Introduction to Neural Networks
Homework #1

機械所
張元睿
N16054629

October 11, 2016

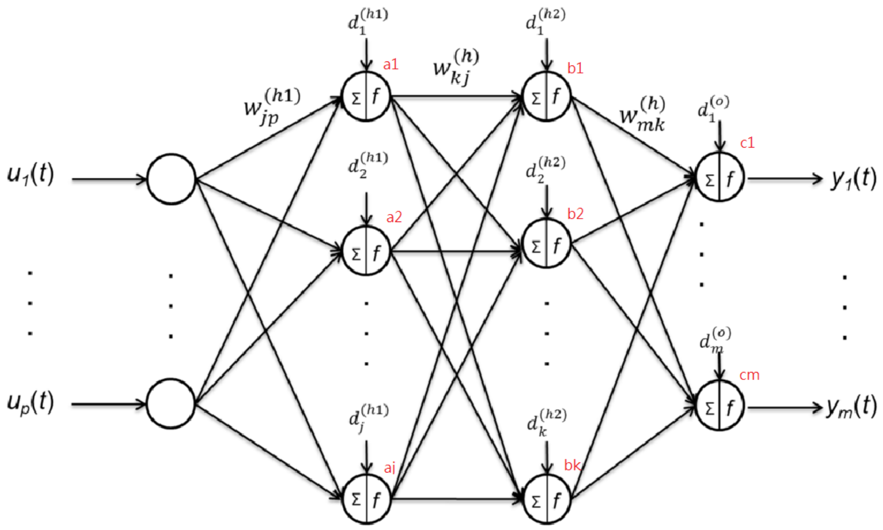

Fig. 1. Structure of three-layer feedforward neural network.

1. Forward path:

$$a_j = \sum_{\alpha=1\sim j,\ \beta=1\sim p} w_{\alpha\beta}^{(h1)} \times u_\beta(t) + d_\alpha^{(h1)}$$

$$Y_{a_j} = sigmoid(a_j) = \frac{1}{1+e^{-a_j}}$$

$$b_k = \sum_{\alpha=1\sim k,\ \beta=1\sim j} w_{\alpha\beta}^{(h2)} \times Y_{a_\beta} + d_\alpha^{(h2)}$$

$$Y_{b_k} = \tanh(a_j) = \frac{e^{b_k} - e^{-b_k}}{e^{b_k} + e^{-b_k}}$$

$$y_m(t) = c_m = \sum_{\alpha=1\sim m,\ \beta=1\sim k} w_{\alpha\beta}^{(h3)} \times y_{b_\beta} + d_\alpha^{(o)}$$

$$E_m(t) = \frac{1}{2} e_m^2(t)$$

$$e_m(t) = d_m(t) - y_m(t)$$

2. Backward propagation

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

$$
\begin{aligned}
w_{mk}^{(h_3)}(t+1) &= w_{mk}^{(h_3)}(t) + \Delta w_{mk}(t) \\
&= w_{mk}^{(h_3)}(t) - \eta \frac{\partial E_m(t)}{\partial w_{mk}^{(h_3)}(t)} \\
&= w_{mk}^{(h_3)}(t) - \eta \frac{\partial E_m(t)}{\partial e_m(t)} \frac{\partial e_m(t)}{\partial y_m(t)} \frac{\partial y_m(t)}{\partial c_m(t)} \frac{\partial c_m(t)}{\partial w_{mk}^{(h_3)}(t)} \\
&= w_{mk}^{(h_3)}(t) - \eta(d_m(t) - y_m(t))(-1)(1)(Y_{b_k}(t)) \\
&= w_{mk}^{(h_3)}(t) + \eta(d_m(t) - y_m(t))(Y_{b_k}(t))
\end{aligned}
$$

