
Artificial Neural Network

■ Logistic Sigmoid

$$\frac{1}{1 + e^{-\frac{x}{a}}}$$

■ Hyperbolic Tangent

$$\frac{e^h - e^{-h}}{e^h + e^{-h}}$$

■ ArcTangent

ArcTan

W_{AB}

The weight of the connection from neuron A to neuron B

$$o[x_1, x_2, \dots, x_n] = \begin{cases} 1 & \text{if } w_0 + w_1 + \dots + w_n > 0 \\ -1 & \text{otherwise} \end{cases}$$

■ Learning in the Perceptron

$$\Delta w_i = \eta (t - o) x_i$$

■ The delta training learning rule

the difference between the desired output t and the actual output o , given input w_i

■ Global error E

between the desired and actual output

■ Gradient Descent

Training error of a weight vector

$$E[\hat{w}] \equiv \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2$$

E is a function of weight vector

D is set of training examples.

t_d is the target output for the training example d .

o_d is the output of the linear unit for the training example d .

The error surface is parabolic

■ Derivation of the Gradient Descent Rule

$$\nabla E[\hat{W}] \equiv \left[\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \dots, \frac{\partial E}{\partial w_n} \right]$$

the partial derivatives of E with respect to each of the w_i

$$W \leftarrow W + \Delta W$$

$$\Delta \hat{W} = -\eta \nabla E[\hat{W}]$$

η is the learning rate (the step size)

$$w_i \leftarrow w_i + \Delta w_i$$

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i}$$

■ Derivation

$$E[W] = \frac{1}{2} \sum (t_d - o[z_d])^2$$

$$\nabla E = \left[\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \dots, \frac{\partial E}{\partial w_n} \right]$$

$$z_d = w_0 + w_1 x_{1d} + \dots + w_n x_{nd}$$

$$o[z_d] = o[w_0 + w_1 x_{1d} + \dots + w_n x_{nd}]$$

$$\begin{aligned} \frac{\partial E}{\partial w_i} &= \frac{\partial \frac{1}{2} \sum (t_d - o[z_d])^2}{\partial w_i} \\ &= \sum (t_d - o[z_d]) \frac{\partial (t_d - o[z_d])}{\partial w_i} \\ &= \sum (t_d - o[z_d]) \frac{\partial (t_d - o[z_d])}{\partial z_d} \frac{\partial z_d}{\partial w_i} \\ &= \sum (t_d - o[z_d]) * -1 * x_i \\ &= -(t - o[z]) x_i \end{aligned}$$

$$\frac{\partial (t_d - o[z_d])}{\partial z_d} = - \frac{\partial o[z_d]}{\partial z_d} = -1$$

$$\frac{\partial z_d}{\partial w_0} = 1, \quad \frac{\partial z_d}{\partial w_1} = x_{1d}, \quad \dots, \quad \frac{\partial z_d}{\partial w_n} = x_{nd}$$

$$\Delta \hat{W} = -\eta \nabla E[\hat{W}]$$

$$\begin{aligned}
 \Delta w_i &= -\eta \frac{\partial E}{\partial w_i} \\
 &= -\eta \sum (t_d - o[z_d]) * -1 * x_{i_d} \\
 &= \eta (t - o) x_i \\
 \Delta w_i &= \eta (t - o) x_i
 \end{aligned}$$

■ Exam

$$o[z_d] = o[w_0 + w_1 x_d + w_2 y_d]$$

$$z_d = w_0 + w_1 x_d + w_2 y_d$$

$$E = \frac{1}{2} \sum (t_d - o[z_d])^2 = \frac{1}{2} \sum (t_d - o[w_0 + w_1 x_d + w_2 y_d])^2$$

$$\nabla E = \left[\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2} \right]$$

$$\frac{\partial E}{\partial w_0} = - (t_d - o[z_d]) \cdot 1$$

$$\frac{\partial E}{\partial w_1} = - (t_d - o[z_d]) \cdot x$$

$$\frac{\partial E}{\partial w_2} = - (t_d - o[z_d]) \cdot y$$

$$w_0 = w_0 + \eta (t_d - o[z_d]) \cdot 1;$$

$$w_1 = w_1 + \eta (t_d - o[z_d]) \cdot x;$$

$$w_2 = w_2 + \eta (t_d - o[z_d]) \cdot y;$$

MATLAB

```

for d = 1 : length (t_d)
    if (w0 + w1 * X1_d + w2 * X2_d) > 0
        Y_d = 1;
    else
        Y_d = -1;
    end

    w0 = w0 + eta (t_d - Y_d) * 1;
    w1 = w1 + eta (t_d - Y_d) * X1_d;
    w2 = w2 + eta (t_d - Y_d) * X2_d;
end

```

Simiar

■ 1. Error 정의

$$E = \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2$$

■ 2. Gradient Descent

$$\nabla E = \left[\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \dots, \frac{\partial E}{\partial w_n} \right]$$

■ 3. Minimum Error

$$\Delta W = -\eta \nabla E$$

■ 4. component

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i}$$

Backpropagation

Initialize all weights to small random numbers Until satisfied, Do

For each training example, Do

1. Input the training example to the network and compute the network outputs

2. For each output unit k

$$\delta_k \frac{df_k[e]}{de} \leftarrow (t_k - o_k) o_k (1 - o_k)$$

3. For each hidden unit h

$$\delta_h \frac{df_h[e]}{de} \leftarrow \left(\sum_{k \in \text{outputs}} w_{hk} \delta_k o_k (1 - o_k) \right) o_h (1 - o_h)$$

