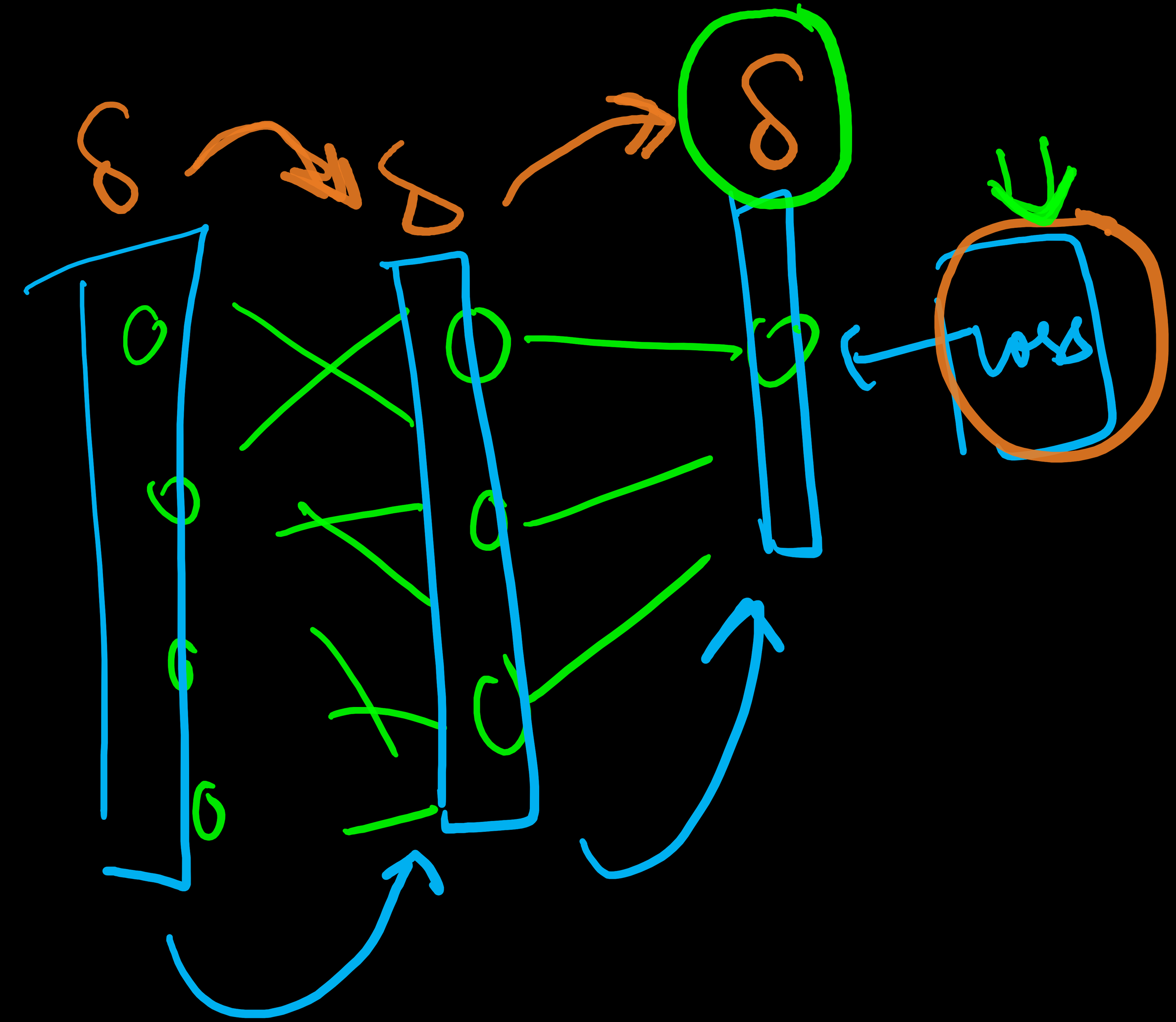


y_i → True Label

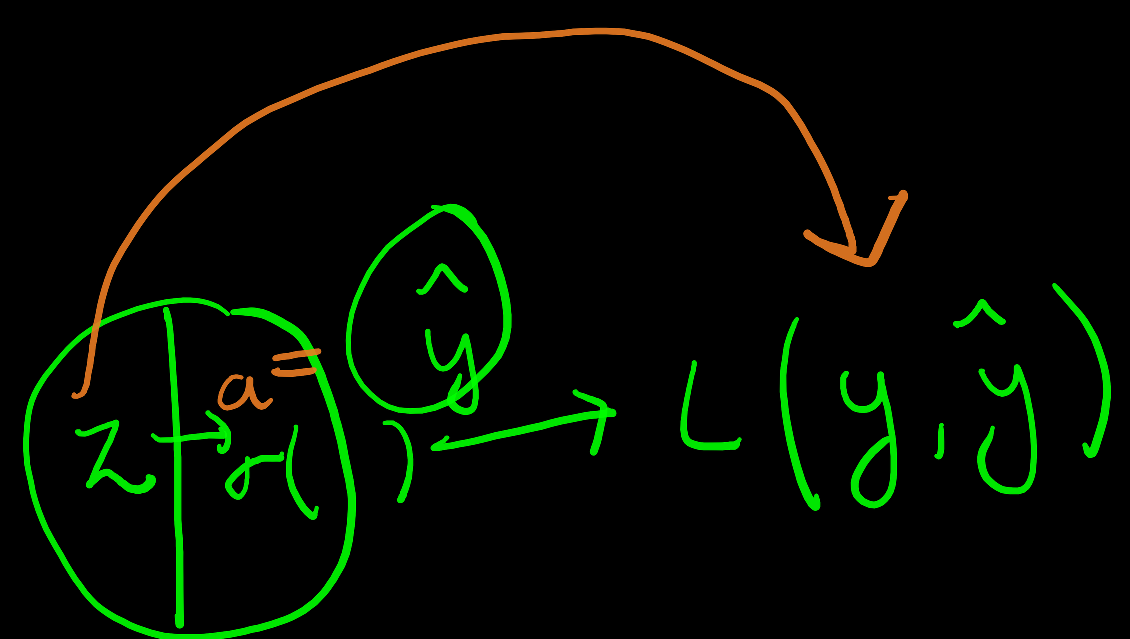
$\hat{y}_i$ → Predict

Binary Cross Entropy

$$-\sum_{i=1}^{m}\left(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right)$$



δ_L change!



$\hat{y} = a$

$$\delta_L = \frac{\partial L}{\partial z^L} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z}$$

$$= \left(-\frac{y_i}{\hat{y}_i} + \frac{1-y_i}{1-\hat{y}_i} \right) \underline{\hat{y}_i (1-\hat{y}_i)}$$

$$a = \sigma(z)$$

$$\frac{\partial a}{\partial z} = \sigma(z)(1-\sigma(z)) = a(1-a) = \hat{y}(1-\hat{y})$$

$$= \frac{(-y_i + \cancel{y_i\hat{y}_i} + \hat{y}_i - \cancel{\hat{y}_i y_i})}{\cancel{\hat{y}(1-\hat{y})}} \cancel{\hat{y}_i (1-\hat{y}_i)}$$

$$\boxed{\boxed{\delta_L = (\hat{y}_i - y_i)}}$$

Cross Entropy Loss.

Categorical Cross Entropy

