1. a) $w_{ji} := w_{ji} - \alpha \dfrac{\partial C}{\partial w_{ji}}$

$$\frac{\partial C}{\partial w_{ji}} = \frac{\partial C}{\partial z_j} \frac{\partial z_j}{\partial w_{ji}}$$

$$= \delta_j \cdot \frac{\partial}{\partial w_{ji}} \left(\sum_i w_{ji} x_i \right)$$

$$= \delta_j \cdot \frac{\partial}{\partial w_{ji}} \left(w_{ji} x_i + \sum_{l \neq i} w_{jl} x_l \right)$$

$$= \underline{\delta_j x_i}$$

$$\delta_j = \frac{\partial C}{\partial z_j} = \sum_k \frac{\partial C}{\partial z_k} \frac{\partial z_k}{\partial a_j} \frac{\partial a_j}{\partial z_j}$$

$$= \sum_k \delta_k \frac{\partial}{\partial a_j} \left(\sum_j w_{kj} a_j \right) \frac{\partial f(z_j)}{\partial z_j}$$

$$= \sum_k \delta_k \frac{\partial}{\partial a_j} \left(w_{kj} a_j + \sum_{m \neq j} w_{km} a_m \right) f'(z_j)$$

$$= \underline{f'(z_j) \sum_k w_{kj} \delta_k}$$

1.b) Hidden layer to output layer:

$w_{kj} := w_{kj} - \alpha \delta_k a_j$

w is a 64×10 matrix

α is a scalar

δ is a Batch_size $\times$ 10 matrix

a is a Batch-size $\times$ 64 matrix

$\underline{w := w - \alpha a^T \delta}$

Input layer to ~~the~~ hidden layer:

$w_{ji} := w_{ji} - \alpha \delta_j x_i$

w is a 785×64 matrix

α is a scalar

δ is a Batch-size $\times$ 64 matrix

X is a Batch_size $\times$ 785 matrix

$\underline{w := w - \alpha X^T \delta}$