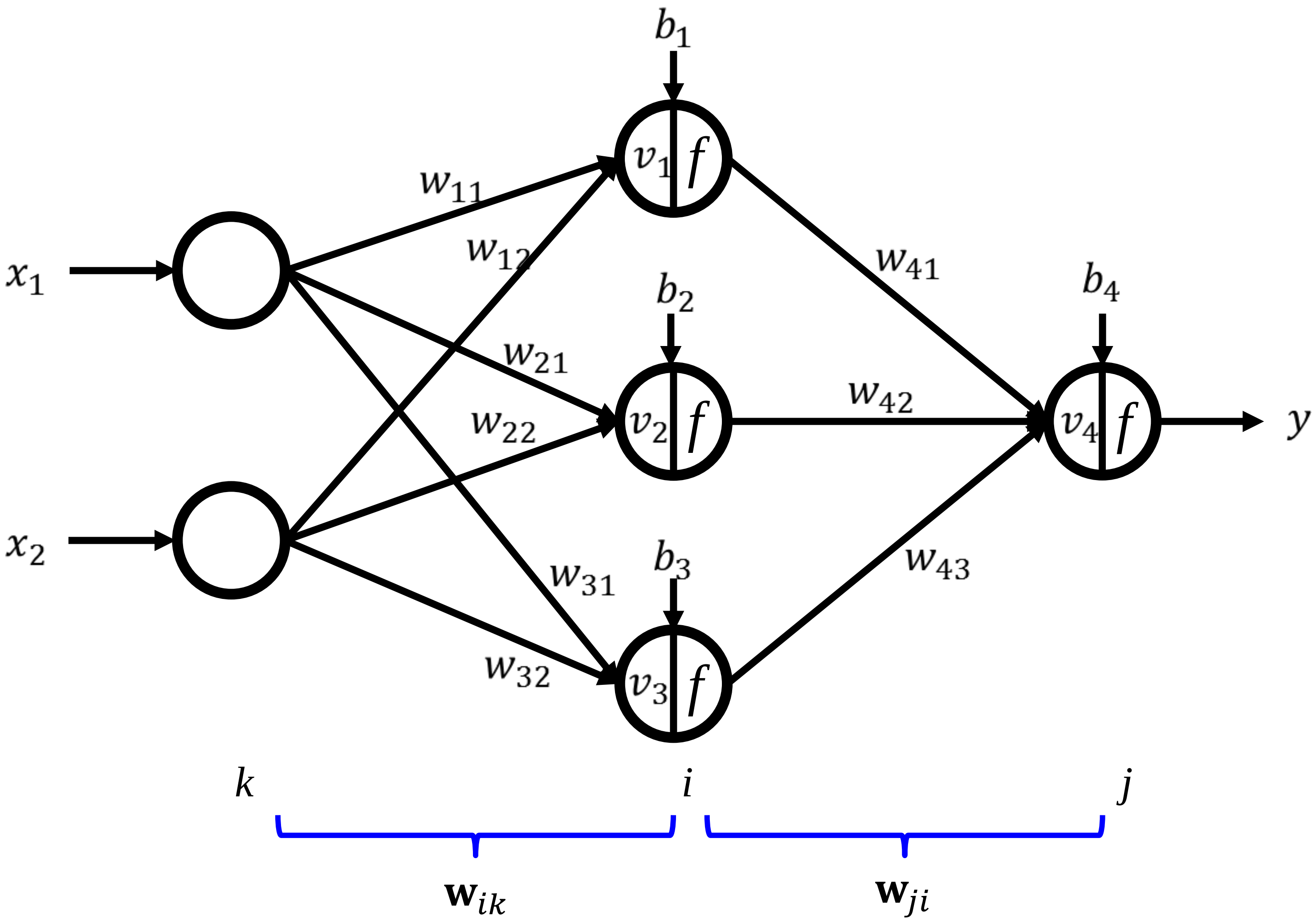
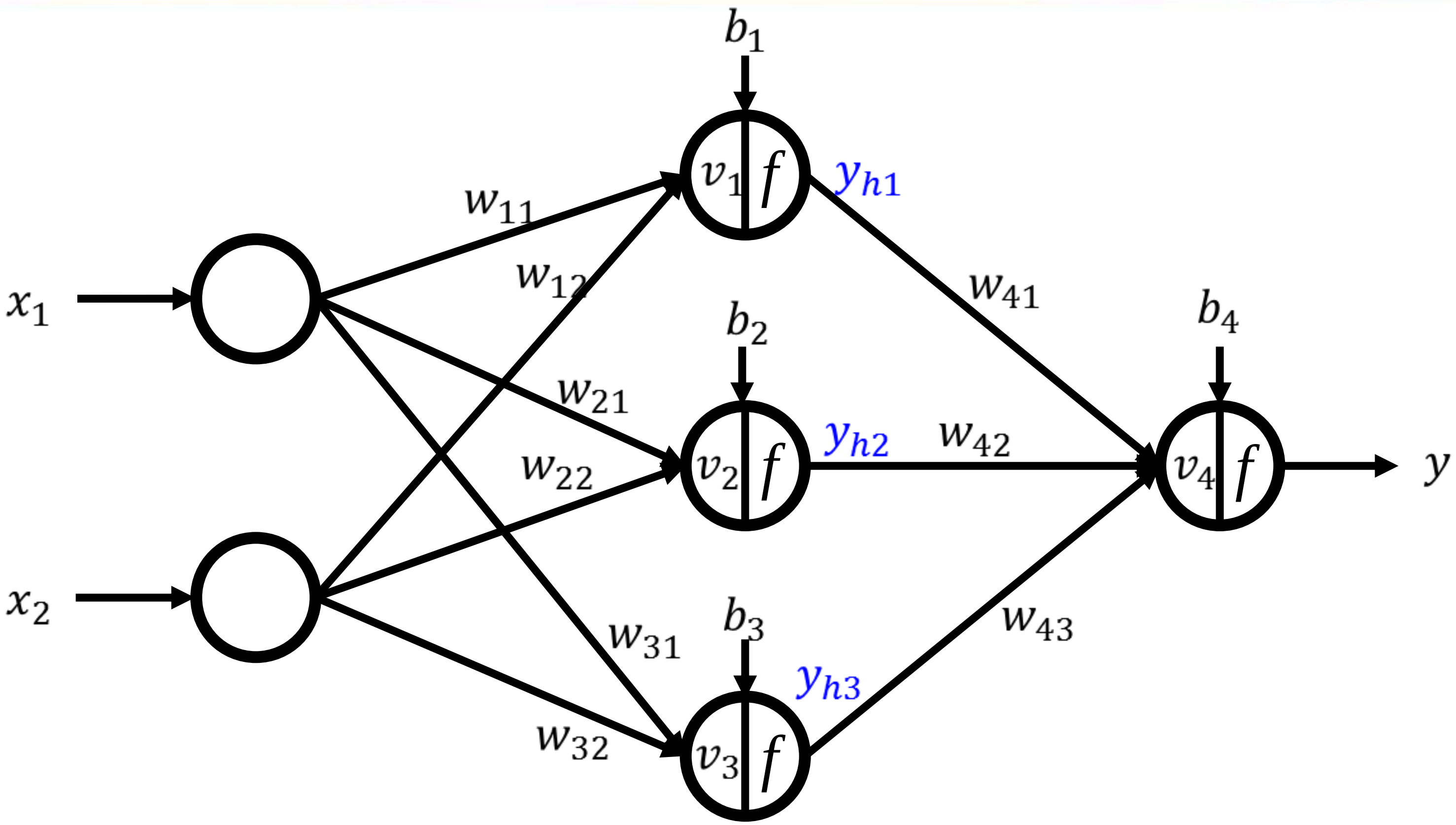
An Example of Derivations for Backpropagation Learning Rules

Jeen-Shing Wang
Department of Electrical Engineering
National Cheng Kung University
October 5, 2016

