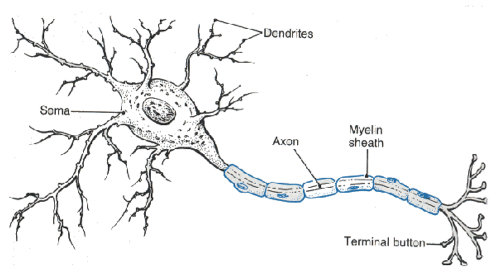
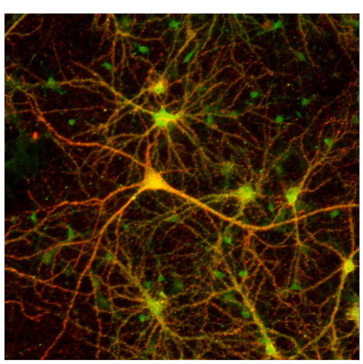
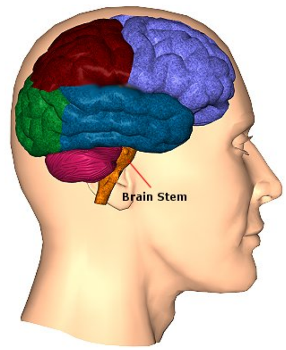
Multilayer Neural Networks

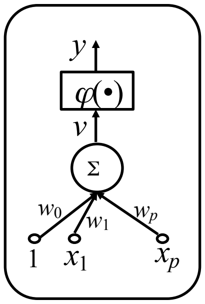
Pengyu Hong
CS 101A

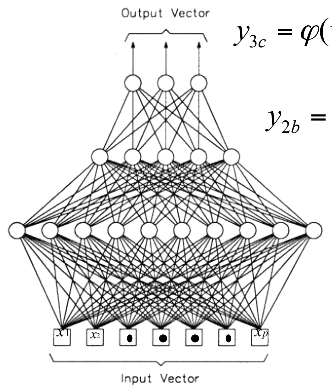






Multilayer Neural Networks

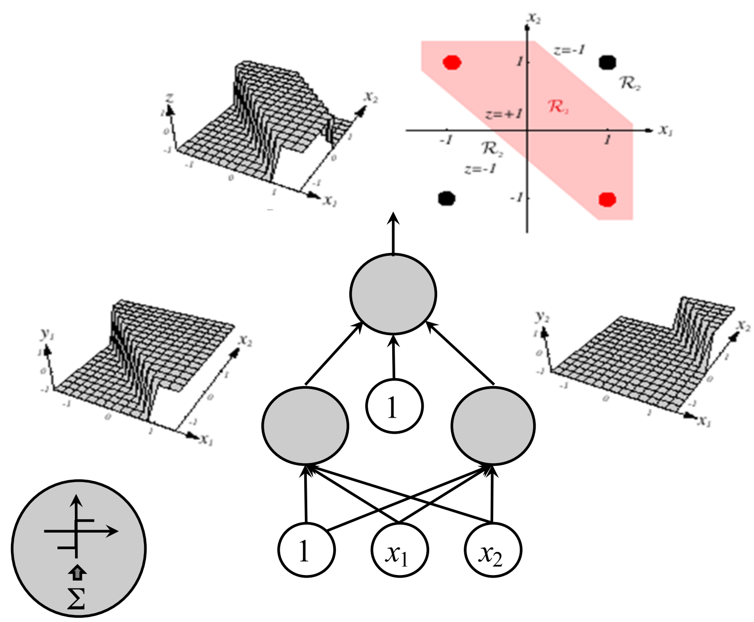




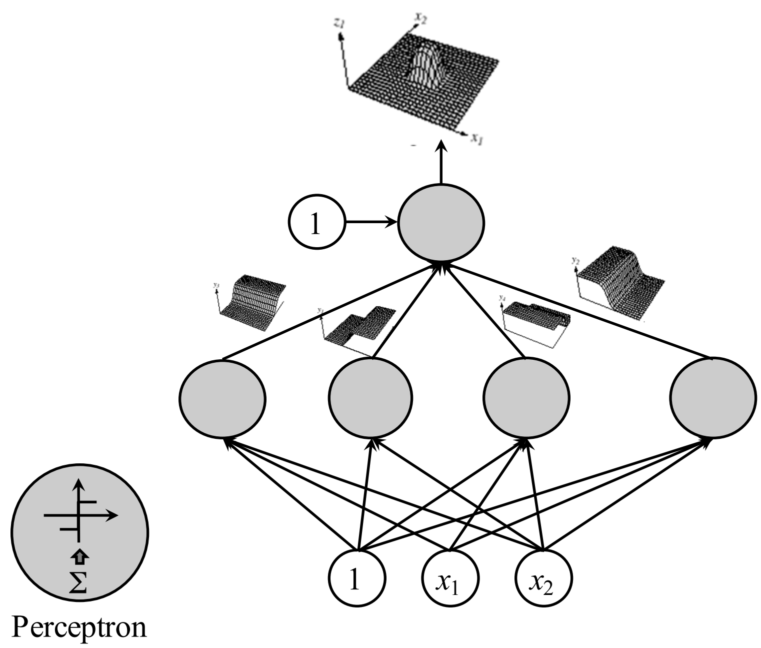
$$y_{3c} = \varphi(v_{3c}) = \varphi(\sum_{b=0}^{n_2} w_{3b} y_{2b})$$

$$y_{2b} = \varphi(v_{2b}) = \varphi(\sum_{a=0}^{n_1} w_{2a} y_{1a})$$

$$y_{1a} = \varphi(v_{1a}) = \varphi(\sum_{i=0}^{p} w_{1i} x_i)$$