4. Update each network weight $w_{i,j}$

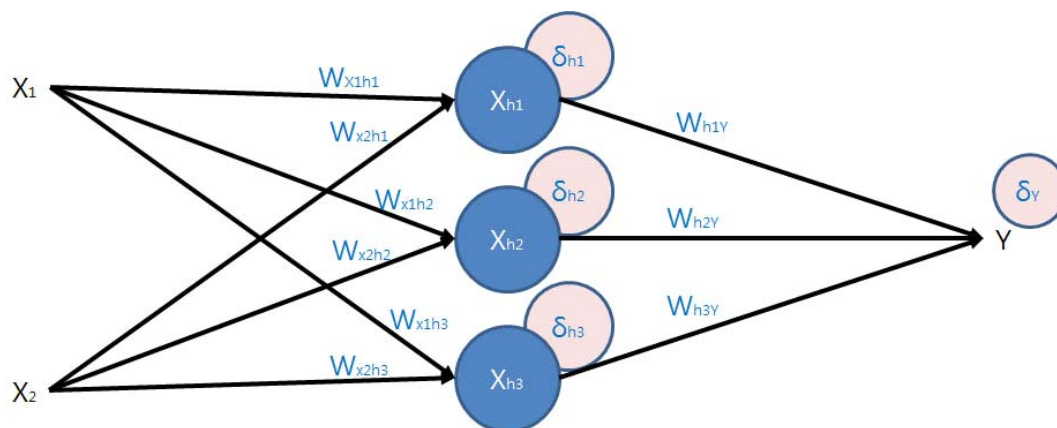
$$w_{ij} \leftarrow w_{ij} + \Delta w_{ij}$$

$$\Delta w_{ij} = \eta \delta_j \frac{df_j[e]}{de} x_i$$

$$E[\hat{w}] = \frac{1}{2} \sum (t_d - o[z_d])^2$$

$$o[z_d] = \frac{1}{1 + e^{-(z_d)}}$$

Derivation



$$E = \frac{1}{2} \sum_{d \in D} (t_d - o[z_d])^2$$

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}$$

$$\Delta w_{ij} = \eta \delta_j \frac{df_j[e]}{de} x_i$$

$$x_k \rightarrow w_{kh} \rightarrow x_h$$

$$x_h = o \left[\sum_k w_{kh} x_k \right]$$

■ w_{ho}

$$\Delta w_{ho} = -\eta \frac{\partial E}{\partial w_{ho}}$$

$$\begin{aligned} \Delta w_{ho} &= -\eta \frac{\partial E}{\partial w_{ho}} = -\eta \frac{\partial \frac{1}{2} \sum_{d \in D} (t_d - o[z_d])^2}{\partial w_{ho}} \\ &= -\eta \frac{\partial \frac{1}{2} \sum_{d \in D} (t_d - o[z_d])^2}{\partial z_d} \frac{\partial z_d}{\partial w_{ho}} \\ &= \eta \sum_{d \in D} (t_d - o[z_d]) \frac{\partial o[z_d]}{\partial z_d} \frac{\partial z_d}{\partial w_{ho}} \\ &= \eta \sum_{d \in D} (t_d - o[z_d]) o[z_d] (1 - o[z_d]) x_h \\ &= \eta \sum_{d \in D} (t - o) o (1 - o) x_h \\ &= \eta \sum_{d \in D} \delta_o o (1 - o) x_h \\ &= \eta \sum_{d \in D} \delta_o \frac{df_o[e]}{de} x_h \end{aligned}$$

$$\blacksquare \Delta w_{ho} = \eta \sum_{d \in D} \delta_o \frac{df_o[e]}{de} x_h$$

$$\therefore x_o = o[z_d] = \frac{1}{1 + e^{-(z_d)}}$$

$$z_d = \sum_h w_{ho} x_h$$

$$\begin{aligned} \therefore \frac{\partial (t_d - o[z_d])}{\partial z_d} &= - \frac{\partial o[z_d]}{\partial z_d} \\ &= - \frac{\partial}{\partial z_d} \frac{1}{1 + e^{-z_d}} = - (-e^{-z_d}) (- (1 + e^{-z_d})^{-2}) \\ &= - \frac{e^{-z_d}}{(1 + e^{-z_d})^2} = - \frac{1}{1 + e^{-z_d}} * \frac{e^{-z_d}}{1 + e^{-z_d}} \\ &= - o[z_d] (1 - o[z_d]) \end{aligned}$$

$$\therefore \frac{\partial z_d}{\partial w_{ho}} = \frac{\partial \sum_h w_{ho} x_h}{\partial w_{ho}} = x_h$$

■ w_{ih}

$$\Delta w_{ih} = -\eta \frac{\partial E}{\partial w_{ih}}$$

$$\begin{aligned} \Delta w_{ih} &= -\eta \frac{\partial E}{\partial w_{ih}} = -\eta \frac{\partial \frac{1}{2} \sum_{d \in D} (t_d - o[z_d])^2}{\partial w_{ih}} \\ &= -\eta \frac{\partial \frac{1}{2} \sum_{d \in D} (t_d - o[z_d])^2}{\partial z_d} \frac{\partial z_d}{\partial w_{ih}} \\ &= \eta \sum_{d \in D} (t_d - o[z_d]) \frac{\partial o[z_d]}{\partial z_d} \frac{\partial z_d}{\partial w_{ih}} \\ &= \eta \sum_{d \in D} (t_d - o[z_d]) o[z_d] (1 - o[z_d]) \frac{\partial z_d}{\partial w_{ih}} \\ &= \eta \sum_{d \in D} (t_d - o[z_d]) o[z_d] (1 - o[z_d]) \frac{\partial z_d}{\partial x_h} \frac{\partial x_h}{\partial u_d} \frac{\partial u_d}{\partial w_{ih}} \\ &= \eta \sum_{d \in D} (t_d - o[z_d]) o[z_d] (1 - o[z_d]) w_{ho} o[u_d] (1 - o[u_d]) x_i \\ &= \eta \sum_{d \in D} \delta_o o[z_d] (1 - o[z_d]) w_{ho} o[u_d] (1 - o[u_d]) x_i \\ &= \eta \sum_{d \in D} \delta_h o[u_d] (1 - o[u_d]) x_i \\ &= \eta \sum_{d \in D} \delta_h \frac{df_h[e]}{de} x_i \end{aligned}$$