2-Layered Feedforward Neural Network



Forward Path



$$v_k = \sum_{k=1\sim3,i=1\sim2} w_{ki}x_i + b_k$$

$$y_{hk} = f(v_k)$$

$$v_4 = w_{41}f(v_1) + w_{42}f(v_2) + w_{43}f(v_3) + b_4$$

$$v_4 = w_{41}y_{h1} + w_{42}y_{h2} + w_{43}y_{h3} + b_4$$

$$y = f(v_4) = v_4$$

Derivative of Sigmoid

$$s(z) = \frac{1}{1 + e^{-z}}$$

$$ds(z)/_{dz} = \left(\frac{1}{1 + e^{-z}}\right)^2 \frac{d}{dz}(1 + e^{-z})$$

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

$$=\left(\frac{1}{1+e^{-z}}\right)\left(\frac{1}{1+e^{-z}}\right) -e^{-x}$$

$$=\left(\frac{1}{1+e^{-z}}\right)\left(\frac{-e^{-z}}{1+e^{-z}}\right)$$

$$=s(z)(1 - s(z))$$

Derivative of Hyperbolic Tangent

$$
\begin{aligned}
g_{\tanh}(z) &= \frac{\sinh(z)}{\cosh(z)} \\
&= \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}
\end{aligned}
$$

$$
\begin{aligned}
g'_{\tanh}(z) &= \frac{\partial}{\partial z}\frac{\sinh(z)}{\cosh(z)} \\
&= \frac{\frac{\partial}{\partial z}\sinh(z)\times\cosh(z)-\frac{\partial}{\partial z}\cosh(z)\times\sinh(z)}{\cosh^2(z)} \\
&= \frac{\cosh^2(z)-\sinh^2(z)}{\cosh^2(z)} \\
&= 1-\frac{\sinh^2(z)}{\cosh^2(z)} \\
&= 1-\tanh^2(z)
\end{aligned}
$$

Some Useful Derivatives

$$\frac{d}{dx}(e^x) = e^x$$

$$\frac{d}{dx}(a^x) = a^x \cdot \ln a$$

$$\frac{d}{dx}(\ln x) = \frac{1}{x}, \qquad x > 0$$

$$\frac{d}{dx}(\ln|x|) = \frac{1}{x}, \qquad x \neq 0$$

$$\frac{d}{dx}(\log_a x) = \frac{1}{x \cdot \ln a}, \qquad x > 0$$

$$\frac{d}{dx}(\sinh x) = \cosh x$$

$$\frac{d}{dx}(\cosh x) = \sinh x$$

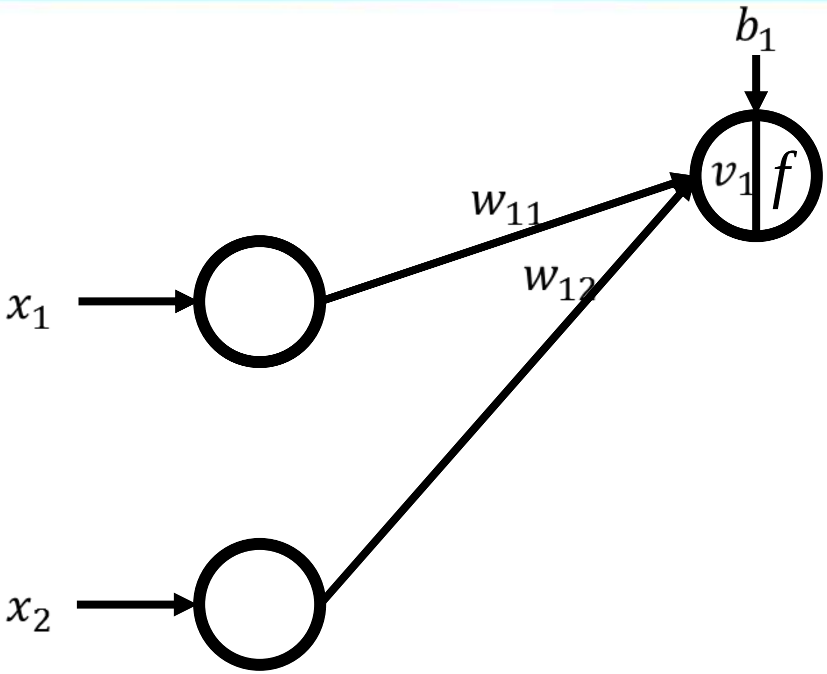
$$\frac{d}{dx}(\tanh x) = \operatorname{sech}^2 x$$

$$\frac{d}{dx}(\coth x) = -\operatorname{csch}^2 x$$

$$\frac{d}{dx}(\operatorname{sech} x) = -\operatorname{sech} x \tanh x$$

$$\frac{d}{dx}(\operatorname{csch} x) = -\operatorname{csch} x \coth x$$

Forward Path

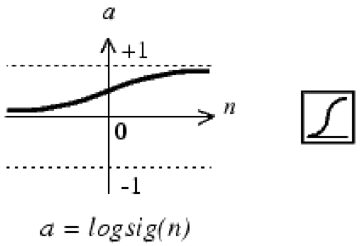


$$\boldsymbol{v_1 = w_{11}x_1 + w_{12}x_2} + b_1$$

Activation function:

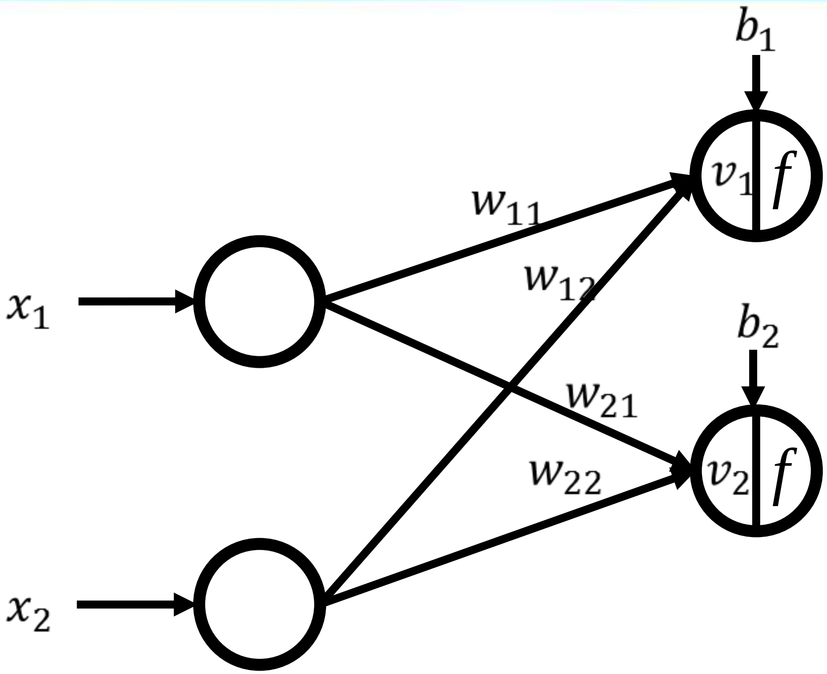
$$f(x) = \frac{1}{1+e^{-x}} \qquad f(v_1) = \frac{1}{1+e^{-v_1}}$$

$$f'(x) = f(x)\big(1 - f(x)\big) \qquad f'(v_1) = f(v_1)\big(1 - f(v_1)\big)$$



Log-Sigmoid Transfer Function

Forward Path



$$v_1 = w_{11}x_1 + w_{12}x_2 + b_1$$

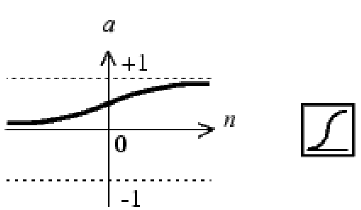
$$y_{h1} = f(v_1) = \frac{1}{1+e^{-v_1}}$$

$$\boldsymbol{v_2 = w_{21}x_1 + w_{22}x_2} + b_2$$

Activation function:

$$f(x) = \frac{1}{1+e^{-x}} \qquad f(v_2) = \frac{1}{1+e^{-v_2}}$$

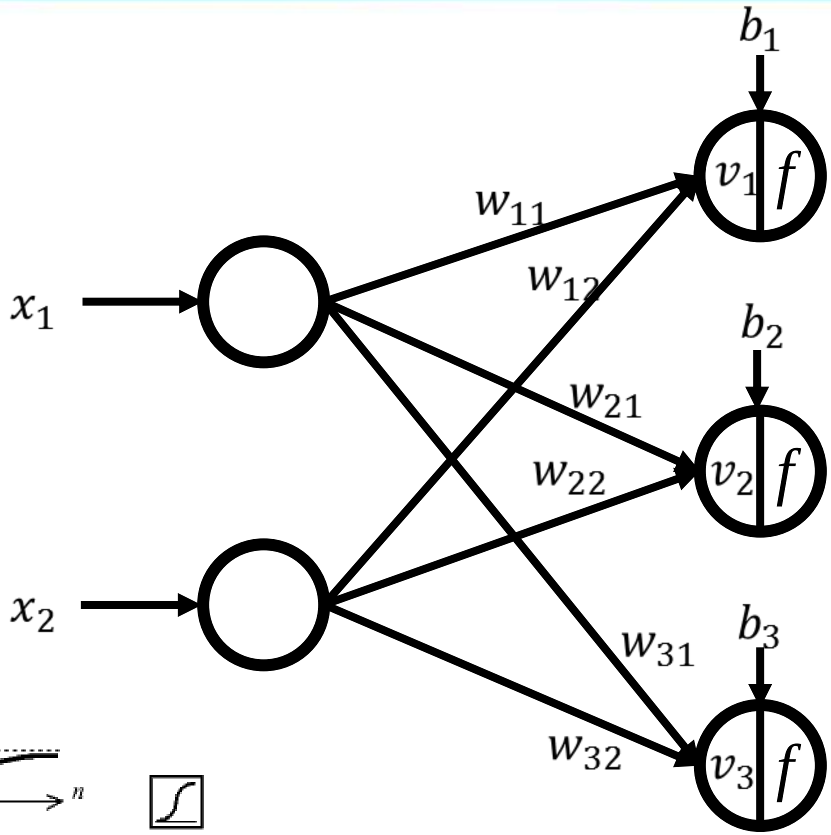
$$f'(x) = f(x)\big(1 - f(x)\big) \qquad f'(v_2) = f(v_2)\big(1 - f(v_2)\big)$$