$$L(y,\hat{y}) = -\frac{1}{N} \sum_{n \in N} \underbrace{\sum_{i \in C} y_{n,i} \log \hat{y}_{n,i}}$$

2 classes

$$= -\frac{1}{N} \sum_{n \in N} \left(y_{n,1} \log \hat{y}_{n,1} + y_{n,0} \log \hat{y}_{n,0} \right)$$

$$y_{n,0} = 1 - y_{n,1}$$

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z}$$

softmax

$$= \left(\hat{y}_i - y_i \right)$$

Predict

Target

CODE

For 1 example

For m examples

Almost

output $\delta^L = (a^L - y^L)$

hidden layers $\delta^l = (w^{l+1}\,\delta^{l+1}) \odot \sigma'(z^l)$

$$\frac{\partial L}{\partial b^l} = \delta^l$$

$$\frac{\partial L}{\partial w^l} = a^{l-1}(\delta^l)^T$$

$W = (h_{l-1}, h_l)$

h_{l-1}

h_l

h_{l-1}, h_l $h_{l-1} \times 1$ $1 \times h_l$

$\delta^L = a^L - y^L$

$\delta^l = (\delta^{l+1} \cdot w^{l+1\,T}) \odot \sigma'(z^l)$

$$\frac{\partial L}{\partial b^l} = \frac{1}{m}\,\text{np.sum}(\delta^l, \text{axis}=0)$$