$$\blacksquare \Delta w_{ih} = \eta \sum_{d \in D} \delta_h \frac{df_h[e]}{de} x_i$$

$$\therefore x_o = o[z_d] = \frac{1}{1 + e^{-(z_d)}}$$

$$z_d = \sum_h w_{h o} x_h$$

$$\therefore x_h = o[u_d] = \frac{1}{1 + e^{-(u_d)}}$$

$$u_d = \sum_i w_{i h} x_i$$

$$\frac{\partial z_d}{\partial w_{i h}} = \frac{\partial z_d}{\partial x_h} \frac{\partial x_h}{\partial w_{i h}} = \frac{\partial z_d}{\partial x_h} \frac{\partial x_h}{\partial u_d} \frac{\partial u_d}{\partial w_{i h}}$$

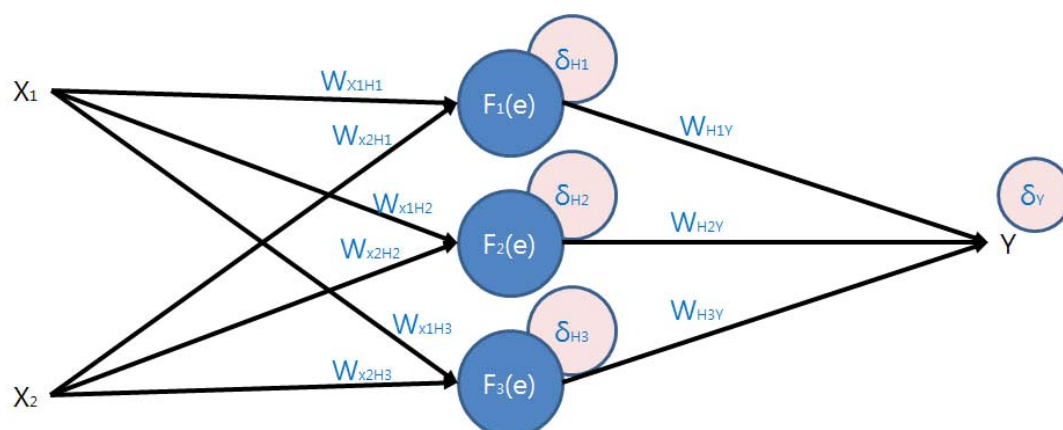
$$\frac{\partial z_d}{\partial x_h} = \frac{\partial \sum_h w_{h o} x_h}{\partial x_h} = w_{h o}$$

$$\frac{\partial x_h}{\partial u_d} = \frac{\partial o[u_d]}{\partial u_d} = o[u_d] (1 - o[u_d])$$

$$\frac{\partial u_d}{\partial w_{i h}} = \frac{\partial \sum_i w_{i h} x_i}{\partial w_{i h}} = x_i$$

$$\therefore \frac{\partial z_d}{\partial w_{i h}} = \frac{\partial z_d}{\partial x_h} \frac{\partial x_h}{\partial u_d} \frac{\partial u_d}{\partial w_{i h}} = w_{h o} o[u_d] (1 - o[u_d]) x_i$$

Example



$$y = o[w_{h1 y} x_{h1} + w_{h2 y} x_{h2} + w_{h3 y} x_{h3}]$$

$$z = w_{h1 y} x_{h1} + w_{h2 y} x_{h2} + w_{h3 y} x_{h3}$$

$$\delta_y = t - y = t - o[z]$$

$$x_{h1} = o[w_{x1 h1} x_1 + w_{x2 h1} x_2]$$

$$u_1 = w_{x1 h1} x_1 + w_{x2 h1} x_2$$

$$\delta_{h1} = w_{h1 y} \delta_y$$

$$x_{h2} = o[w_{x1 h2} x_1 + w_{x2 h2} x_2]$$

$$u_2 = w_{x1 h2} x_1 + w_{x2 h2} x_2$$

$$\delta_{h2} = w_{h2 y} \delta_y$$

$$x_{h3} = o[w_{x1 h3} x_1 + w_{x2 h3} x_2]$$

$$u_3 = w_{x1\ h3} x_1 + w_{x2\ h3} x_2$$

$$\delta_{h3} = w_{h3\ y} \delta_y$$

$$\blacksquare x_k \rightarrow w_{k\ h} \rightarrow x_h$$

$$\delta_h \rightarrow w_{h\ o} \rightarrow \delta_o$$

$$\therefore x_h = o \left[\sum_k x_k w_{k\ h} \right]$$

$$\therefore \delta_h = \sum_k w_{h\ o} \delta_o;$$

$$\blacksquare y$$

$$y = o[w_{h1\ y} x_{h1} + w_{h2\ y} x_{h2} + w_{h3\ y} x_{h3}] = \frac{1}{1 + e^{-z_d}}$$

$$z_d = w_{h1\ y} x_{h1} + w_{h2\ y} x_{h2} + w_{h3\ y} x_{h3}$$

$$\delta = t - y = t - o[z]$$

$$\begin{aligned} w_{h1\ y} &= w_{h1\ y} - \eta \frac{\partial E}{\partial w_{h1\ y}} \\ &= w_{h1\ y} + \eta (t - o[z]) \frac{d o[z]}{dz} \frac{\partial z_d}{\partial w_{h1\ y}} \\ &= w_{h1\ y} + \eta \delta o[z] (1 - o[z]) \frac{\partial z_d}{\partial w_{h1\ y}} \\ &= w_{h1\ y} + \eta \delta_y x_{h1} \end{aligned}$$