$a = logsig(n)$

Log-Sigmoid Transfer Function

Forward Path

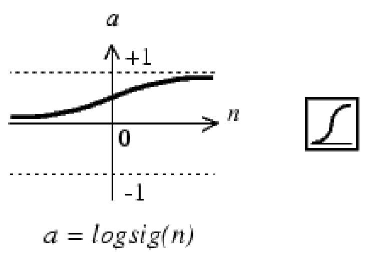


$$v_1 = w_{11}x_1 + w_{12}x_2 + b_1$$

$$y_{h1} = f(v_1) = \frac{1}{1+e^{-v_1}}$$

$$v_2 = w_{21}x_1 + w_{22}x_2 + b_2$$

$$y_{h2} = f(v_2) = \frac{1}{1+e^{-v_2}}$$



Log-Sigmoid Transfer Function

$$\boldsymbol{v_3 = w_{31}x_1} + w_{32}x_2 + b_3$$

Activation function:

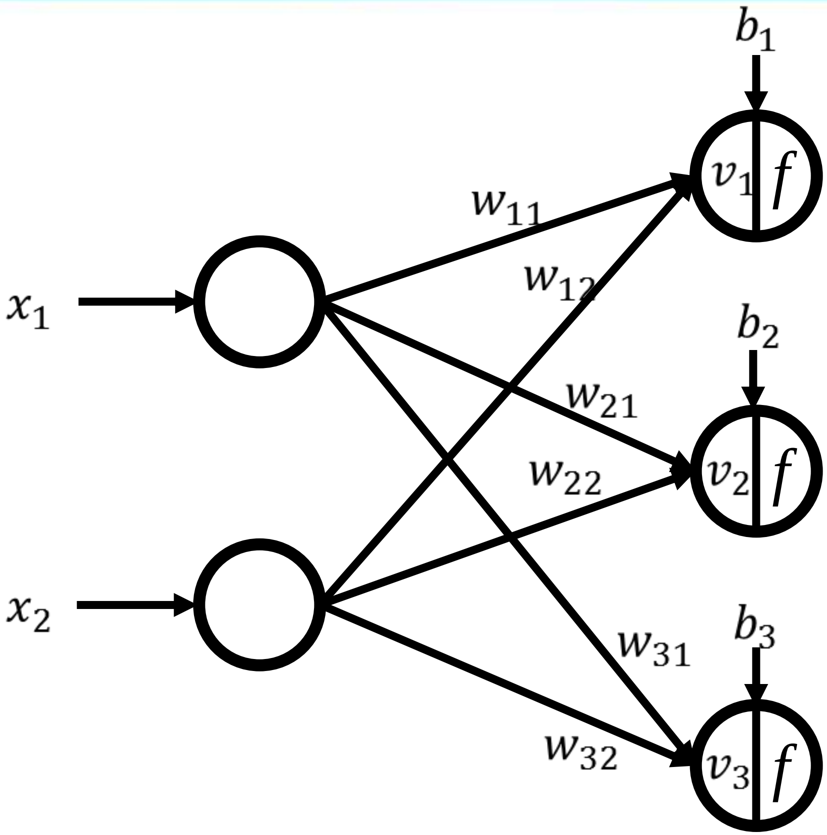
$$f(x) = \frac{1}{1+e^{-x}}$$

$$f'(x) = f(x)\big(1 - f(x)\big)$$

$$f(v_3) = \frac{1}{1+e^{-v_3}}$$

$$f'(v_3) = f(v_3)\big(1 - f(v_3)\big)$$

Forward Path



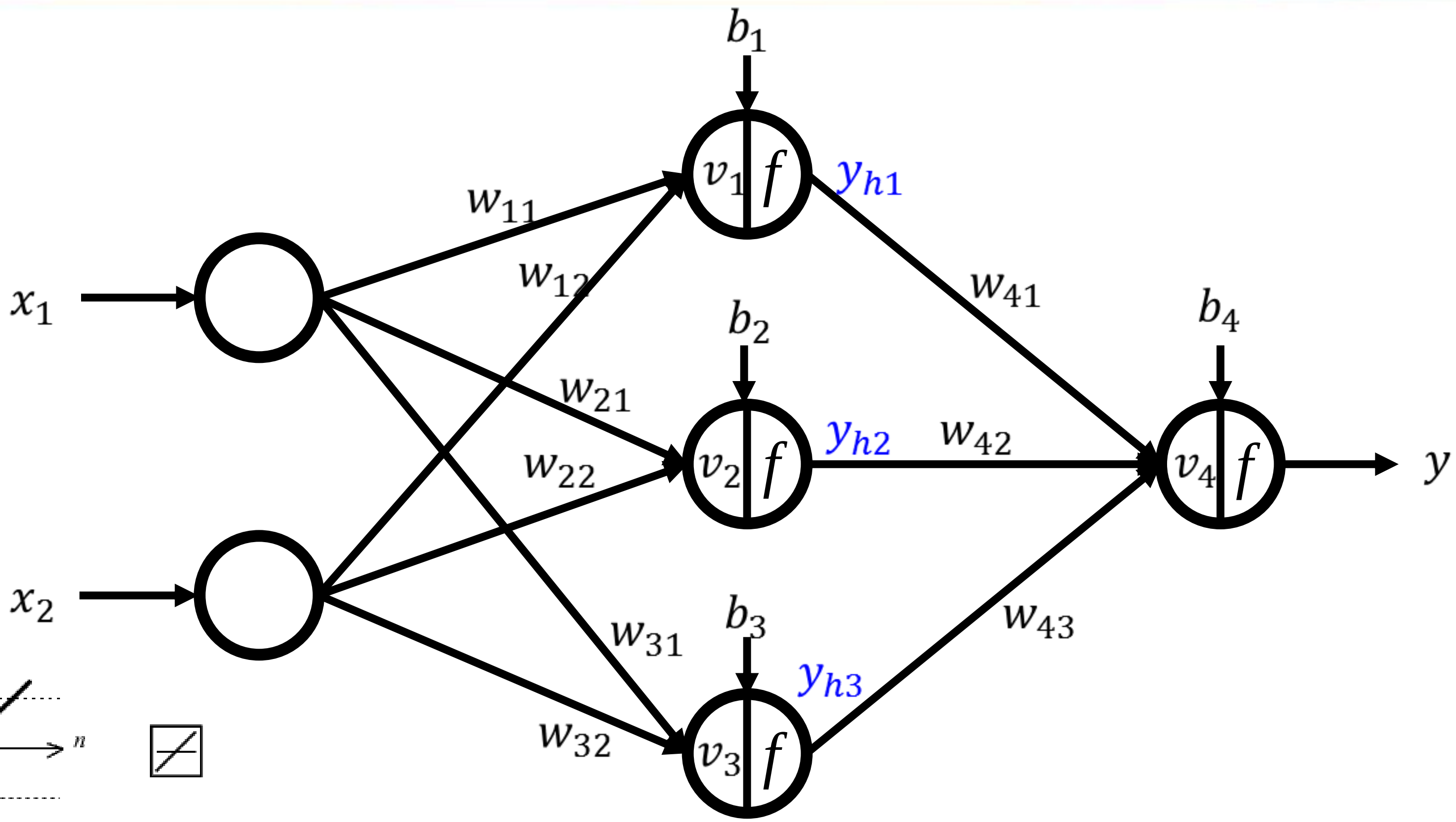
$$v_1 = w_{11}x_1 + w_{12}x_2 + b_1$$
$$y_{h1} = f(v_1) = \frac{1}{1+e^{-v_1}}$$

