$$C = \frac{1}{2n} \sum_{i=1}^{n} \| \underset{\sim}{y}_i - \underset{\sim}{a}_i^L \|^2 \approx \underbrace{\frac{1}{2m} \sum_{i=1}^{m} \| \underset{\sim}{y}_i - \underset{\sim}{a}_i^L \|^2}_{\text{For Mini Batch}}$$

Let, p_j^l be a parameter of jth neuron in lth layer.

$$\therefore \frac{\partial C}{\partial p_j^l} \approx \frac{1}{2m} \sum_{i=1}^{m} \frac{\partial}{\partial p_j^l} \| \underset{\sim}{y}_i - \underset{\sim}{a}_i^L \|^2$$

Now, we do derivative for every data point in Mini Batch. So, we can drop the summation, index i when calculating derivative for a single data point. We also ignore the constant $\frac{1}{2m}$.

So, we get,

$$\frac{\partial}{\partial p_j^l} \| \underset{\sim}{y} - \underset{\sim}{a}^L \|^2 = \frac{\partial Q}{\partial p_j^l}$$

$$(Q = C \text{ if } m = 1)$$

Now,

$$\frac{\partial Q}{\partial p_j^l} = \frac{\partial Q}{\partial z_j^l} \cdot \frac{\partial z_j^l}{\partial p_j^l} = \delta_j^l \cdot \frac{\partial z_j^l}{\partial p_j^l} \quad \text{— } (*)$$

where, $\delta_j^l = \dfrac{\partial Q}{\partial z_j^l}$

$$z_j^l = w_{j1} a_1^{l-1} + w_{j2} a_2^{l-1} + w_{j3} a_3^{l-1} + \dots + w_{jN_{l-1}} a_{N_{l-1}}^{l-1} + b_j^l$$

We calculate δ_j^l for $l = L$, then using that calculate δ_t^{l-1}, so on.

For $l = L$

$$\delta_j^L = \frac{\partial Q}{\partial z_j^L} = \frac{\partial Q}{\partial a_j^L} \cdot \frac{\partial a_j^L}{\partial z_j^L}$$

$$= \frac{\partial}{\partial a_j^L} \| \underset{\sim}{y} - \underset{\sim}{a}^L \|^2 \frac{\partial}{\partial z_j^L} \sigma(z_j^L)$$

$$= \frac{\partial}{\partial a_j^L} \sum_{s=1}^{p} (y_s - a_s^L)^2 \cdot \frac{\partial}{\partial z_j^L} \sigma(z_j^L)$$

(In our case $p = 10$)

$$= (a_j^L - y_s)\ \sigma'(z_j^L)$$

$$\therefore \quad \delta^L = (\underset{\sim}{a}^L - \underset{\sim}{y}) \circ \sigma'(z^L) \quad \text{--- (1)}$$

$\circ$ is Hadamard product.

For any layer l,

$$\delta_j^l = \frac{\partial Q}{\partial z_j^l}$$

$$= \sum_{t=1}^{N_{l+1}} \frac{\partial Q}{\partial z_t^{l+1}} \frac{\partial z_t^{l+1}}{\partial z_j^l}$$

$$= \sum_{t=1}^{N_{l+1}} \delta_t^{l+1} \frac{\partial z_t^{l+1}}{\partial z_j^l} \quad \text{--- (* *)}$$

$$z_t^{l+1} = \sum_{s=1}^{N_l} w_{ts}^{l+1} a_s^l + b_t^{l+1}$$

$$= \sum_{s=1}^{N_l} w_{ts}^{l+1} \sigma(z_s^l) + b_t^{l+1}$$

$$\therefore \quad \frac{\partial z_t^{l+1}}{\partial z_j^l} = w_{tj}^{l+1} \sigma'(z_j^l)$$

$\therefore$ From (* *) $\delta_j^l = \sum_{t=1}^{N_{l+1}} \delta_t^{l+1} w_{tj}^{l+1} \sigma'(z_j^l)$

$$\therefore \quad \delta^{l} = \left((W^{l+1})^{T} \delta^{l+1} \right) \circ \sigma'(z^{l}) \quad — (2)$$

Now, from $(*)$

$$\frac{\partial Q}{\partial p_j^l} = \delta_j^l \cdot \frac{\partial z_j^l}{\partial p_j^l}$$

$$= \delta_j^l \cdot \frac{\partial}{\partial p_j^l} \left(w_{j1} a_1^{l-1} + w_{j2} a_2^{l-1} + w_{j3} a_3^{l-1} + \dots + w_{jN_{l-1}} a_{N_{l-1}}^{l-1} + b_j^l \right)$$

If, $p_j^l = b_j^l$

then, $$\frac{\partial Q}{\partial b_j^l} = \delta_j^l \quad — (3)$$

If, $p_j^l = w_{jr}^l$

then, $$\frac{\partial Q}{\partial w_{jr}^l} = a_r^{l-1} \quad — (4)$$

In my code ~~Back~~ backpropagation function implements only (1) - (4), update function adds the derivatives and divide the sum by Mini Batch size.