$$\therefore w_{h1\ y} = w_{h1\ y} + \eta \delta_y x_{h1}$$

$$\therefore w_{h2\ y} = w_{h2\ y} + \eta \delta_y x_{h2}$$

$$\therefore w_{h3\ y} = w_{h3\ y} + \eta \delta_y x_{h3}$$

$$\blacksquare x_{h1}$$

$$y = o_d = o[z_d] = o[w_{h1\ y} x_{h1} + w_{h2\ y} x_{h2} + w_{h3\ y} x_{h3}] = \frac{1}{1 + e^{-z}}$$

$$z_d = w_{h1\ y} x_{h1} + w_{h2\ y} x_{h2} + w_{h3\ y} x_{h3}$$

$$z_d = w_{h1\ y} o[w_{x1\ h1} x_1 + w_{x2\ h1} x_2] + w_{h2\ y} o[w_{x1\ h2} x_1 + w_{x2\ h2} x_2] + w_{h3\ y} o[w_{x1\ h3} x_1 + w_{x2\ h3} x_2]$$

$$x_{h1} = o[w_{x1\ h1} x_1 + w_{x2\ h1} x_2] = \frac{1}{1 + e^{-u_1}}$$

$$u_1 = w_{x1\ h1} x_1 + w_{x2\ h1} x_2$$

$$\delta_{h1} = w_{h1\ y} \delta_y$$

$$\begin{aligned}
\frac{\partial E}{\partial w_{x1\ h1}} &= - \sum (t_d - o_d) \frac{\partial o_d}{\partial w_{x1\ h1}} \\
&= - \sum (t_d - o_d) \frac{\partial o_d}{\partial z_d} \frac{\partial z_d}{\partial w_{x1\ h1}} \\
&= - \sum (t_d - o_d) o_d (1 - o_d) \sum \frac{\partial z_d}{\partial x_{h1}} \frac{\partial x_{h1}}{\partial w_{x1\ h1}} \\
&= -\delta_y \sum \frac{\partial z_d}{\partial x_{h1}} \frac{\partial x_{h1}}{\partial w_{x1\ h1}} \\
&= -\delta_y \sum w_{h1\ y} \frac{\partial x_{hi}}{\partial w_{x1\ h1}} \\
&= -\delta_y \sum w_{h1\ y} \frac{\partial x_{hi}}{\partial u_1} \frac{\partial u_1}{\partial w_{x1\ h1}} \\
&= -\delta_y \sum w_{h1\ y} o_h (1 - o_h) \frac{\partial u_1}{\partial w_{x1\ h1}} \\
&= -\delta_y \sum w_{h1\ y} o_h (1 - o_h) x_1 \\
&= -o_h (1 - o_h) \sum \delta_y w_{h1\ y} x_1 \\
&= -\delta_{h1} x_1
\end{aligned}$$

$$\delta_h \leftarrow o_h (1 - o_h) \sum w_{h\ k} \delta_k$$

$$\begin{aligned}
w_{x1\ h1} &= w_{x1\ h1} - \eta \frac{\partial E}{\partial w_{x1\ h1}} \\
&= w_{x1\ h1} + \eta o_{h1} (1 - o_{h1}) \sum w_{h1\ y} \delta_y x_1 \\
&= w_{x1\ h1} + \eta \delta_{h1} x_1
\end{aligned}$$

$$\therefore w_{x1\ h1} = w_{x1\ h1} + \eta \delta_{h1} x_1$$

$$\therefore w_{x2\ h1} = w_{x2\ h1} + \eta \delta_{h1} x_2$$

$$\therefore w_{x1\ h2} = w_{x1\ h2} + \eta \delta_{h2} x_1$$

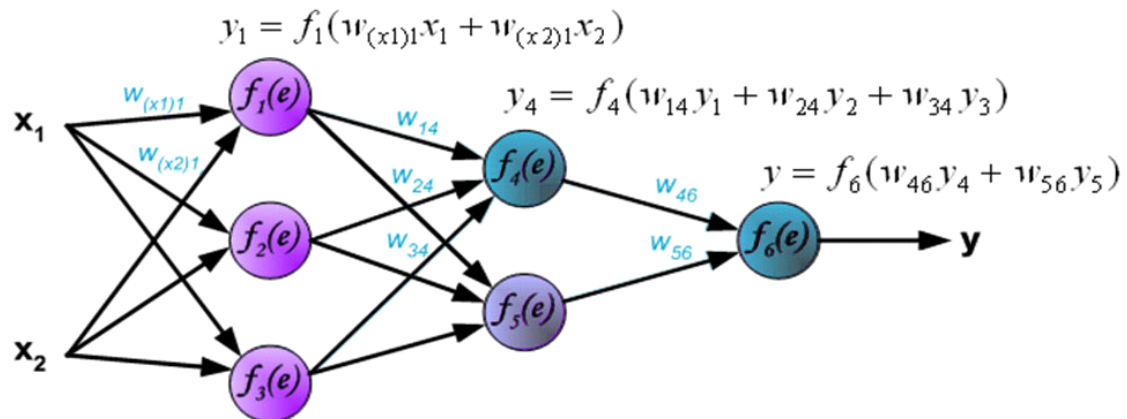
$$\therefore w_{x2\ h2} = w_{x2\ h2} + \eta \delta_{h2} x_2$$