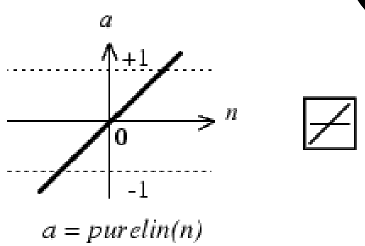
$$v_2 = w_{21}x_1 + w_{22}x_2 + b_2$$
$$y_{h2} = f(v_2) = \frac{1}{1+e^{-v_2}}$$

$$v_3 = w_{31}x_1 + w_{32}x_2 + b_3$$
$$y_{h3} = f(v_3) = \frac{1}{1+e^{-v_3}}$$

$$v_i = \sum_k w_{ik}x_k + b_i \qquad y_{hi} = f(v_i) = \frac{1}{1+e^{-v_i}} \qquad f'(v_i) = f(v_i)\big(1 - f(v_i)\big)$$

Forward Path





Linear Transfer Function

$$\boldsymbol{v_4 = w_{41}y_{h1} + w_{42}y_{h2} + w_{43}y_{h3}} + b_4$$

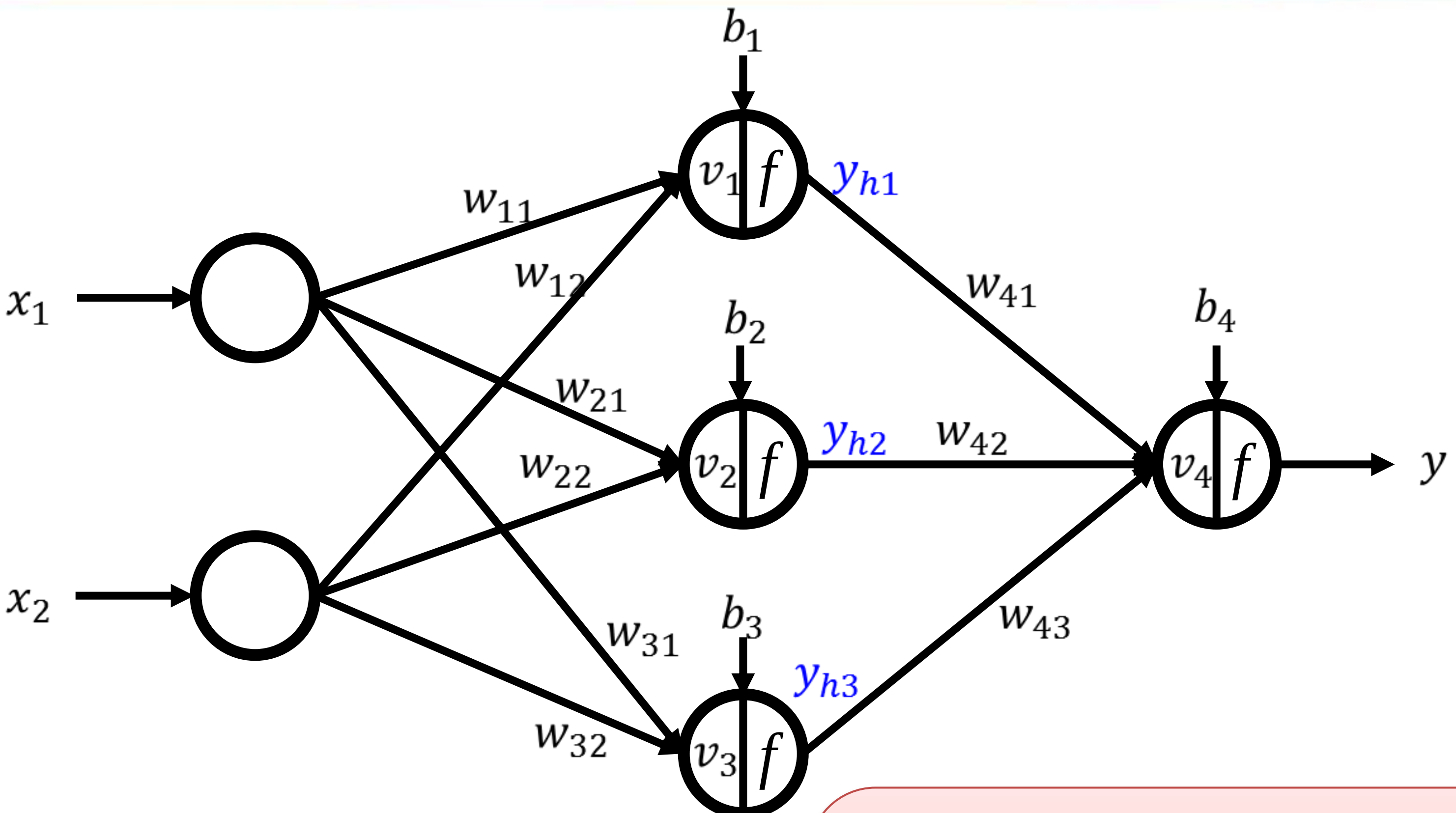
$$v_4 = w_{41}f(v_1) + w_{42}f(v_2) + w_{43}f(v_3) + b_4$$

Activation function:

$$f(x) = x \qquad f(v_4) = v_4$$

$$f'(x) = 1 \qquad f'(v_4) = 1$$

Forward Path



$$v_4 = w_{41}f(v_1) + w_{42}f(v_2) + w_{43}f(v_3) + b_4$$

$$v_4 = w_{41}y_{h1} + w_{42}y_{h2} + w_{43}y_{h3} + b_4$$

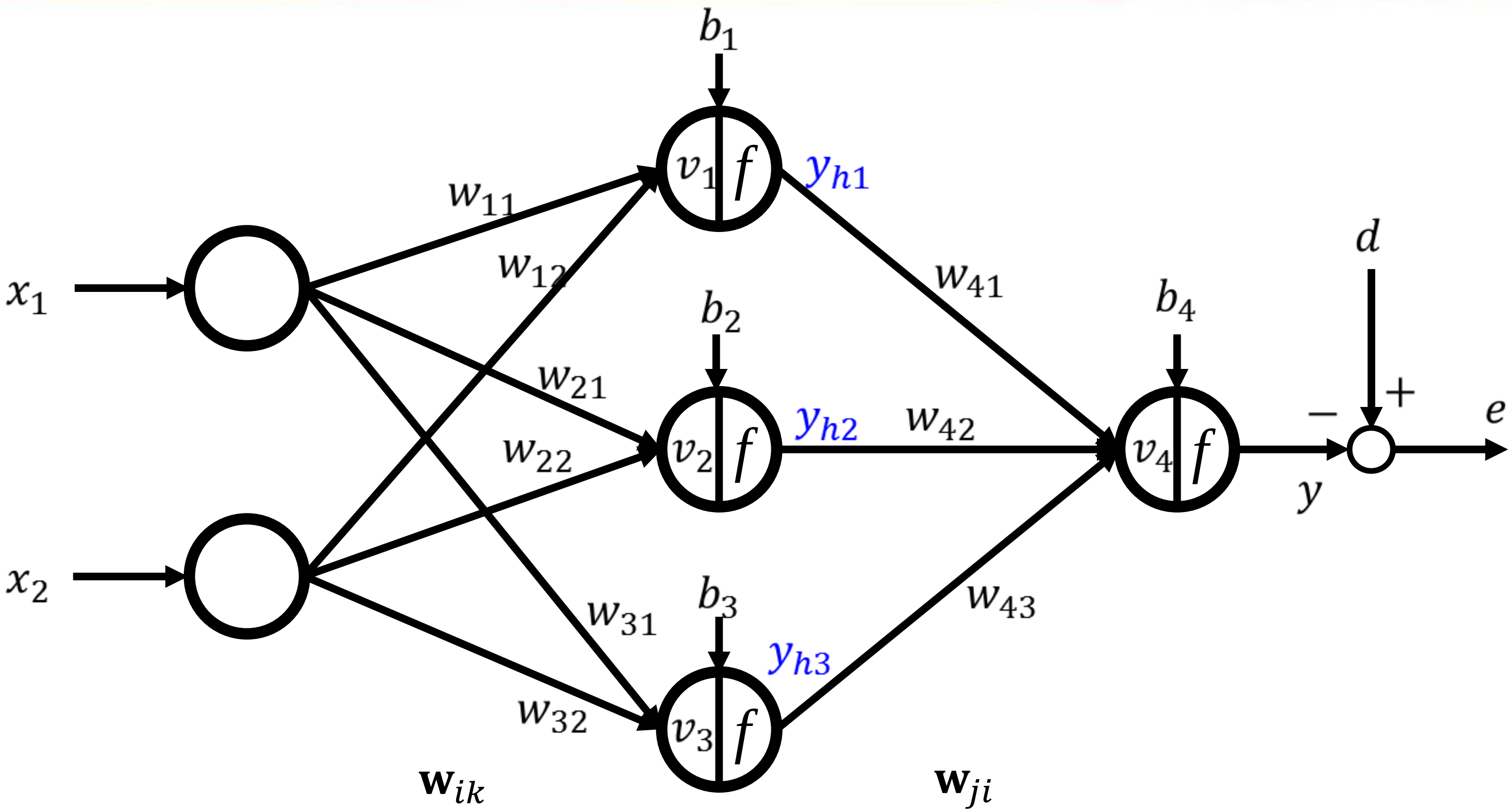
$$y = f(v_4) = v_4$$

$$v_j = \sum_i w_{ji}y_{hi} + b_j = \sum_i w_{ji}f(v_i) + b_j$$

$$y_j = f(v_j) = v_j$$

$$f'(v_j) = 1$$

Forward Path



Cost function:

$$e_j(n) = d_j(n) - y_j(n)$$

$$E_j(n) = \frac{1}{2} e_j(n)^2$$
$$= \frac{1}{2} \left(d_j(n) - y_j(n)\right)^2$$

