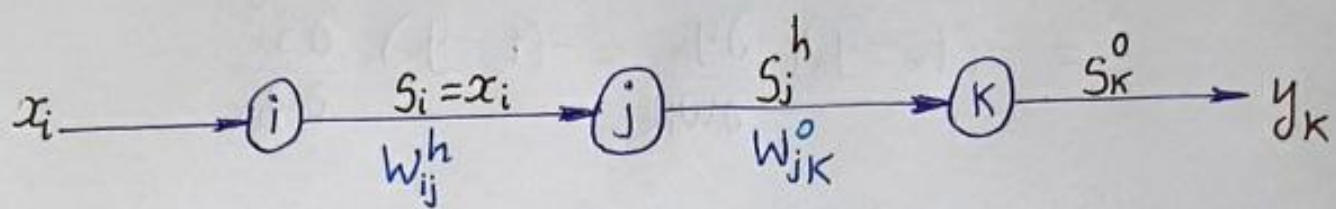


# Back Propagation Algorithm



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

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

$$\Rightarrow S_K^o = f_K^o(a_K^o) \quad \text{--- (3)}$$

$$a_K^o = \sum_{j=1}^J w_{jK}^o S_j^h - \theta_K^o \quad \text{--- (4)}$$

$$y_K = S_K^o = f_K^o \left[ \sum_{j=1}^J w_{jK}^o f_j^h \left[ \sum_{i=1}^I w_{ij}^h S_i - \theta_j^h \right] - \theta_K^o \right] \quad \text{--- (5)}$$

## I. Function Approximation:

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

Stochastic Gradient Descent:

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

$$\Delta w_{jk} \Rightarrow$$

$$\Delta w_{jk}^o = -\eta \frac{\partial \bar{E}}{\partial w_{jk}^o}$$

$$\text{--- (6)}$$

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

$$\begin{aligned}\frac{\partial E}{\partial w_{jk}^0} &= \frac{-\partial}{\partial w_{jk}} (t_k - y_k) \cdot y_k \\ &= -(t_k - y_k) \cdot \frac{\partial y_k}{\partial w_{jk}} = -(t_k - y_k) \cdot \frac{\partial s_k^0}{\partial w_{jk}}\end{aligned}$$

$$\frac{\partial s_k^0}{\partial w_{jk}^0} = \frac{df_k^0(a_k^0)}{da_k^0} \cdot \frac{\partial a_k^0}{\partial w_{jk}} \quad \text{--- --- --- (7)}$$

Substituting (4) in (7),

$$\frac{\partial s_k^0}{\partial w_{jk}^0} = \frac{df_k^0(a_k^0)}{da_k^0} \cdot s_j^h$$

Then Equ. (6) becomes,

$$\Delta w_{jk}^0 = -\eta (t_k - y_k) \frac{df_k^0(a_k^0)}{da_k^0} \cdot s_j^h$$

$$\boxed{\Delta w_{jk}^0 = -\eta \cdot \delta_k^0 s_j^h} \quad \text{--- --- --- (8)}$$

$\eta \rightarrow$  learning rate

$\delta_k^0 \rightarrow$  local gradient  $= (t_k - y_k) \cdot \frac{df_k^0(a_k^0)}{da_k^0}$

$s_j^h \rightarrow$  Input from previous hidden layer



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

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

$$\Delta w_{ij}^h = -\eta \cdot (t_k - y_k) \cdot \frac{\partial s_k^0}{\partial w_{ij}^h} \text{ --- --- --- (9)}$$

$$= -\eta \cdot (t_k - y_k) \cdot \frac{df_k^0(a_k^0)}{da_k^0} \cdot \frac{\partial a_k^0}{\partial w_{ij}^h}$$

$$= -\eta \cdot (t_k - y_k) \cdot \underbrace{\frac{df_k^0(a_k^0)}{da_k^0}}_{\delta_k^0} \cdot \frac{\partial}{\partial w_{ij}^h} \left[ \sum_{j=1}^J w_{jk} s_j^h - \theta_k^0 \right]$$

$$= -\eta \cdot (t_k - y_k) \cdot \frac{df_k^0(a_k^0)}{da_k^0} \cdot w_{jk} \cdot \frac{ds_j^h}{da_j^h} \cdot \frac{\partial a_j^h}{\partial w_{ij}^h}$$

Final update equation of layer i-j,

$$\Delta w_{ij}^h = -\eta \delta_k^0 w_{jk} \cdot \frac{\partial f(a_j^h)}{\partial a_j^h} \cdot \frac{\partial a_j^h}{\partial w_{ij}^h}$$

$$\Delta w_{ij}^h = -\eta \underbrace{\delta_k^0 w_{jk} \cdot \frac{\partial f(a_j^h)}{\partial a_j^h}}_{\delta_j^h} \cdot s_i$$

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