$$\therefore w_{x1\ h3} = w_{x1\ h3} + \eta \delta_{h3} x_1$$

$$\therefore w_{x2\ h3} = w_{x2\ h3} + \eta \delta_{h3} x_2$$

Learning Algorithm: Backpropagation

$$\square \therefore x_h = 0 \left[\sum_k x_k w_{kh} \right]$$



$$y_1 = f_1[e] = f_1[w_{(x1)1}x_1 + w_{(x2)1}x_2]$$

$$y_2 = f_2[w_{(x1)2}x_1 + w_{(x2)2}x_2]$$

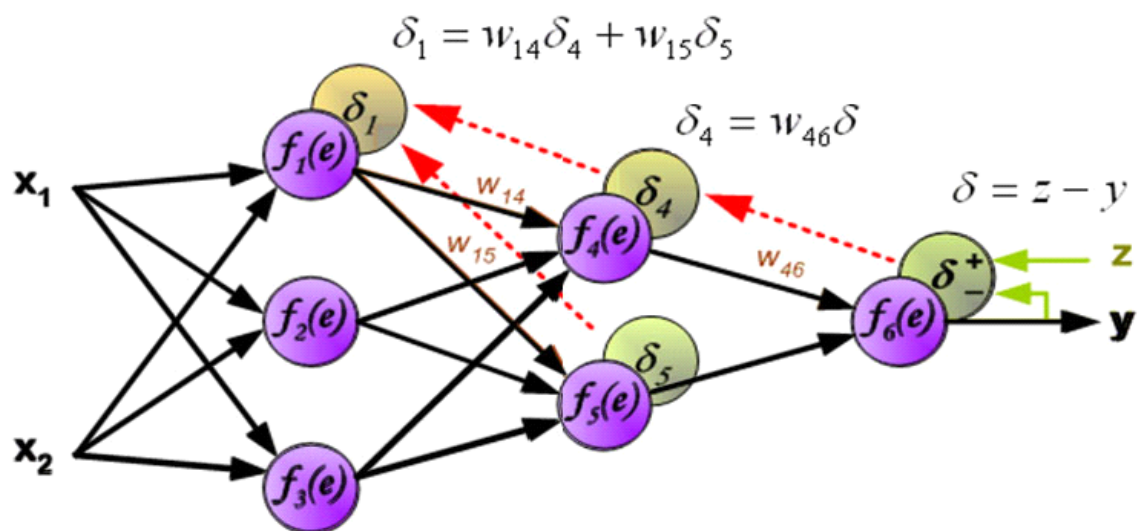
$$y_3 = f_3[w_{(x1)3}x_1 + w_{(x2)3}x_2]$$

$$y_4 = f_4[w_{14}y_1 + w_{24}y_2 + w_{34}y_3]$$

$$y_5 = f_5[w_{15}y_1 + w_{25}y_2 + w_{35}y_3]$$

$$y = f_6[w_{46}y_4 + w_{56}y_5]$$

$$\square \therefore \delta_h = \sum_o w_{ho} \delta_o$$



$$\delta = z - y$$

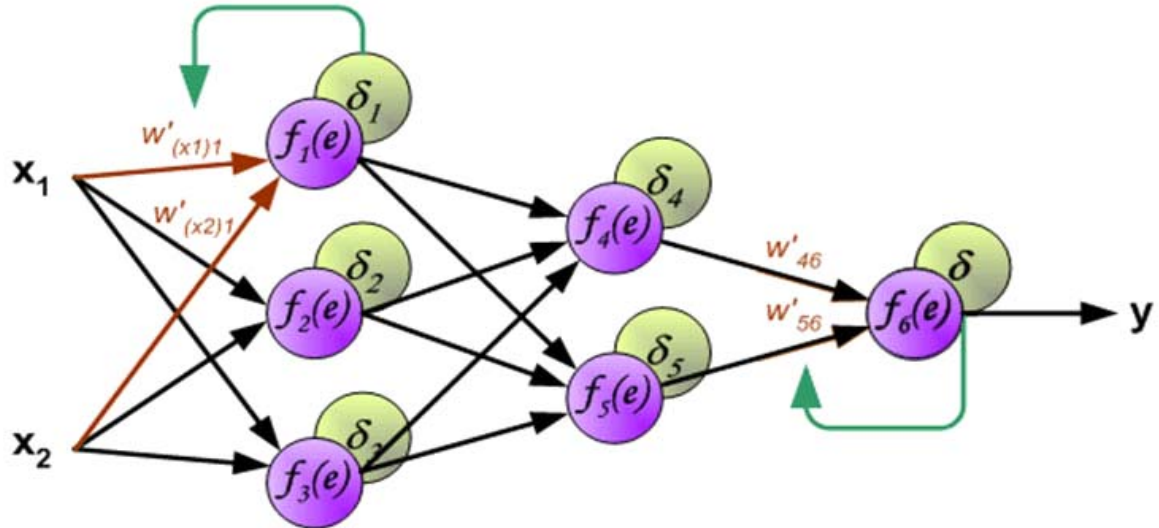
$$\delta_4 = w_{46} \delta$$

$$\delta_5 = w_{56} \delta$$

$$\delta_1 = w_{14} \delta_4 + w_{15} \delta_5$$

$$\delta_2 = w_{24} \delta_4 + w_{25} \delta_5$$

$$\delta_3 = w_{34} \delta_4 + w_{35} \delta_5$$



$$w'_{(x1)1} = w_{(x1)1} + \eta \delta_1 \frac{\partial f_1[e]}{\partial e} x_1$$

$$w'_{(x2)1} = w_{(x2)1} + \eta \delta_1 \frac{\partial f_1[e]}{\partial e} x_2$$

$$w'_{(x1)2} = w_{(x1)2} + \eta \delta_2 \frac{\partial f_2[e]}{\partial e} x_1$$

$$w'_{(x2)2} = w_{(x2)2} + \eta \delta_2 \frac{\partial f_2[e]}{\partial e} x_2$$

$$w'_{(x1)3} = w_{(x1)3} + \eta \delta_3 \frac{\partial f_3[e]}{\partial e} x_1$$

$$w'_{(x2)3} = w_{(x2)3} + \eta \delta_3 \frac{\partial f_3[e]}{\partial e} x_2$$