Backward Path

(1) Update rule for the weights of the output neurons:

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n)$$

$$= w_{ji}(n) - \eta \frac{\partial E_j(n)}{\partial w_{ji}(n)}$$

$$= w_{ji}(n) - \eta \frac{\partial E_j(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)}$$

$$= w_{ji}(n) - \eta e_j(n)(-1) f'\big(v_j(n)\big) f\big(v_i(n)\big)$$

$$= w_{ji}(n) - \eta e_j(n)(-1)(1) f\big(v_i(n)\big)$$

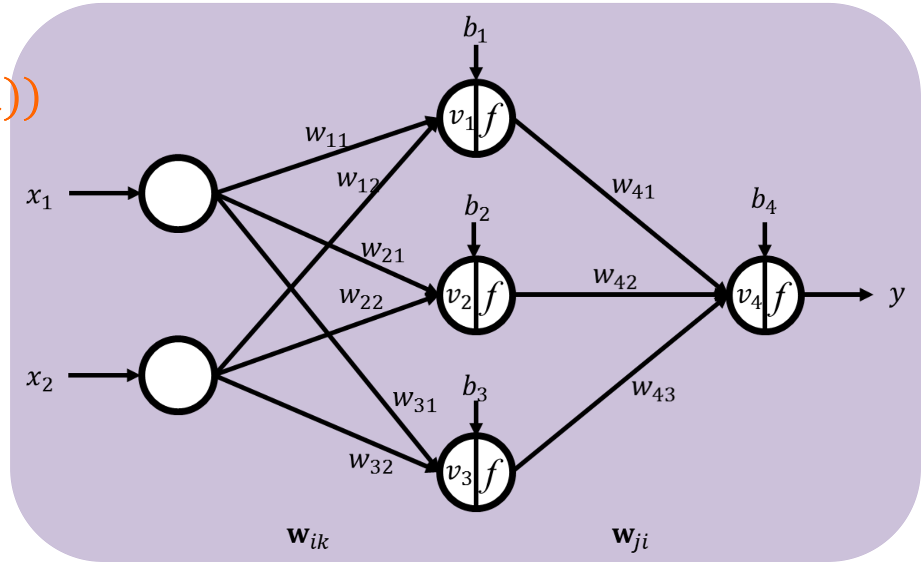
$$= w_{ji}(n) + \eta \big(d_j(n) - y_j(n)\big) f\big(v_i(n)\big)$$

$$E_j(n) = \frac{1}{2} e_j(n)^2$$

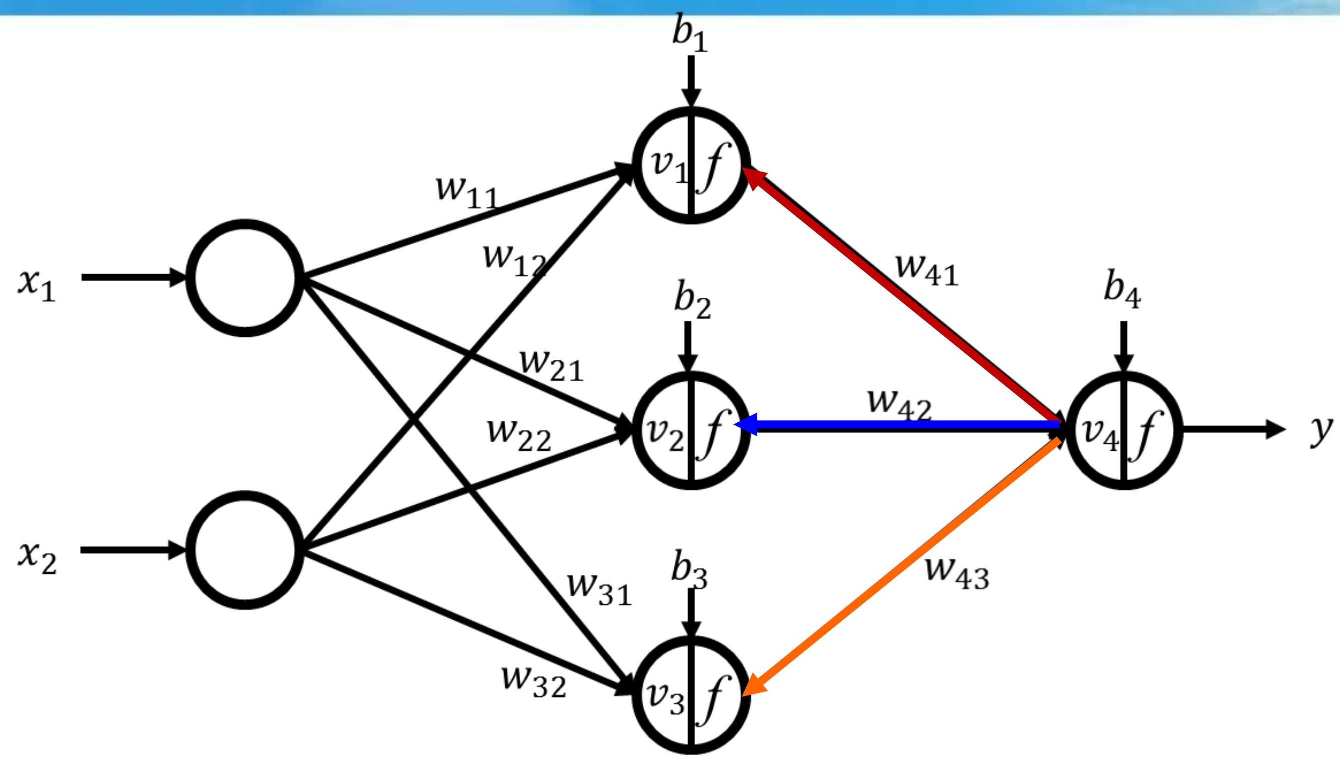
$$e_j(n) = d_j(n) - y_j(n)$$

$$y_j = f(v_j) = v_j$$

$$v_j = \sum_i w_{ji} f(v_i) + b_j$$



Backward Path



$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n) = w_{ji}(n) - \eta \frac{\partial E_j(n)}{\partial w_{ji}(n)} = w_{ji}(n) + \eta\big(d_j(n) - y_j(n)\big)f(v_i(n))$$

$$w_{41}(n+1) = w_{41}(n) + \Delta w_{41}(n) = w_{41}(n) + \eta\big(d(n) - y(n)\big)f(v_1(n))$$

$$w_{42}(n+1) = w_{42}(n) + \Delta w_{42}(n) = w_{42}(n) + \eta\big(d(n) - y(n)\big)f(v_2(n))$$

$$w_{43}(n+1) = w_{43}(n) + \Delta w_{43}(n) = w_{43}(n) + \eta\big(d(n) - y(n)\big)f(v_3(n))$$

