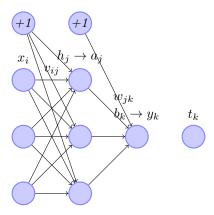
MLP

Wayne Wang

Principles of Backpropagation 1



Suppost we have a simple binary classification problem that needs to be solved by MLP. The structure of the MLP is shown as above which has 3 input nodes for input layer, 3 hidden nodes for one hidden layer and 1 output node

The activation function is sigmoid function: $g(x) = \frac{1}{1+e^{-x}}$ and $g'(x) = \frac{1}{1+e^{-x}}$ g(x)(1-g(x)).

Step 1: Going forward

$$h_{j} = \sum_{j} x_{i} \cdot v_{ij}$$

$$a_{j} = g(h_{j})$$

$$b_{k} = \sum_{k} a_{j} \cdot w_{jk}$$

$$(1)$$

$$(2)$$

$$(3)$$

$$a_j = g(h_j) \tag{2}$$

$$b_k = \sum_k a_j \cdot w_{jk} \tag{3}$$

$$y_k = g(b_k) (4)$$

Step 2: Compute Error

$$E(\mathbf{v}, \mathbf{w}) = \frac{1}{2} \sum_{k} (t_k - y_k)^2$$

$$= \frac{1}{2} \sum_{k} (t_k - g(b_k)^2)$$

$$= \frac{1}{2} \sum_{k} (t_k - g(\sum_{j} a_j \cdot w_{jk}))^2$$

$$= \frac{1}{2} \sum_{k} (t_k - g(\sum_{j} g(h_j) \cdot w_{jk})^2$$

$$= \frac{1}{2} \sum_{k} (t_k - g(\sum_{j} g(\sum_{i} x_i \cdot v_{ij}) \cdot w_{jk})^2$$
(5)

Step 3: Compute the derivative of Error function respect to \mathbf{w} and \mathbf{v}

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial}{\partial w_{jk}} \left(\frac{1}{2} \sum_{k} (t_k - y_k)^2 \right)
= \sum_{k} (t_k - y_k) \cdot \frac{\partial}{\partial w_{jk}} (t_k - y_k)
= \sum_{k} (t_k - y_k) \cdot \frac{\partial}{\partial w_{jk}} (-g(b_k))
= \sum_{k} (t_k - y_k) \cdot -g'(b_k) \cdot \frac{\partial}{\partial w_{jk}} \left(\sum_{j} a_j \cdot w_{jk} \right)
= \sum_{k} (y_k - t_k) \cdot y_k \cdot (1 - y_k) \cdot a_j$$
(6)

$$\frac{\partial E}{\partial v_{ij}} = \frac{\partial E}{\partial a_j} \cdot \frac{\partial a_j}{\partial v_{ij}}
= \frac{\partial}{\partial a_j} \left(\frac{1}{2} \sum_k (t_k - g(\sum_j a_j \cdot w_{jk}))^2 \right) \cdot \frac{\partial}{\partial v_{ij}} (g(\sum_i x_i \cdot v_{ij}))
= \left(\sum_k (y_k - t_k) \cdot y_k \cdot (1 - y_k) \cdot w_{jk} \right) \cdot a_j \cdot (1 - a_j) \cdot x_i$$
(7)

Step 4: Update weight \mathbf{w} and \mathbf{v}

$$w_j k = w_j k - \eta \frac{\partial E}{\partial w_{jk}} \tag{8}$$

$$v_i j = v_i j - \eta \frac{\partial E}{\partial v_{ij}} \tag{9}$$