$$w'_{14} = w_{14} + \eta \delta_4 \frac{\partial f_4[e]}{\partial e} y_1$$

$$w'_{24} = w_{24} + \eta \delta_4 \frac{\partial f_4[e]}{\partial e} y_2$$

$$w'_{34} = w_{34} + \eta \delta_4 \frac{\partial f_4[e]}{\partial e} y_3$$

$$w'_{15} = w_{15} + \eta \delta_5 \frac{\partial f_5[e]}{\partial e} y_1$$

$$w'_{25} = w_{25} + \eta \delta_5 \frac{\partial f_5[e]}{\partial e} y_2$$

$$w'_{35} = w_{35} + \eta \delta_5 \frac{\partial f_5[e]}{\partial e} y_3$$

$$w'_{46} = w_{46} + \eta \delta \frac{df_6[e]}{de} y_4$$

$$w'_{56} = w_{56} + \eta \delta \frac{df_6[e]}{de} y_5$$

$$\therefore w_{h1y} = w_{h1y} + \eta \sum_d \delta_y x_{h1}$$

$$\therefore w_{h2y} = w_{h2y} + \eta \sum_d \delta_y x_{h2}$$

$$\therefore w_{h3y} = w_{h3y} + \eta \sum_d \delta_y x_{h3}$$

$$\therefore w_{x1h1} = w_{x1h1} + \eta \sum_d \delta_{h1} x_1$$

$$\therefore w_{x2h1} = w_{x2h1} + \eta \sum_d \delta_{h1} x_2$$

$$\therefore w_{x1h2} = w_{x1h2} + \eta \sum_d \delta_{h2} x_1$$

$$\therefore w_{x2h2} = w_{x2h2} + \eta \sum_d \delta_{h2} x_2$$

$$\therefore w_{x1h3} = w_{x1h3} + \eta \sum_d \delta_{h3_d} x_{1_d}$$

$$\therefore w_{x2h3} = w_{x2h3} + \eta \sum_d \delta_{h3_d} x_{2_d}$$

MATLAB

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} H1 \\ H2 \\ H3 \end{pmatrix} \rightarrow (Y)$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{w_{xihj}} \begin{pmatrix} H1 \\ H2 \\ H3 \end{pmatrix} \xrightarrow{w_{hjiy}} (Y)$$

■ MATLAB

```

hElt = xIn * wh
hOut = 1. / (1 + Exp[-hElt])

oElt = hOut * wo
oOut = 1. / (1 + Exp[-oElt])

deltaOut = oOut * (1 - oOut) * (tOut - oOut)
deltaHid = hOut * (1 - hOut) * (deltaOut * wo')

wo = wo + η * hOut'.deltaOut
wh = wh + η * xIn'.deltaHid

SSE = 0.5 * Sqrt[(tOut - oOut)'.(tOut - oOut)]

```

■ matrix

$$wh = \begin{pmatrix} w_{x1\ h1} & w_{x1\ h2} & w_{x1\ h3} \\ w_{x2\ h1} & w_{x2\ h2} & w_{x2\ h3} \end{pmatrix};$$

$$wo = \begin{pmatrix} w_{h1\ y} \\ w_{h2\ y} \\ w_{h3\ y} \end{pmatrix};$$

$$xIn = \begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \\ x_{14} & x_{24} \end{pmatrix};$$

$$tOut = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{pmatrix};$$

■ hElt=xIn.wh

$$\begin{aligned}
hElt = xIn.wh &= \begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \\ x_{14} & x_{24} \end{pmatrix} \cdot \begin{pmatrix} w_{x1\ h1} & w_{x1\ h2} & w_{x1\ h3} \\ w_{x2\ h1} & w_{x2\ h2} & w_{x2\ h3} \end{pmatrix} \\
&= \begin{pmatrix} w_{x1\ h1} x_{11} + w_{x2\ h1} x_{21} & w_{x1\ h2} x_{11} + w_{x2\ h2} x_{21} & w_{x1\ h3} x_{11} + w_{x2\ h3} x_{21} \\ w_{x1\ h1} x_{12} + w_{x2\ h1} x_{22} & w_{x1\ h2} x_{12} + w_{x2\ h2} x_{22} & w_{x1\ h3} x_{12} + w_{x2\ h3} x_{22} \\ w_{x1\ h1} x_{13} + w_{x2\ h1} x_{23} & w_{x1\ h2} x_{13} + w_{x2\ h2} x_{23} & w_{x1\ h3} x_{13} + w_{x2\ h3} x_{23} \\ w_{x1\ h1} x_{14} + w_{x2\ h1} x_{24} & w_{x1\ h2} x_{14} + w_{x2\ h2} x_{24} & w_{x1\ h3} x_{14} + w_{x2\ h3} x_{24} \end{pmatrix}; \\
hOut &= \frac{1}{1 + e^{-hElt}} = \begin{pmatrix} x_{h1_1} & x_{h2_1} & x_{h3_1} \\ x_{h1_2} & x_{h2_2} & x_{h3_2} \\ x_{h1_3} & x_{h2_3} & x_{h3_3} \\ x_{h1_4} & x_{h2_4} & x_{h3_4} \end{pmatrix};
\end{aligned}$$

■ **oElt=hOut.wo**

$$\begin{aligned} \text{oElt} = \text{hOut.wo} &= \begin{pmatrix} x_{h1_1} & x_{h2_1} & x_{h3_1} \\ x_{h1_2} & x_{h2_2} & x_{h3_2} \\ x_{h1_3} & x_{h2_3} & x_{h3_3} \\ x_{h1_4} & x_{h2_4} & x_{h3_4} \end{pmatrix} \cdot \begin{pmatrix} w_{h1y} \\ w_{h2y} \\ w_{h3y} \end{pmatrix} \\ &= \begin{pmatrix} w_{h1y} x_{h1_1} + w_{h2y} x_{h2_1} + w_{h3y} x_{h3_1} \\ w_{h1y} x_{h1_2} + w_{h2y} x_{h2_2} + w_{h3y} x_{h3_2} \\ w_{h1y} x_{h1_3} + w_{h2y} x_{h2_3} + w_{h3y} x_{h3_3} \\ w_{h1y} x_{h1_4} + w_{h2y} x_{h2_4} + w_{h3y} x_{h3_4} \end{pmatrix}; \\ \text{oOut} &= \frac{1}{1 + e^{-\text{oElt}}} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}; \end{aligned}$$