Backward Path

(2) Update rule for the biases of the output neurons:

$$b_j(n+1) = b_j(n) + \Delta b_j(n)$$

$$= b_j(n) - \eta \frac{\partial E_j(n)}{\partial b_j(n)}$$

$$= b_j(n) - \eta \frac{\partial E_j(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial b_j(n)}$$

$$= b_j(n) - \eta e_j(n)(-1) f'\big(v_j(n)\big)(1)$$

$$= b_j(n) - \eta e_j(n)(-1)(1)(1)$$

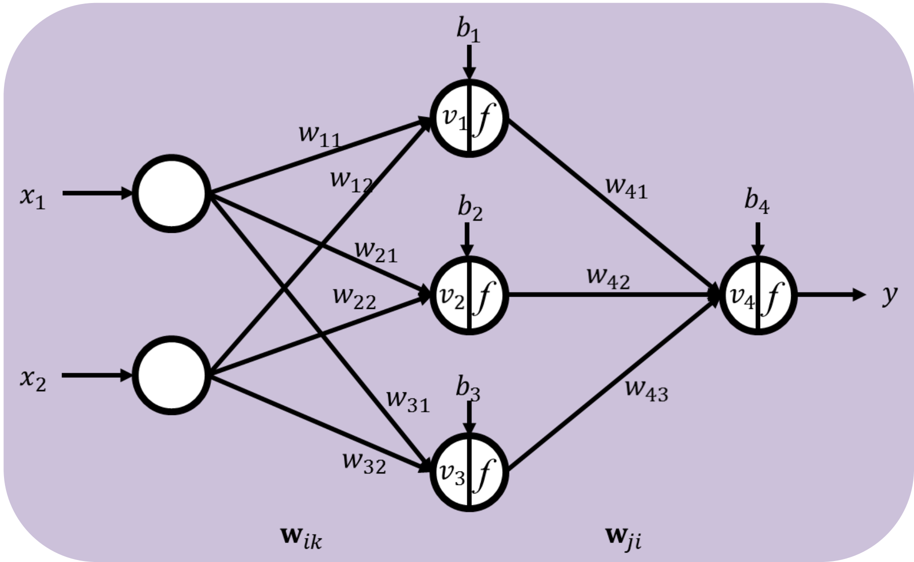
$$= b_j(n) + \eta\big(d_j(n) - y_j(n)\big)$$

$$E_j(n) = \frac{1}{2} e_j(n)^2$$

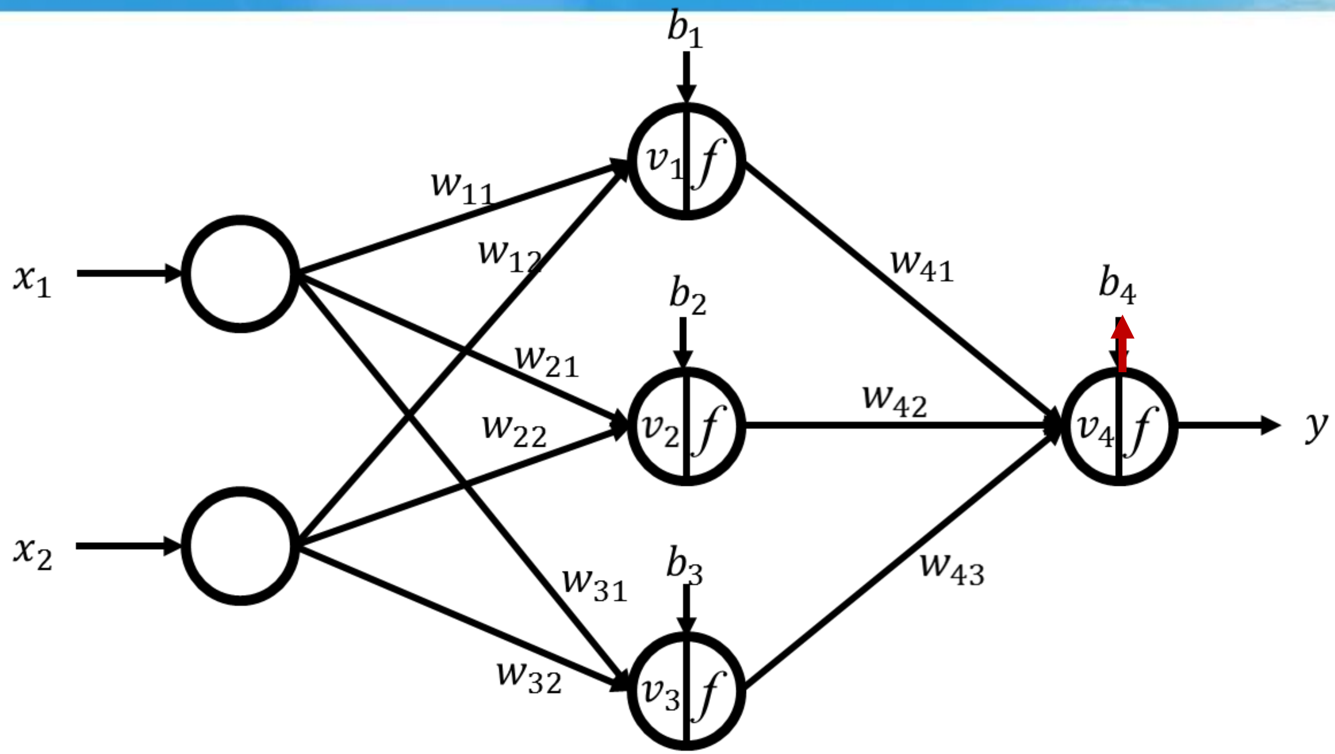
$$e_j(n) = d_j(n) - y_j(n)$$

$$y_j = f\big(v_j\big) = v_j$$

$$v_j = \sum_i w_{ji} f(v_i) + b_j$$



Backward Path



$$b_j(n+1) = b_j(n) + \Delta b_j(n) = b_j(n) - \eta \frac{\partial E_j(n)}{\partial b_j(n)} = b_j(n) + \eta\big(d_j(n) - y_j(n)\big)$$

$$b_4(n+1) = b_4(n) + \Delta b_4(n) = b_4(n) + \eta\big(d(n) - y(n)\big)$$

Backward Path

(3) Update rule for the weights of the hidden neurons:

$$w_{ik}(n+1) = w_{ik}(n) + \Delta w_{ik}(n)$$

$$= w_{ik}(n) - \eta \frac{\partial E_j(n)}{\partial w_{ik}(n)}$$

$$= w_{ik}(n) - \eta \frac{\partial E_j(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial f(v_i(n))} \frac{\partial f(v_i(n))}{\partial v_i(n)} \frac{\partial v_i(n)}{\partial w_{ik}(n)}$$

$$= w_{ik}(n) - \eta e_j(n)(-1) f'(v_j(n)) w_{ji}(n) f'(v_i(n)) x_k(n)$$

$$= w_{ik}(n) - \eta e_j(n)(-1)(1) w_{ji}(n) \left[f(v_i(n))\left(1 - f(v_i(n))\right) \right] x_k(n)$$

$$= w_{ik}(n) + \eta \left(d_j(n) - y_j(n)\right) w_{ji}(n) \left[f(v_i(n))\left(1 - f(v_i(n))\right) \right] x_k(n)$$

