# Back Propagation Algorithm

$$x_i \longrightarrow (i) \xrightarrow[w_{ij}^h]{s_i = x_i} (j) \xrightarrow[w_{jk}^o]{s_j^h} (k) \xrightarrow{s_k^o} y_k$$

$$\Rightarrow \quad s_j^h = f_i^h(a_j^h) \qquad \text{----} ①$$

$$a_j^h = \sum_{i=1}^{I} w_{ij}^h s_i - \theta_j^h \quad , \text{ where } s_i = x_i \qquad \text{----} ②$$

$$\Rightarrow \quad s_k^o = f_k^o(a_k^o) \qquad \text{----} ③$$

$$a_k^o = \sum_{j=1}^{J} w_{jk}^o s_j^h - \theta_k^o \qquad \text{----} ④$$

$$y_k = s_k^o = f_k^o\left[\sum_{j=1}^{J} w_{jk}^o f_j^h\left[\sum_{i=1}^{I} w_{ij}^h s_i - \theta_j^h\right] - \theta_k^o\right] \qquad \text{---} ⑤$$

## I. Function Approximation:

The general error, $\hat{E}_n = \frac{1}{2}\sum_{k=1}^{K}(t_{nk} - y_{nk})^2$

Stochastic Gradient Descent:

$$\Delta W(m) = -\eta \frac{\partial \bar{E}(m)}{\partial W}$$

$\Delta w_{jk} \Rightarrow$

$$\boxed{\Delta w_{jk}^o = -\eta \frac{\partial E}{\partial w_{jk}}} \qquad \text{----} ⑥$$

$$\frac{\partial E}{\partial w_{jk}^o} = \frac{\partial}{\partial w_{jk}^o} * \frac{1}{2} * \sum_{l=1}^{K}(t_l - y_l)^2$$

$$\frac{\partial E}{\partial w_{jk}^{o}} = \frac{-\partial}{\partial w_{jk}} (t_k - y_k) \cdot y_k$$

$$= -(t_k - y_k) \cdot \frac{\partial y_k}{\partial w_{jk}} = -(t_k - y_k) \cdot \frac{\partial s_k^{o}}{\partial w_{jk}}$$

$$\frac{\partial s_k^{o}}{\partial w_{jk}^{o}} = \frac{d f_k^{o}(a_k^{o})}{\partial a_k^{o}} \cdot \frac{\partial a_k^{o}}{\partial w_{jk}} \quad \text{-----} \; (7)$$

Substituting (4) in (7),

$$\frac{\partial s_k^{o}}{\partial w_{jk}^{o}} = \frac{d f_k^{o}(a_k^{o})}{\partial a_k^{o}} \cdot s_j^{h}$$

Then Equ. (6) becomes,

$$\Delta w_{jk}^{o} = -\eta \, (t_k - y_k) \frac{d f_k^{o}(a_k^{o})}{d a_k^{o}} \cdot s_j^{h}$$

$$\boxed{\Delta w_{jk}^{o} = -\eta \cdot \delta_k^{o} \, s_j^{h}} \quad \text{-----} \; (8)$$

$\eta \longrightarrow$ learning rate

$\delta_k^{o} \longrightarrow$ local gradient $= (t_k - y_k) \cdot \dfrac{d f_k^{o}(a_k^{o})}{d a_k^{o}}$

$s_j^{h} \longrightarrow$ Input from previous hidden layer

$$\Delta w_{ij}^h \Rightarrow \quad \Delta w_{ij}^h = -\eta \cdot \frac{\partial \bar{E}}{\partial w_{ij}^h}$$

$$= -\eta (t_k - y_k) \cdot \frac{\partial y_k}{\partial w_{ij}^h}$$

$$\Delta w_{ij}^h = -\eta \cdot (t_k - y_k) \cdot \frac{\partial s_k^o}{\partial w_{ij}^h} \quad \text{-----} (9)$$

$$= -\eta \cdot (t_k - y_k) \cdot \frac{d f_k^o (a_k^o)}{d a_k^o} \cdot \frac{\partial a_k^o}{\partial w_{ij}^h}$$

$$= -\eta \cdot \underbrace{(t_k - y_k) \cdot \frac{d f_k^o (a_k^o)}{d a_k^o}}_{\delta_k^o} \cdot \frac{\partial}{\partial w_{ij}^h} \left[ \sum_{j=1}^{J} w_{jk} s_j^h - \theta_k^o \right]$$

$$= -\eta \cdot (t_k - y_k) \cdot \frac{d f_k^o (a_k^o)}{d a_k^o} \cdot w_{jk} \cdot \frac{d s_j^h}{d a_j^h} \cdot \frac{\partial a_j^h}{\partial w_{ij}^h}$$

Final update equation of layer i-j,

$$\Delta w_{ij}^h = -\eta \delta_k^o w_{jk} \cdot \frac{\partial f(a_j^h)}{\partial a_j^h} \cdot \frac{\partial a_j^h}{\partial w_{ij}^h}$$

$$\Delta w_{ij}^h = -\eta \underbrace{\delta_k^o w_{jk} \cdot \frac{\partial f(a_j^h)}{\partial a_j^h}}_{\delta_j^h} \cdot s_i$$

$$\boxed{\Delta w_{ij}^h = -\eta \delta_j^h s_i} \quad \text{-----} (10)$$