Perceptron

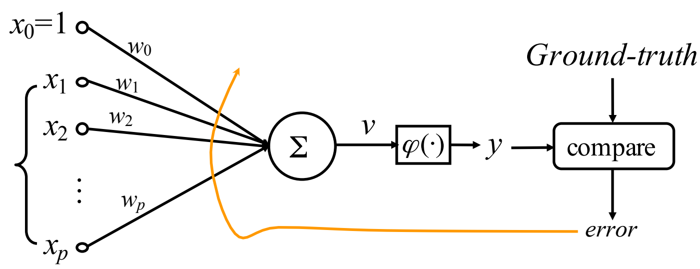


Universal Approximation Theorem

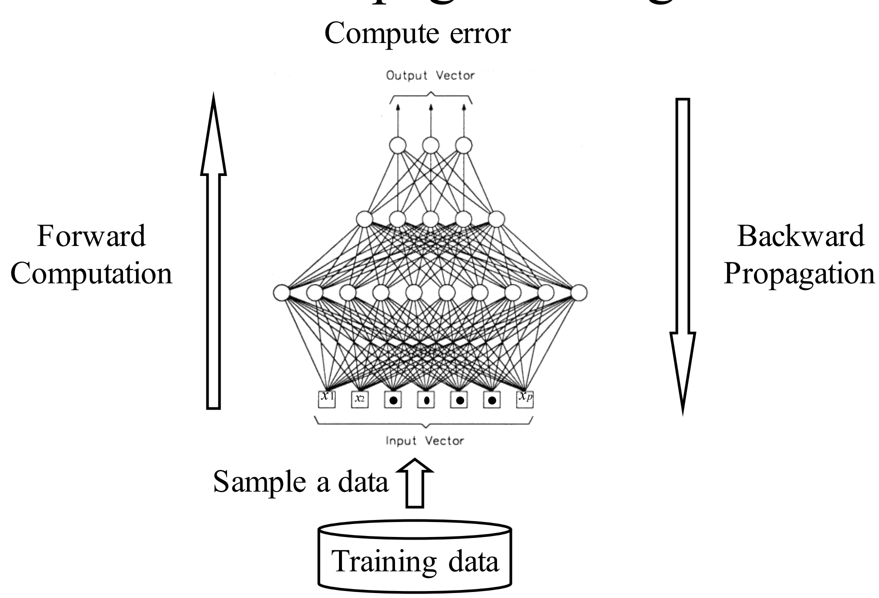
Any mapping function from input to output can be implemented as a three-layer neural network.

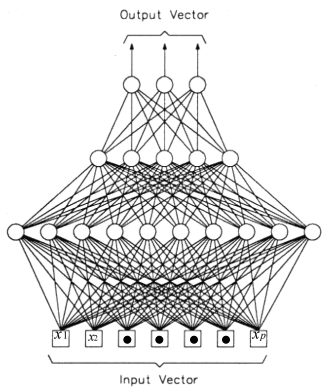
Neural Networks (2nd ed). Haykin. p. 208-209 and 249

Error-Correction Learning



The Back-Propagation Algorithm





Error of a single output neuron j

$$e_j(t) = d_j(t) - y_j(t)$$
$$= d_j(t) - \varphi_j(v_j(t))$$
$$= d_j(t) - \varphi_j\Big(\sum_{i=0}^{m} w_{ji}(t) y_i(t)\Big)$$