$$E_j(n) = \frac{1}{2} e_j(n)^2$$

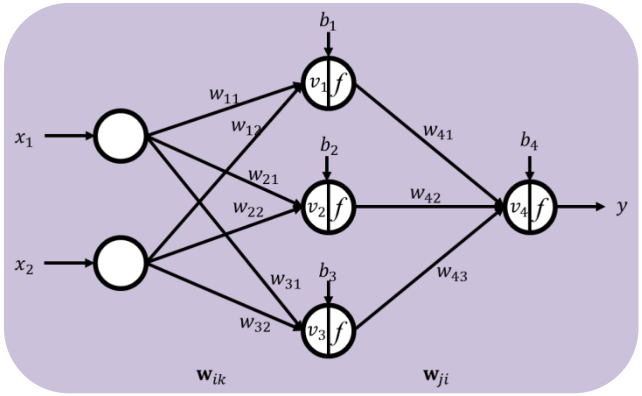
$$e_j(n) = d_j(n) - y_j(n)$$

$$y_j = f(v_j) = v_j$$

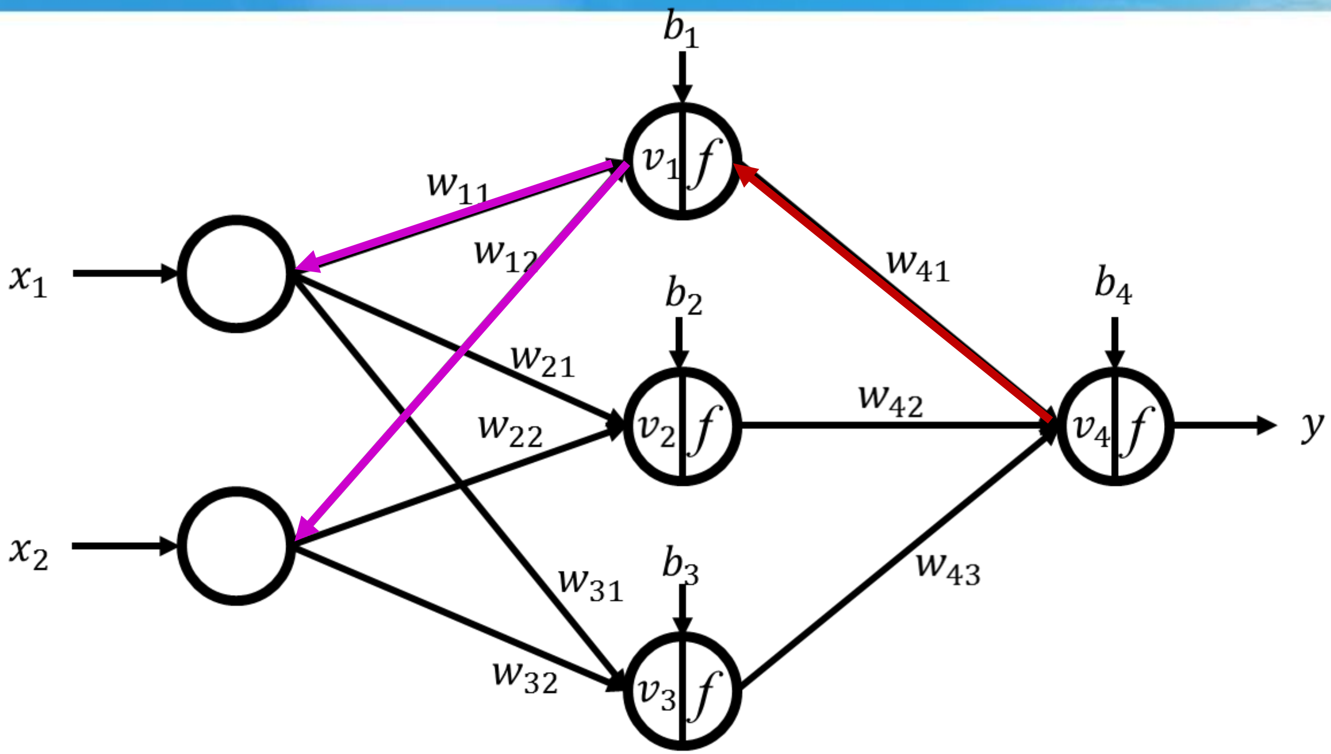
$$v_j = \sum_i w_{ji} f(v_i) + b_j$$

$$y_{hi} = f(v_i) = \frac{1}{1 + e^{-v_i}}$$

$$v_i = \sum_k w_{ik} x_k + b_i$$



Backward Path

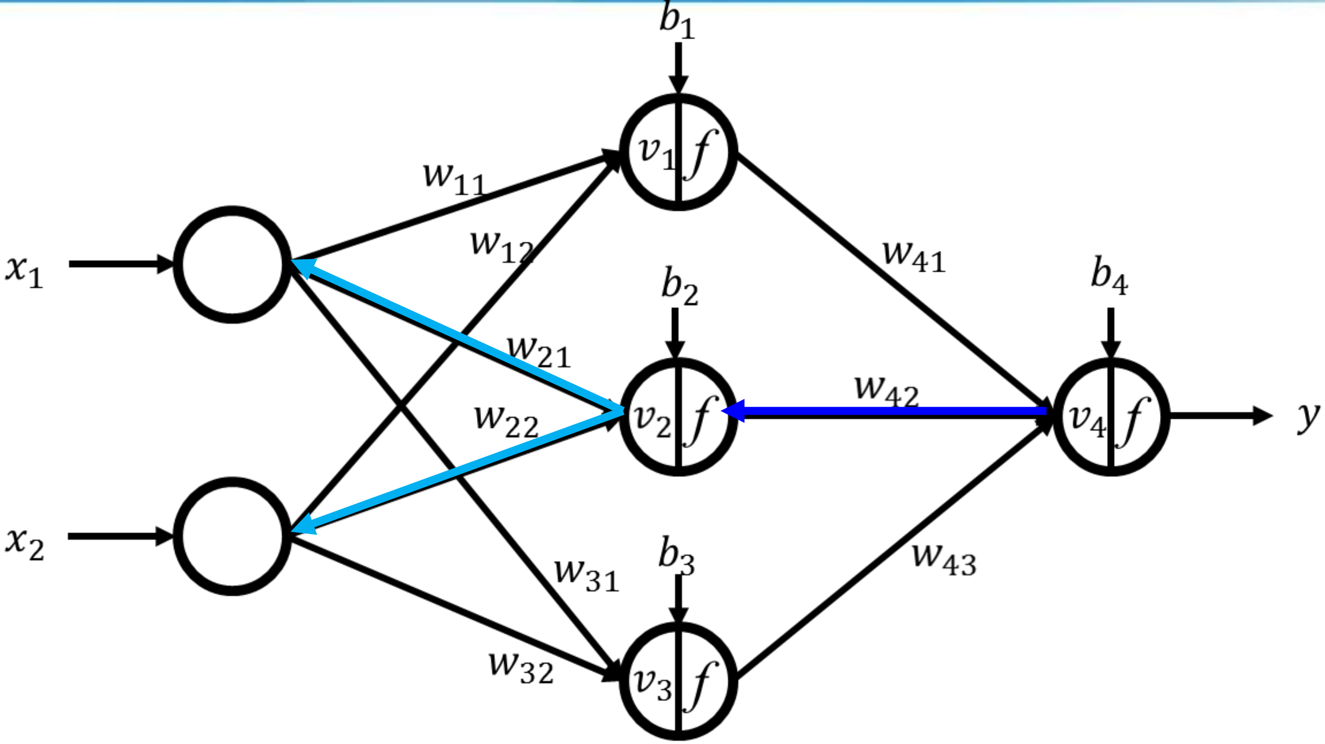


$$w_{ik}(n+1) = w_{ik}(n) + \Delta w_{ik}(n) = w_{ik}(n) + \eta\big(d_j(n) - y_j(n)\big)w_{ji}(n)\big[f(v_i(n))\big(1 - f(v_i(n))\big)\big]x_k(n)$$

$$\begin{aligned} w_{11}(n+1) &= w_{11}(n) + \Delta w_{11}(n) \\ &= w_{11}(n) + \eta\big(d(n) - y(n)\big)w_{41}(n)f(v_1(n))\big(1 - f(v_1(n))\big)x_1(n) \end{aligned}$$

$$\begin{aligned} w_{12}(n+1) &= w_{12}(n) + \Delta w_{12}(n) \\ &= w_{12}(n) + \eta\big(d(n) - y(n)\big)w_{41}(n)f(v_1(n))\big(1 - f(v_1(n))\big)x_2(n) \end{aligned}$$

Backward Path

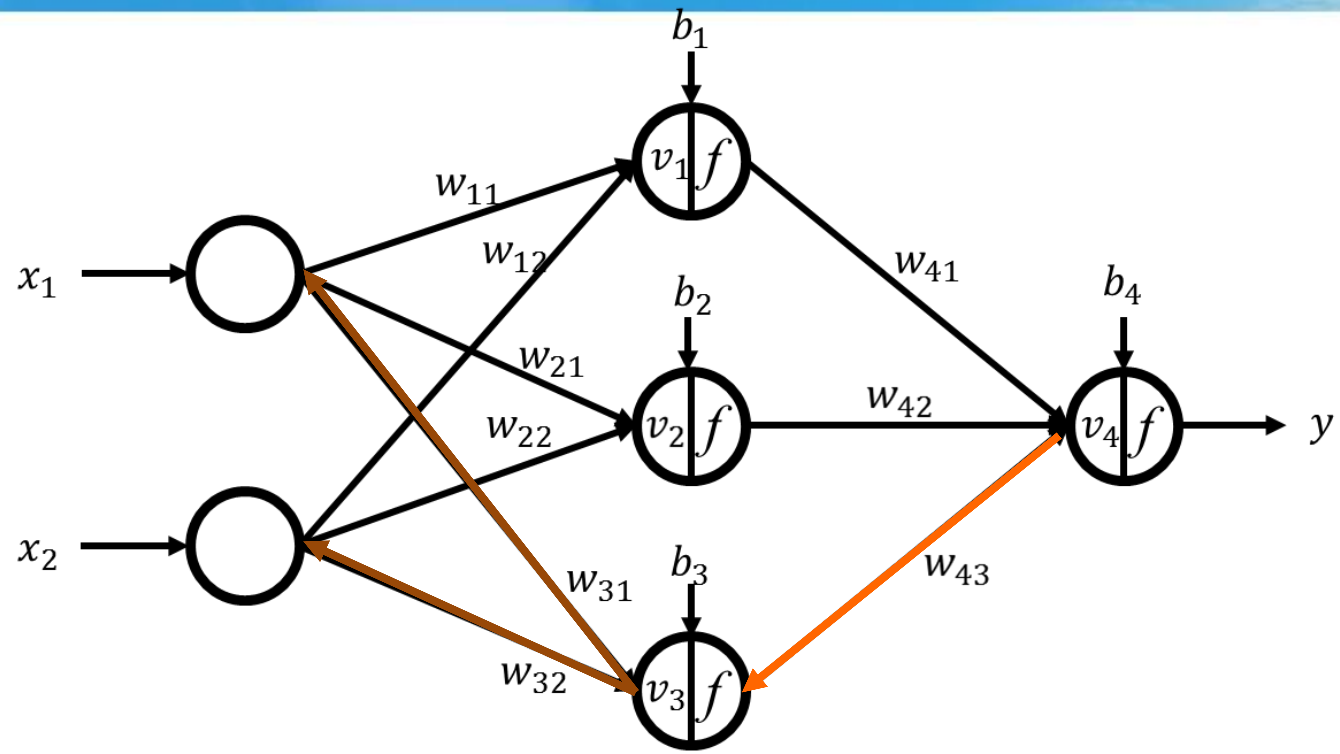


$$w_{ik}(n+1) = w_{ik}(n) + \Delta w_{ik}(n) = w_{ik}(n) + \eta\big(d_j(n) - y_j(n)\big)w_{ji}(n)\big[f(v_i(n))\big(1 - f(v_i(n))\big)\big]x_k(n)$$

$$\begin{aligned} w_{21}(n+1) &= w_{21}(n) + \Delta w_{21}(n) \\ &= w_{21}(n) + \eta\big(d(n) - y(n)\big)w_{42}(n)f(v_2(n))\big(1 - f(v_2(n))\big)x_1(n) \end{aligned}$$

$$\begin{aligned} w_{22}(n+1) &= w_{22}(n) + \Delta w_{22}(n) \\ &= w_{22}(n) + \eta\big(d(n) - y(n)\big)w_{42}(n)f(v_2(n))\big(1 - f(v_2(n))\big)x_2(n) \end{aligned}$$

