(a)

$$a_j = f(z_j)$$

$$z_j = \sum_{i=0}^{I} w_{ji} x_i$$

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

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

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

$$= \delta_j \times \frac{\partial z_j}{\partial w_{ji}}$$

$$= \delta_j \times x_i$$

$$w_{ji} := w_{ji} - \alpha \frac{\partial C}{\partial w_{ji}}$$

$\therefore$ $w_{ji} = w_{ji} - \alpha \delta_j x_i$ (shown)

$$z_{kj} = \sum_{j=0}^{J} w_{kj} x_j$$

$$C = \sum_k C_k$$

$$\frac{\partial C}{\partial C_k} = 1$$

$$\frac{\partial C}{\partial z_k} = \sum_k \frac{\partial C_k}{\partial z_k}$$

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

$$= \sum_k \frac{\partial C_k}{\partial z_k} \times \frac{\partial z_k}{\partial a_j} \times \frac{\partial a_j}{\partial z_j}$$

$$\frac{\partial C}{\partial C_k} \times \frac{\partial C_k}{\partial z_k} = \frac{\partial C}{\partial z_k}$$

$$1 \times \frac{\partial C_k}{\partial z_k} = \frac{\partial C}{\partial z_k}$$

$$\frac{\partial C}{\partial z_k} = \frac{\partial C_k}{\partial z_k}$$

$$\therefore \quad \delta_j = \sum_k \frac{\partial C}{\partial z_k} \times \frac{\partial z_k}{\partial a_j} \times \frac{\partial a_j}{\partial z_j}, \text{ with } \delta_k = \frac{\partial C}{\partial z_k}$$

$$z_k = \sum_j w_{kj} a_j$$

$$\frac{\partial z_k}{\partial a_j} = w_{kj}$$

$$a_j = f(z_j)$$

$$\frac{\partial a_j}{\partial z_j} = f'(z_j)$$

$$\therefore \quad \delta_j = \sum_k \delta_k \times w_{kj} \times f'(z_j)$$

$$= f'(z_j) \sum_k w_{kj} \delta_k \qquad \text{(Shown)}$$

1b)

Hidden layer – Output layer

$[m, n]$ means a $m \times n$ matrix

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

$$\delta_k = (\hat{y}_k - y_k)$$

$$= \begin{bmatrix} \hat{y}_1 - y_1 \\ \hat{y}_2 - y_2 \\ \vdots \\ \hat{y}_k - y_k \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_k \end{bmatrix} \rightarrow [k, 1]$$

$$a_j = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_j \end{bmatrix} \rightarrow [j, 1]$$

$$a_j^T = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_j \end{bmatrix} \rightarrow [1, j]$$

$$\alpha \delta_k a_j^T = \alpha \begin{bmatrix} \delta_1 a_1 & \delta_1 a_2 & \cdots & \delta_1 a_j \\ \delta_2 a_1 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \delta_k a_1 & \cdots & \cdots & \delta_k a_j \end{bmatrix} \rightarrow [k, j]$$

α is a scalar

$$w_{kj} := \begin{bmatrix} w_{11} - \alpha \delta_1 a_1 & w_{12} - \alpha \delta_1 a_2 & \cdots & w_{1j} - \alpha \delta_1 a_j \\ w_{21} - \alpha \delta_2 a_1 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ w_{k1} - \alpha \delta_k a_1 & \cdots & \cdots & w_{kj} - \alpha \delta_k a_j \end{bmatrix} \rightarrow [k, j]$$

Input to hidden

$w_{ji} := w_{ji} - \alpha \, \delta_j x_i^T$ where α is a scalar.

$$w_{ji} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1i} \\ w_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ w_{j1} & \cdots & \cdots & w_{ji} \end{bmatrix} \rightarrow [j, i]$$

$$x_i = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \end{bmatrix} \rightarrow [i, 1]$$

$$x_i^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_i \end{bmatrix} \rightarrow [1, i]$$

$$\delta_j = \begin{bmatrix} f'(z_1) \sum_k \delta_k w_{kj} \\ f'(z_2) \sum_k \delta_k w_{kj} \\ \vdots \\ f'(z_j) \sum_k \delta_k w_{kj} \end{bmatrix} \rightarrow [j, 1]$$

δ_k defined earlier.

$$w_{kj} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1j} \\ w_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ w_{k1} & \cdots & \cdots & w_{kj} \end{bmatrix} \rightarrow [k, j]$$

$$\sum_k \delta_k w_{kj} = \begin{bmatrix} \sum_k \delta_k w_{kj} \\ \sum_k \delta_k w_{kj} \\ \vdots \\ \sum_k \delta_k w_{kj} \end{bmatrix} \rightarrow [j, 1]$$

$$f'(z_j) = \begin{bmatrix} f'(z_1) \\ f'(z_2) \\ \vdots \\ f'(z_j \end{bmatrix} \rightarrow [j, 1]$$