Error of the output layer

$$Err(t) = \frac{1}{2} \sum_{j \in \{\text{output Neurons}\}} e_j^2(t)$$

$$\frac{\partial Err(t)}{\partial w_{ji}(t)} = \frac{\partial Err(t)}{\partial e_j(t)} \frac{\partial e_j(t)}{\partial y_j(t)} \frac{\partial y_j(t)}{\partial v_j(t)} \frac{\partial v_j(t)}{\partial w_{ji}(t)}$$

$$\frac{\partial Err(t)}{\partial w_{ji}(t)} = \frac{\partial Err(t)}{\partial e_j(t)} \frac{\partial e_j(t)}{\partial y_j(t)} \frac{\partial y_j(t)}{\partial v_j(t)} \frac{\partial v_j(t)}{\partial w_{ji}(t)}$$

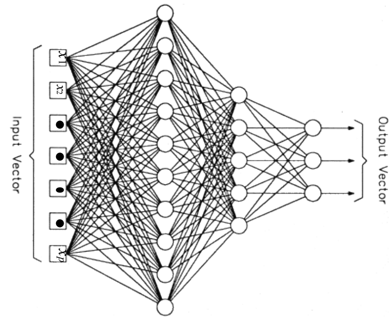
$$\frac{\partial Err(t)}{\partial e_j(t)} = \frac{\partial \left(\frac{1}{2} \sum_{j=1}^{n_o} e_j^2(t) \right)}{\partial e_j(t)} = e_j(t) \qquad \frac{\partial e_j(t)}{\partial y_j(t)} = \frac{\partial \big(d_j(t) - y_j(t)\big)}{\partial y_j(t)} = -1$$

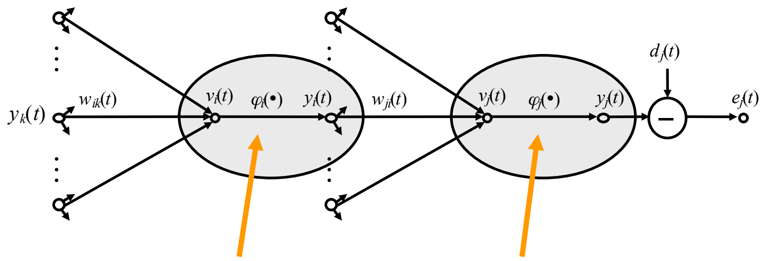
$$\frac{\partial y_j(t)}{\partial v_j(t)} = \frac{\partial \varphi_j(v_j(t))}{\partial v_j(t)} = \varphi_j'(v_j(t)) \qquad \frac{\partial v_j(t)}{\partial w_{ji}(t)} = \frac{\partial \left(\sum_{i=0}^{m} w_{ji} y_i(t) \right)}{\partial w_{ji}(t)} = y_i(t)$$

$$\frac{\partial Err(t)}{\partial w_{ji}(t)} = -e_j(t)\varphi_j'(v_j(t)) y_i(t) \qquad \Delta w_{ji} = -\eta \frac{\partial Err(t)}{\partial w_{ji}(t)} = \eta \delta_j(t) y_i(t)$$

$$\delta_j(t) = \frac{\partial Err(t)}{\partial v_j(t)} = \frac{\partial Err(t)}{\partial y_j(t)} \frac{\partial y_j(t)}{\partial v_j(t)} = e_j(t)\varphi_j'(v_j(t))$$

Local gradient





$$\delta_i(t) = \frac{\partial Err(t)}{\partial y_i(t)} \frac{\partial y_i(t)}{\partial v_i(t)} \qquad \delta_j(t) = \varphi_j'(v_j(t))e_j(t)$$

$$\delta_i(t) = \frac{\partial Err(t)}{\partial y_i(t)} \frac{\partial y_i(t)}{\partial v_i(t)} \qquad Err(t) = \frac{1}{2}\sum_{j=1}^{n_o} e_j^2(t)$$