Backward Path



$$w_{ik}(n+1) = w_{ik}(n) + \Delta w_{ik}(n) = w_{ik}(n) + \eta\big(d_j(n) - y_j(n)\big)w_{ji}(n)\big[f(v_i(n))\big(1 - f(v_i(n))\big)\big]x_k(n)$$

$$\begin{aligned} w_{31}(n+1) &= w_{31}(n) + \Delta w_{31}(n) \\ &= w_{31}(n) + \eta\big(d(n) - y(n)\big)w_{43}(n)f(v_3(n))\big(1 - f(v_3(n))\big)x_1(n) \end{aligned}$$

$$\begin{aligned} w_{32}(n+1) &= w_{32}(n) + \Delta w_{32}(n) \\ &= w_{32}(n) + \eta\big(d(n) - y(n)\big)w_{43}(n)f(v_3(n))\big(1 - f(v_3(n))\big)x_2(n) \end{aligned}$$

Backward Path

(4) Update rule for the biases of the hidden neurons:

$$b_i(n+1) = b_i(n) + \Delta b_i(n)$$

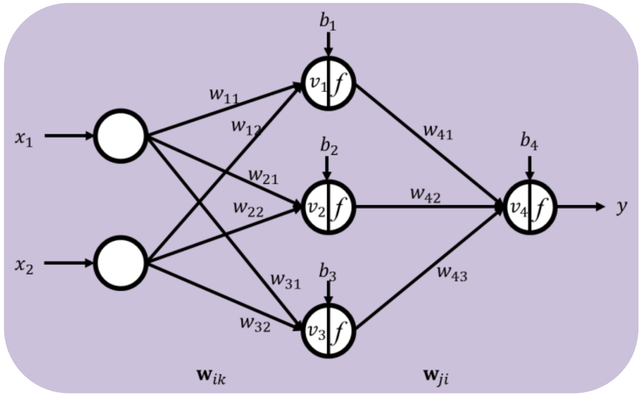
$$= b_i(n) - \eta \frac{\partial E_j(n)}{\partial b_i(n)}$$

$$= b_i(n) - \eta \frac{\partial E_j(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial f(v_i(n))} \frac{\partial f(v_i(n))}{\partial v_i(n)} \frac{\partial v_i(n)}{\partial b_i(n)}$$

$$= b_i(n) - \eta e_j(n)(-1) f'\big(v_j(n)\big) w_{ji}(n) f'(v_i(n))(1)$$

$$= b_i(n) - \eta e_j(n)(-1)(1) w_{ji}(n)\big[f(v_i(n))\big(1 - f(v_i(n))\big)\big](1)$$

$$= b_i(n) + \eta\big(d_j(n) - y_j(n)\big) w_{ji}(n)\big[f(v_i(n))\big(1 - f(v_i(n))\big)\big]$$



$$E_j(n) = \frac{1}{2} e_j(n)^2$$

$$e_j(n) = d_j(n) - y_j(n)$$

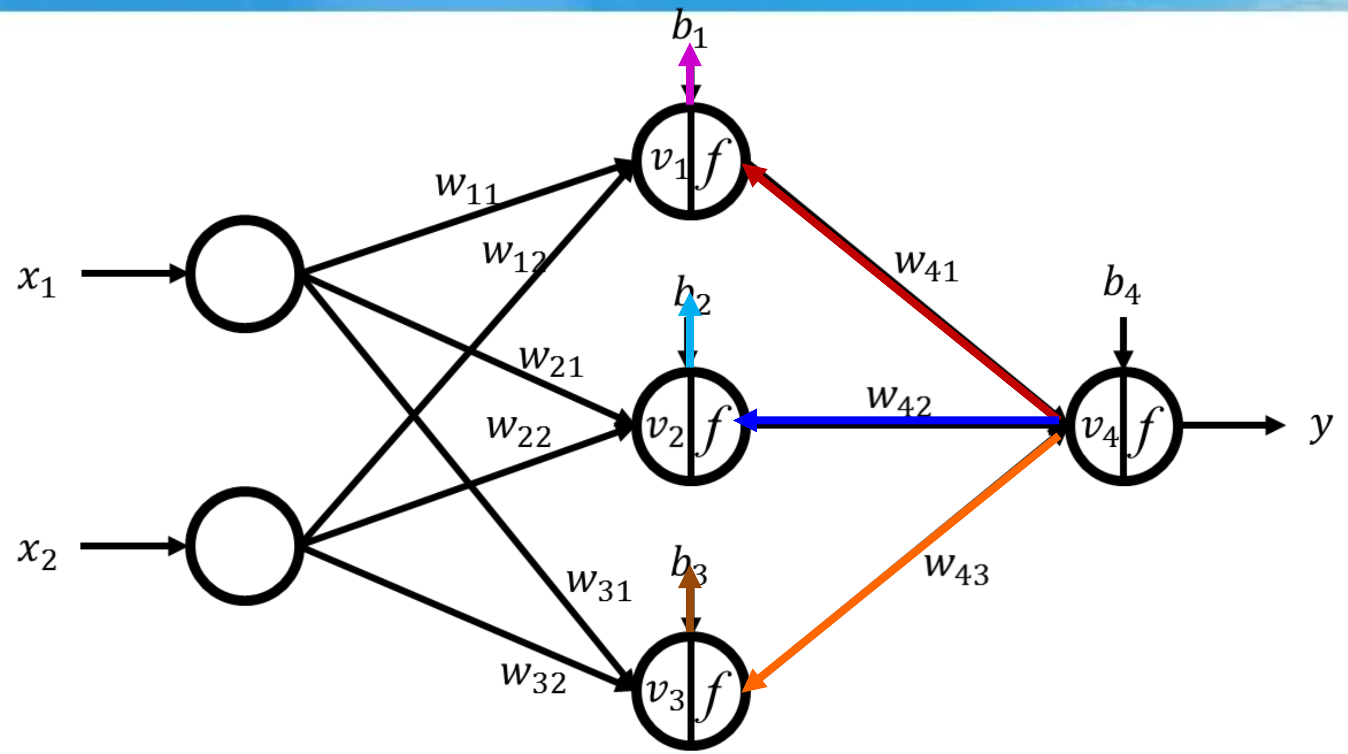
$$y_j = f(v_j) = v_j$$

$$v_j = \sum_i w_{ji} f(v_i) + b_j$$

$$y_{hi} = f(v_i) = \frac{1}{1 + e^{-v_i}}$$

$$v_i = \sum_k w_{ik} x_k + b_i$$

Backward Path



$$b_i(n+1) = b_i(n) + \Delta b_i(n) = b_i(n) + \eta\big(d_j(n) - y_j(n)\big)w_{ji}(n)\big[f(v_i(n))\big(1 - f(v_i(n))\big)\big]$$

$$b_1(n+1) = b_1(n) + \Delta b_1(n) = b_1(n) + \eta\big(d(n) - y(n)\big)w_{41}(n)f(v_1(n))\big(1 - f(v_1(n))\big)$$

$$b_2(n+1) = b_2(n) + \Delta b_2(n) = b_2(n) + \eta\big(d(n) - y(n)\big)w_{42}(n)f(v_2(n))\big(1 - f(v_2(n))\big)$$

$$b_3(n+1) = b_3(n) + \Delta b_3(n) = b_3(n) + \eta\big(d(n) - y(n)\big)w_{43}(n)f(v_3(n))\big(1 - f(v_3(n))\big)$$

Thanks for your attention !!

Any Question?