1

The Error of N/w

$$E(w) = \frac{1}{2}\sum_{k=1}^{N}(y_k - t_k)^2 = \frac{1}{2}\sum_{k=1}^{N}\left[g\left(\sum_{j=0}^{M} w_{jk} a_j\right) - t_k\right]^2$$

Err in terms of input units.

$$= \frac{1}{2}\sum_{k=1}^{N}\left[g\left(\sum_{i=0}^{L} w_{ik} x_i\right) - t_k\right]^2$$

$$\frac{\partial E}{\partial w_{ik}} = \frac{\partial}{\partial w_{ik}}\left(\frac{1}{2}\sum_{k=1}^{N}(y_k - t_k)^2\right)$$

$$= \frac{1}{\cancel{2}}\sum_{k=1}^{N}\cancel{2}(y_k - t_k)\frac{\partial}{\partial w_{ik}}\left(\left(\sum_{i=0}^{L} w_{ik} x_i\right) - t_k\right)^2$$

$$= \sum_{k=1}^{N}(t_k - y_k)(-x_i) \qquad \frac{\partial(t_k)}{\partial w_{ik}} = 0.$$

So weight update.

$$\boxed{w_{ik} = w_{ik} + \eta(t_k - y_k) \cdot x_i}$$

Activation function

$$a = g(h) = \frac{1}{1+e^{-\beta h}} = \frac{d}{dh}(1 + e^{-\beta h})^{-1}$$

$$= -1 \cdot (1+e^{-\beta h})^{-2} \cdot \frac{d(e^{-\beta h})}{dh}$$

$$= -1\,(1+e^{-\beta h})^{-2}\,(\cdot e^{-\beta h}) \cdot (-\beta)$$

β $\leftarrow$

$$= \frac{\beta e^{-\beta h}}{(1+e^{-\beta h})^2} = \beta\left(\cancel{\frac{1 + e^{-\beta h} - 1}{(1+e^{-\beta h})^2}}\right)$$

$$= \frac{\beta e^{-\beta h}}{(1+e^{-\beta h})^2} = \beta \cdot \left(\frac{1}{1+e^{-\beta h}}\right) \cdot \left(\frac{1 + e^{-\beta h} - 1}{1+e^{-\beta h}}\right)$$

$$= \beta \cdot g(h) \cdot \left(\left(\frac{1+e^{-\beta h}}{1+e^{-\beta h}}\right) - \frac{1}{1+e^{-\beta h}}\right)$$

$$= \underline{\underline{\beta \cdot a \cdot (1-a)}}.$$

Back-propagation of error

$$\frac{\partial E}{\partial w_{ik}} = \frac{\partial E}{\partial h_k} \cdot \frac{\partial h_k}{\partial w_{ik}} \quad \text{--- (1)}$$

(2)

$h_k = \sum_{j=0}^{M} w_{jk} \cdot a_j$ is the input to output neuron k.

ie the sum of activations of hidden layer neurons multiplied by second layer weight.

- It tell that if we want to ~~change~~ know how the err at the output changes as we vary the second layer weight, in turn how the err changes as we vary the input to output neuron, also how those input values change as we vary the weights (first layer weight).

(2)

(1) - $$\frac{\partial h_k}{\partial w_{jk}} = \frac{\partial \left(\sum_{j=0}^{M} w_{jk} \cdot a_j \right)}{\partial w_{jk}}$$

$$= \sum_{j=0}^{M} \frac{\partial (w_{jk} \cdot a_j)}{\partial w_{jk}} = \underline{\underline{a_j}}$$

(2) $$\frac{\partial E}{\partial h_k} = \delta_o(k) = \frac{\partial E}{\partial y_k} \cdot \frac{\partial y_k}{\partial h_k} = \frac{\partial E}{\partial h_k}.$$

$$y_k = g(h_k^{output}) = g\left(\sum_{j=0}^{M} w_{jk} \cdot a_j^{hidden} \right)$$

$$\delta_o(k) = \frac{\partial E}{\partial g(h_k^{output})} \cdot \frac{\partial g(h_k^{output})}{\partial h_k^{output}}$$

$$= \frac{\partial E}{\partial g(h_k^{output})} g'(h_k^{output})$$

$$= \frac{\partial}{\partial g(h_k^{output})} \overbrace{\left[\frac{1}{2} \sum_{k=1}^{N} \left(g(h_k^{output}) - t_k \right)^2 \right]}^{E} g'(h_k^{output})$$

$$= \left(g(h_k^{output}) - t_k \right) g'(h_k^{output})$$

$$= (y_k - t_k) \quad \text{"} \quad y_k(1 - y_k)$$

Lactivation function derivative.

in case of sigmoid function.

$$\boxed{\delta_o(k) = (y_k - t_k) \, y_k(1 - y_k)}$$

So the weight update.

$$w_{jk} = w_{jk} - \eta \frac{\partial E}{\partial w_{jk}}$$

$$= w_{jk} - \eta \, \delta_o(k) \cdot a_j$$

- because to go downhill the error.

Same way first layer err. and weight updates

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial h_j} \cdot \frac{\partial h_j}{\partial w_{ij}}$$

① $$\frac{\partial h_j}{\partial w_{ij}} = \frac{\partial \sum_{i=0}^{L} w_{ij} \cdot x_i}{\partial w_{ij}} = \sum_{i=0}^{L} \frac{\partial (w_{ij} \cdot x_i)}{\partial w_{ij}}$$

$$= x_i$$

② $$\delta_h(j) = \frac{\partial E}{\partial h_j} = a_j(1-a_j) \cdot \sum_{k=1}^{N} w_{jk} \cdot \delta_o(k).$$

$$\boxed{w_{ij} = w_{ij} - \eta \cdot \delta_h(j) \cdot x_i}$$

③

1st layer weight updates

$$\delta_h(j) = \sum_{k=1}^{N} \frac{\partial E}{\partial h_k^{output}} \cdot \frac{\partial h_k^{output}}{\partial h_j^{hidden}}$$

② ①

$$h_k^{output} = \sum_{j=0}^{M} w_{jk} \cdot g(h_j^{hidden}).$$

① $$\frac{\partial h_k^{output}}{\partial h_j^{hidden}} = \frac{\partial \sum_{j=0}^{M} \left(w_{jk} \cdot g(h_j^{hidden}) \right)}{\partial h_j^{hidden}} =$$

$$= w_{jk} \cdot g'(a_j)$$ — hidden layers always have sigmoidal activation function

$$g'(a_j) = \beta \cdot a_j (1 - a_j).$$

$$\delta_h(j) = \boxed{\beta \cdot a_j (1 - a_j)} \sum_{k=1}^{N} \cdot \boxed{\delta_o(k)} \cdot \boxed{w_{jk}}$$

② ①