■ **δOut=oOut*(1-oOut)*(tOut-oOut)**

$$\begin{aligned} \delta \text{Out} &= \text{oOut} * (1 - \text{oOut}) * (\text{tOut} - \text{oOut}) \\ &= \begin{pmatrix} Y_1 (1 - Y_1) (t_1 - Y_1) \\ Y_2 (1 - Y_2) (t_2 - Y_2) \\ Y_3 (1 - Y_3) (t_3 - Y_3) \\ Y_4 (1 - Y_4) (t_4 - Y_4) \end{pmatrix} \\ &= \begin{pmatrix} \delta_{Y_1} \\ \delta_{Y_2} \\ \delta_{Y_3} \\ \delta_{Y_4} \end{pmatrix}; \end{aligned}$$

$$\delta_k = o_k (1 - o_k) (t_k - o_k);$$

■ **δHid=hOut*(1-hOut)*(δOut.wo^T)**

$$\begin{aligned} \delta \text{Hid} &= \text{hOut} * (1 - \text{hOut}) * (\delta \text{Out.wo}^T) \\ &= \begin{pmatrix} x_{h1_1} (1 - x_{h1_1}) & x_{h2_1} (1 - x_{h2_1}) & x_{h3_1} (1 - x_{h3_1}) \\ x_{h1_2} (1 - x_{h1_2}) & x_{h2_2} (1 - x_{h2_2}) & x_{h3_2} (1 - x_{h3_2}) \\ x_{h1_3} (1 - x_{h1_3}) & x_{h2_3} (1 - x_{h2_3}) & x_{h3_3} (1 - x_{h3_3}) \\ x_{h1_4} (1 - x_{h1_4}) & x_{h2_4} (1 - x_{h2_4}) & x_{h3_4} (1 - x_{h3_4}) \end{pmatrix} * \begin{pmatrix} \delta_{Y_1} \\ \delta_{Y_2} \\ \delta_{Y_3} \\ \delta_{Y_4} \end{pmatrix} \cdot (w_{h1y} \quad w_{h2y} \quad w_{h3y}) \\ &= \begin{pmatrix} x_{h1_1} (1 - x_{h1_1}) & x_{h2_1} (1 - x_{h2_1}) & x_{h3_1} (1 - x_{h3_1}) \\ x_{h1_2} (1 - x_{h1_2}) & x_{h2_2} (1 - x_{h2_2}) & x_{h3_2} (1 - x_{h3_2}) \\ x_{h1_3} (1 - x_{h1_3}) & x_{h2_3} (1 - x_{h2_3}) & x_{h3_3} (1 - x_{h3_3}) \\ x_{h1_4} (1 - x_{h1_4}) & x_{h2_4} (1 - x_{h2_4}) & x_{h3_4} (1 - x_{h3_4}) \end{pmatrix} * \begin{pmatrix} w_{h1y} \delta_{Y_1} & w_{h2y} \delta_{Y_1} & w_{h3y} \delta_{Y_1} \\ w_{h1y} \delta_{Y_2} & w_{h2y} \delta_{Y_2} & w_{h3y} \delta_{Y_2} \\ w_{h1y} \delta_{Y_3} & w_{h2y} \delta_{Y_3} & w_{h3y} \delta_{Y_3} \\ w_{h1y} \delta_{Y_4} & w_{h2y} \delta_{Y_4} & w_{h3y} \delta_{Y_4} \end{pmatrix} \\ &= \begin{pmatrix} x_{h1_1} (1 - x_{h1_1}) w_{h1y} \delta_{Y_1} & x_{h2_1} (1 - x_{h2_1}) w_{h2y} \delta_{Y_1} & x_{h3_1} (1 - x_{h3_1}) w_{h3y} \delta_{Y_1} \\ x_{h1_2} (1 - x_{h1_2}) w_{h1y} \delta_{Y_2} & x_{h2_2} (1 - x_{h2_2}) w_{h2y} \delta_{Y_2} & x_{h3_2} (1 - x_{h3_2}) w_{h3y} \delta_{Y_2} \\ x_{h1_3} (1 - x_{h1_3}) w_{h1y} \delta_{Y_3} & x_{h2_3} (1 - x_{h2_3}) w_{h2y} \delta_{Y_3} & x_{h3_3} (1 - x_{h3_3}) w_{h3y} \delta_{Y_3} \\ x_{h1_4} (1 - x_{h1_4}) w_{h1y} \delta_{Y_4} & x_{h2_4} (1 - x_{h2_4}) w_{h2y} \delta_{Y_4} & x_{h3_4} (1 - x_{h3_4}) w_{h3y} \delta_{Y_4} \end{pmatrix} \\ &= \begin{pmatrix} \delta_{h1_1} & \delta_{h2_1} & \delta_{h3_1} \\ \delta_{h1_2} & \delta_{h2_2} & \delta_{h3_2} \\ \delta_{h1_3} & \delta_{h2_3} & \delta_{h3_3} \\ \delta_{h1_4} & \delta_{h2_4} & \delta_{h3_4} \end{pmatrix} \end{aligned}$$