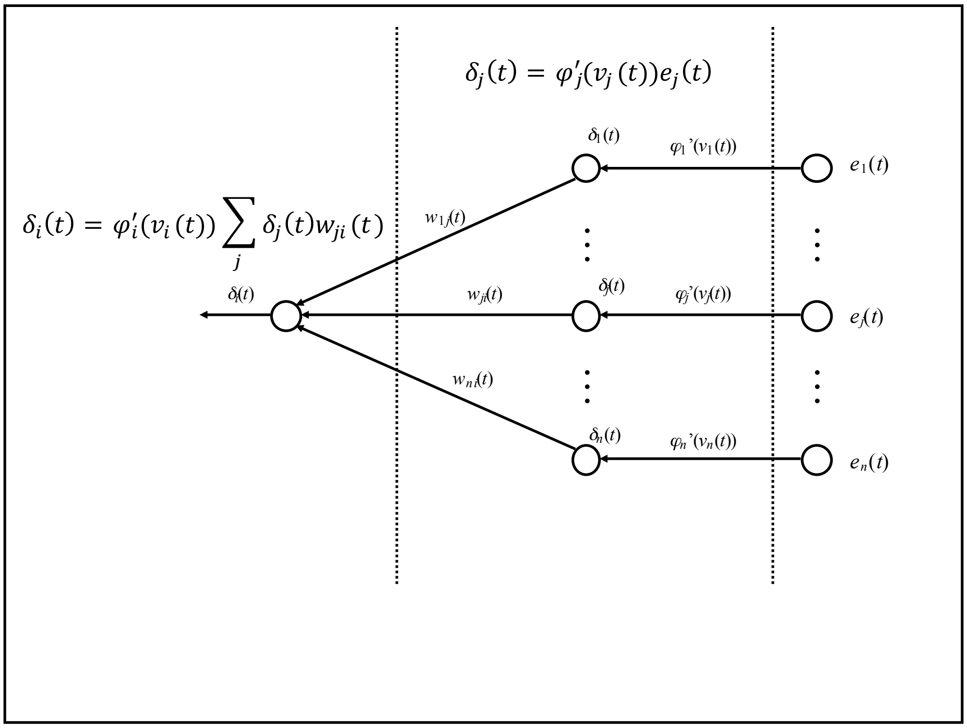
$$\frac{\partial Err(t)}{\partial y_i(t)} = \sum_j \frac{\partial Err(t)}{\partial e_j(t)} \frac{\partial e_j(t)}{\partial y_i(t)} = \sum_j e_j(t) \frac{\partial e_j(t)}{\partial y_i(t)} = \sum_j e_j(t) \frac{\partial e_j(t)}{\partial v_j(t)} \frac{\partial v_j(t)}{\partial y_i(t)}$$

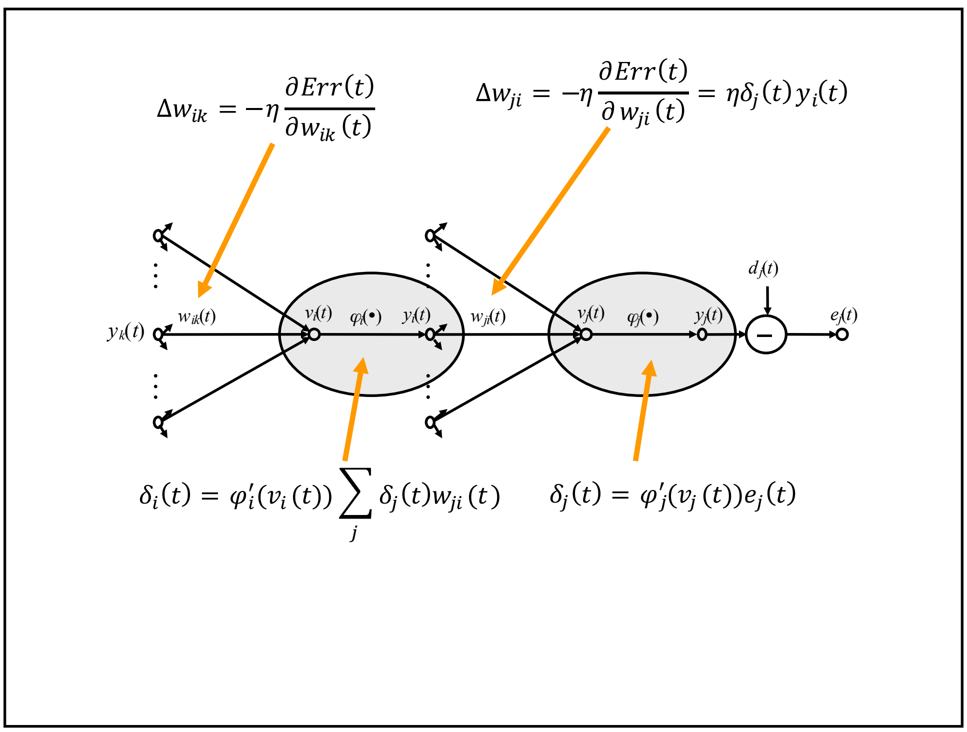
$$\frac{\partial e_j(t)}{\partial v_j(t)} = -\frac{\partial \varphi_j(v_j(t))}{\partial v_j(t)} = -\varphi_j'(v_j(t))$$