$$\frac{\partial L}{\partial w^l} = a^{l-1\,T} \cdot \delta^l$$

Do make dimensions

$$\frac{\partial L}{\partial w^l} = \quad h^{l-1} \times m \quad m \times h^l$$

$= (h^{l-1}, h^l)$

Prob

$$a_L = \begin{bmatrix} \vert \\ a^{(i)} \\ \vdots \\ \vert \\ K \end{bmatrix} \Big\} \text{ Argmax}$$

$$a_L = \begin{bmatrix} \text{---} \; a^{(1)} \; \text{---} \\ \text{---} \; a^{(2)} \; \text{---} \\ \text{---} \; a^{(i)} \; \text{---} \\ \vdots \\ \text{---} \; a^{(m)} \; \text{---} \end{bmatrix} \quad m \times c$$

$$y = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{matrix} \leftarrow \text{Cat} \\ \leftarrow \text{dog} \\ \leftarrow \text{horse} \end{matrix}$$

$$y = \begin{bmatrix} 0\,0\,0\,1\,0\,0\,0 \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{bmatrix} \quad m \times c$$

$$b^l = \begin{bmatrix} b_1^l \\ b_2^l \\ b_3^l \end{bmatrix} \qquad \delta^l = \begin{bmatrix} \delta b_1 \\ \delta b_2 \\ \delta b_3 \end{bmatrix}$$

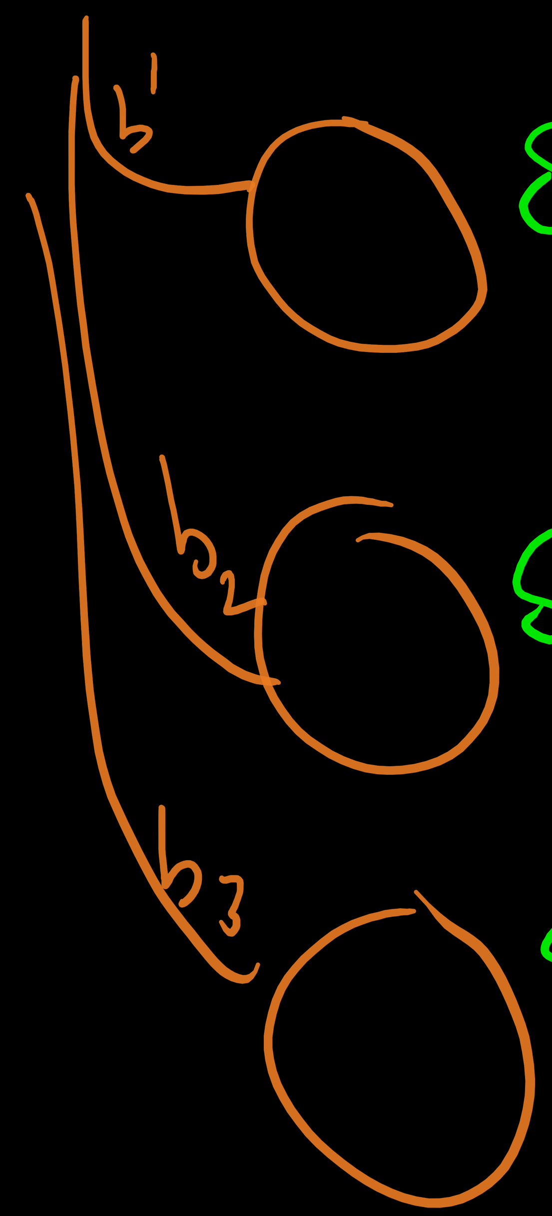
$$b^l = b^l - \eta \cdot \delta^l$$

$$\delta^L = m \times c$$

$$= \begin{bmatrix} \text{---} \; \delta^{l(1)} \; \text{---} \\ \text{---} \; \delta^{l(2)} \; \text{---} \\ \vdots \\ \text{---} \; \delta^{l(m)} \; \text{---} \end{bmatrix}$$

$b_1 =$ $\quad b_2$

$$= \begin{bmatrix} \delta_1^{l(1)} & \delta_2^{l(1)} & \delta_3^{l(3)} & \cdots \\ \delta_1^{l(2)} & & & \end{bmatrix}$$



$\delta_1^{l(1)}, \delta^{l(2)} \dots$

$\delta_2^{l(1)}, \delta^{l(2)} \dots$

$\delta_3^{l(1)}, \delta^{l(3)} \dots$

$\delta_1^{l(3)}$

$\delta_1^{l(4)}$

$\vdots$

$\delta_1^{l(m)}$ ...

b_1

$$b_1^l = b_1^l - \eta \cdot \sum_{i=1}^{m} \frac{\partial L}{\partial b_1^{l(i)}}$$