$$\delta_{h2_3} = o_{h2_3} (1 - o_{h2_3}) w_{h2y} \delta_{Y_3};$$

$$\delta_h = \sum_{k \in \text{outputs}} w_{h k} \delta_k;$$

$$\delta_h = o_h (1 - o_h) \sum_{k \in \text{outputs}} w_{h k} \delta_k;$$

$$\blacksquare \text{ wo} = \text{wo} + \eta \text{ hOut}^T \delta \text{Out};$$

$$\text{wo} = \begin{pmatrix} w_{h1 y} \\ w_{h2 y} \\ w_{h3 y} \end{pmatrix};$$

$$\begin{aligned} \begin{pmatrix} w_{h1 y} \\ w_{h2 y} \\ w_{h3 y} \end{pmatrix} &= \begin{pmatrix} w_{h1 y} \\ w_{h2 y} \\ w_{h3 y} \end{pmatrix} + \eta * \begin{pmatrix} x_{h1_1} & x_{h1_2} & x_{h1_3} & x_{h1_4} \\ x_{h2_1} & x_{h2_2} & x_{h2_3} & x_{h2_4} \\ x_{h3_1} & x_{h3_2} & x_{h3_3} & x_{h3_4} \end{pmatrix} \cdot \begin{pmatrix} \delta_{y_1} \\ \delta_{y_2} \\ \delta_{y_3} \\ \delta_{y_4} \end{pmatrix} \\ &= \begin{pmatrix} w_{h1 y} \\ w_{h2 y} \\ w_{h3 y} \end{pmatrix} + \eta * \begin{pmatrix} x_{h1_1} \delta_{y_1} + x_{h1_2} \delta_{y_2} + x_{h1_3} \delta_{y_3} + x_{h1_4} \delta_{y_4} \\ x_{h2_1} \delta_{y_1} + x_{h2_2} \delta_{y_2} + x_{h2_3} \delta_{y_3} + x_{h2_4} \delta_{y_4} \\ x_{h3_1} \delta_{y_1} + x_{h3_2} \delta_{y_2} + x_{h3_3} \delta_{y_3} + x_{h3_4} \delta_{y_4} \end{pmatrix} \\ &= \begin{pmatrix} w_{h1 y} \\ w_{h2 y} \\ w_{h3 y} \end{pmatrix} + \eta * \begin{pmatrix} \sum_d x_{h1_d} \delta_{y_d} \\ \sum_d x_{h2_d} \delta_{y_d} \\ \sum_d x_{h3_d} \delta_{y_d} \end{pmatrix} \end{aligned}$$

$$\blacksquare \text{ wh} = \text{wh} + \eta \text{ xIn}^T \delta \text{Hid};$$

$$\text{wh} = \begin{pmatrix} w_{x1 h1} & w_{x1 h2} & w_{x1 h3} \\ w_{x2 h1} & w_{x2 h2} & w_{x2 h3} \end{pmatrix};$$

$$\begin{aligned} \begin{pmatrix} w_{x1 h1} & w_{x1 h2} & w_{x1 h3} \\ w_{x2 h1} & w_{x2 h2} & w_{x2 h3} \end{pmatrix} &= \begin{pmatrix} w_{x1 h1} & w_{x1 h2} & w_{x1 h3} \\ w_{x2 h1} & w_{x2 h2} & w_{x2 h3} \end{pmatrix} + \eta * \begin{pmatrix} x_{1_1} & x_{1_2} & x_{1_3} & x_{1_4} \\ x_{2_1} & x_{2_2} & x_{2_3} & x_{2_4} \end{pmatrix} \cdot \begin{pmatrix} \delta_{h1_1} & \delta_{h2_1} & \delta_{h3_1} \\ \delta_{h1_2} & \delta_{h2_2} & \delta_{h3_2} \\ \delta_{h1_3} & \delta_{h2_3} & \delta_{h3_3} \\ \delta_{h1_4} & \delta_{h2_4} & \delta_{h3_4} \end{pmatrix} \\ &= \begin{pmatrix} w_{x1 h1} & w_{x1 h2} & w_{x1 h3} \\ w_{x2 h1} & w_{x2 h2} & w_{x2 h3} \end{pmatrix} + \eta \begin{pmatrix} x_{1_1} \delta_{h1_1} + x_{1_2} \delta_{h1_2} + x_{1_3} \delta_{h1_3} + x_{1_4} \delta_{h1_4} & x_{1_1} \delta_{h2_1} + x_{1_2} \delta_{h2_2} + x_{1_3} \delta_{h2_3} + x_{1_4} \delta_{h2_4} & x_{1_1} \delta_{h3_1} + x_{1_2} \delta_{h3_2} + x_{1_3} \delta_{h3_3} + x_{1_4} \delta_{h3_4} \\ x_{2_1} \delta_{h1_1} + x_{2_2} \delta_{h1_2} + x_{2_3} \delta_{h1_3} + x_{2_4} \delta_{h1_4} & x_{2_1} \delta_{h2_1} + x_{2_2} \delta_{h2_2} + x_{2_3} \delta_{h2_3} + x_{2_4} \delta_{h2_4} & x_{2_1} \delta_{h3_1} + x_{2_2} \delta_{h3_2} + x_{2_3} \delta_{h3_3} + x_{2_4} \delta_{h3_4} \end{pmatrix} \\ &= \begin{pmatrix} w_{x1 h1} & w_{x1 h2} & w_{x1 h3} \\ w_{x2 h1} & w_{x2 h2} & w_{x2 h3} \end{pmatrix} + \eta \begin{pmatrix} \sum_d x_{1_d} \delta_{h1_d} & \sum_d x_{1_d} \delta_{h2_d} & \sum_d x_{1_d} \delta_{h3_d} \\ \sum_d x_{2_d} \delta_{h1_d} & \sum_d x_{2_d} \delta_{h2_d} & \sum_d x_{2_d} \delta_{h3_d} \end{pmatrix} \end{aligned}$$