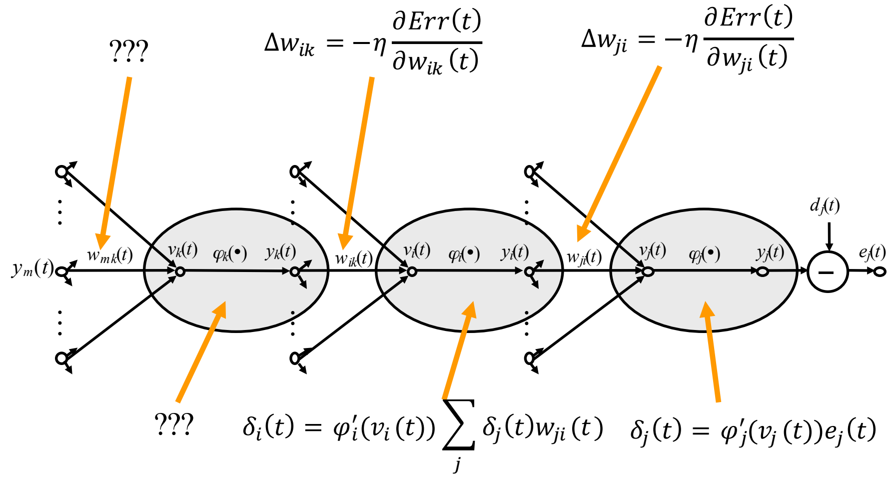
$$v_j(t) = \sum_{i=0}^{m} w_{ji}(t) y_i(t) \qquad \frac{\partial v_j(t)}{\partial y_i(t)} = w_{ji}(t)$$

$$\frac{\partial Err(t)}{\partial y_i(t)} = -\sum_j e_j(t)\varphi_j'(v_j(t))w_{ji}(t) = -\sum_j \delta_j(t) w_{ji}(t)$$

$$\delta_i(t) = \varphi_i'(v_i(t))\sum_j \delta_j(t) w_{ji}(t)$$

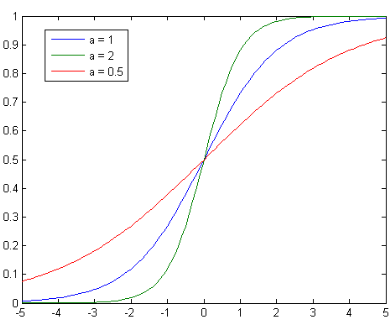

$\delta_j(t) = \varphi_j'(v_j(t))e_j(t)$
$\delta_i(t) = \varphi_i'(v_i(t))\sum_j \delta_j(t)w_{ji}(t)$
$\delta_1(t)$
$\varphi_1'(v_1(t))$
$e_1(t)$
$w_{1i}(t)$
$\delta_i(t)$
$w_{ji}(t)$
$\delta_j(t)$
$\varphi_j'(v_j(t))$
$e_j(t)$
$w_{ni}(t)$
$\delta_n(t)$
$\varphi_n'(v_n(t))$
$e_n(t)$


$\Delta w_{ik} = -\eta \frac{\partial Err(t)}{\partial w_{ik}(t)}$
$\Delta w_{ji} = -\eta \frac{\partial Err(t)}{\partial w_{ji}(t)} = \eta \delta_j(t) y_i(t)$
$d_j(t)$
$y_k(t)$
$w_{ik}(t)$
$v_i(t)$
$\varphi_i(\bullet)$
$y_i(t)$
$w_{ji}(t)$
$v_j(t)$
$\varphi_j(\bullet)$
$y_j(t)$
$e_j(t)$
$\delta_i(t) = \varphi_i'(v_i(t))\sum_j \delta_j(t)w_{ji}(t)$
$\delta_j(t) = \varphi_j'(v_j(t))e_j(t)$



Activation Functions

- Logistic Function $\varphi(v) = \dfrac{1}{1+\exp(-av)}$ $\quad a > 0, \; -\infty < v < \infty$



$$\varphi'(v_j) = \frac{a \cdot \exp(-av_j)}{\left[1 + \exp(-av_j)\right]^2}$$
$$= ay_j(1 - y_j)$$

Output neuron

$$\delta(v_j) = a \times e_j \times y_j(1 - y_j)$$

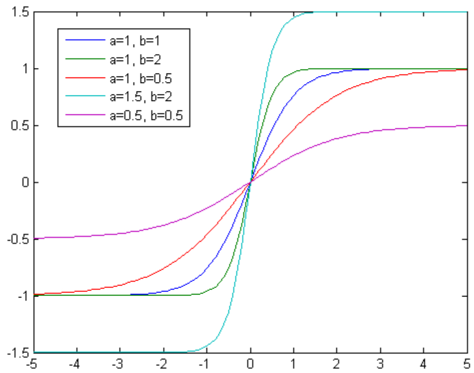
Hidden neuron

$$\delta(v_i) = a \times y_i(1 - y_i) \times \sum_j \delta_j w_{ji}$$

Activation Functions

- Hyperbolic Tangent Function

$$\varphi(v) = a\tanh(bv) = a\frac{e^{bv} - e^{-bv}}{e^{bv} + e^{-bv}} \qquad a > 0,\ b > 0$$



$$y = \varphi(v) = a\tanh(bv) = a\frac{e^{bv} - e^{-bv}}{e^{bv} + e^{-bv}}$$

$$\frac{\partial\varphi(v)}{\partial v} = a\frac{b(e^{bv} + e^{-bv})(e^{bv} + e^{-bv}) - (e^{bv} - e^{-bv})b(e^{bv} - e^{-bv})}{(e^{bv} + e^{-bv})^2}$$

$$= ab\frac{(e^{bv} + e^{-bv})^2 - (e^{bv} - e^{-bv})^2}{(e^{bv} + e^{-bv})^2}$$

$$= \frac{4ab}{(e^{bv} + e^{-bv})^2} = \frac{b}{a}\frac{2ae^{bv} \times 2ae^{-bv}}{(e^{bv} + e^{-bv})^2}$$

$$= \frac{b}{a}\frac{2ae^{bv}}{(e^{bv} + e^{-bv})}\frac{2ae^{-bv}}{(e^{bv} + e^{-bv})}$$

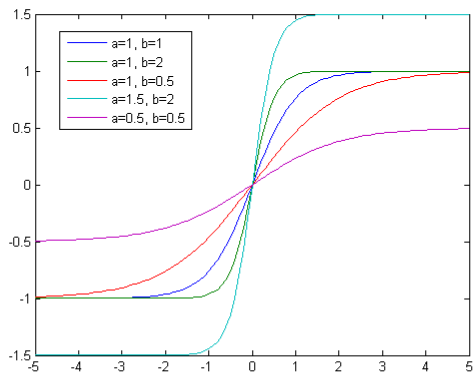
$$= \frac{b}{a}\left(a + a\frac{e^{bv} - e^{-bv}}{e^{bv} + e^{-bv}}\right)\left(a - a\frac{e^{bv} - e^{-bv}}{e^{bv} + e^{-bv}}\right)$$

$$= \frac{b}{a}(a - y)(a + y)$$

Activation Functions

- Hyperbolic Tangent Function

$$\varphi(v) = a\tanh(bv) = a\frac{e^{bv} - e^{-bv}}{e^{bv} + e^{-bv}} \qquad a > 0,\ b > 0$$



$$\varphi'(v_j) = \frac{b}{a}(a - y_j)(a + y_j)$$

Output neuron

$$\delta(v_j) = \frac{b}{a}e_j(a - y_j)(a + y_j)$$

Hidden neuron

$$\delta(v_i) = \frac{b}{a}(a - y_i)(a + y_i)\times\sum_j \delta_j w_{ji}$$

Sequential *vs* Batch

- Sequential mode $\quad \Delta w_{ji} = -\eta\frac{\partial Err}{\partial w_{ji}} = -\frac{\eta}{N}\sum_{n=1}^{N} e_j(n)\frac{\partial e_j(n)}{\partial w_{ji}}$
- Batch mode

$$Err = \frac{1}{2N}\sum_{n \in samples}\ \sum_{\substack{j \in output \\ neurons}} e_{nj}^2$$

$$\Delta w_{ji} = -\eta\frac{\partial Err(t)}{\partial w_{ji}(t)} = \eta\sum_n e_{nj}\frac{\partial e_{nj}}{\partial w_{ji}}$$

Online, memory, implementation, stochastic

- Number of hidden neurons?