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

$$
\begin{aligned}
d_m^{(o)}(t+1) &= d_m^{(o)}(t) + \Delta d_m(t) \\
&= d_m^{(o)}(t) - \eta \frac{\partial E_m(t)}{\partial d_m^{(o)}(t)} \\
&= d_m^{(o)}(t) - \eta \frac{\partial E_m(t)}{\partial e_j(t)} \frac{\partial e_m(t)}{\partial y_m(t)} \frac{\partial y_m(t)}{\partial c_m(t)} \frac{\partial c_m(t)}{\partial d_m^{(o)}(t)} \\
&= d_m^{(o)}(t) - \eta(d_m(t) - y_m(t))(-1)(1)(1) \\
&= d_m^{(o)}(t) + \eta(d_m(t) - y_m(t))
\end{aligned}
$$

(c) Update rule for the weights of the second hidden neurons:

$$
\begin{aligned}
w_{kj}^{(h_2)}(t+1) &= w_{kj}^{(h_2)}(t) + \Delta w_{kj}(t) \\
&= w_{kj}^{(h_2)}(t) - \eta \frac{\partial E_m(t)}{\partial w_{kj}^{(h_2)}(t)} \\
&= w_{kj}^{(h_2)}(t) - \eta \frac{\partial E_m(t)}{\partial e_m(t)} \frac{\partial e_m(t)}{\partial y_m(t)} \frac{\partial y_m(t)}{\partial c_m(t)} \frac{\partial c_m(t)}{\partial Y_{b_k}(t)} \frac{\partial Y_{b_k}(t)}{\partial b_k(t)} \frac{\partial b_k(t)}{\partial w_{kj}^{(h_2)}(t)} \\
&= w_{kj}^{(h_2)}(t) - \eta(d_m(t) - y_m(t))(-1)(1)(w_{mk}^{(h_3)}(t))[1 - \tanh^2(b_k(t))](Y_{a_j}(t)) \\
&= w_{kj}^{(h_2)}(t) + \eta(d_m(t) - y_m(t))(w_{mk}^{(h_3)}(t))[1 - \tanh^2(b_k(t))](Y_{a_j}(t))
\end{aligned}
$$

(d) Update rule for the biases of the second hidden neurons:

$$
\begin{aligned}
d_k^{(h_2)}(t+1) &= d_k^{(h_2)}(t) + \Delta d_k(t) \\
&= d_k^{(h_2)}(t) - \eta \frac{\partial E_m(t)}{\partial d_k^{(h_2)}(t)} \\
&= d_k^{(h_2)}(t) - \eta \frac{\partial E_m(t)}{\partial e_j(t)} \frac{\partial e_m(t)}{\partial y_m(t)} \frac{\partial y_m(t)}{\partial c_m(t)} \frac{\partial c_m(t)}{\partial Y_{b_k}(t)} \frac{\partial Y_{b_k}(t)}{\partial b_k(t)} \frac{\partial b_k(t)}{\partial d_j^{(h_1)}(t)} \\
&= d_k^{(h_2)}(t) - \eta(d_m(t) - y_m(t))(-1)(1)(w_{mk}^{(h_3)}(t))[1 - \tanh^2(b_k(t))](1) \\
&= d_k^{(h_2)}(t) + \eta(d_m(t) - y_m(t))(w_{mk}^{(h_3)}(t))[1 - \tanh^2(b_k(t))]
\end{aligned}
$$

(e) Update rule for the weights of the first hidden neurons:

$$
\begin{aligned}
w_{jp}^{(h_1)}(t+1) &= w_{jp}^{(h_1)}(t) + \Delta w_{jp}(t) \\
&= w_{jp}^{(h_1)}(t) - \eta \frac{\partial E_m(t)}{\partial w_{jp}^{(h_1)}(t)} \\
&= w_{jp}^{(h_1)}(t) - \eta \frac{\partial E_m(t)}{\partial e_m(t)} \frac{\partial e_m(t)}{\partial y_m(t)} \frac{\partial y_m(t)}{\partial c_m(t)} \frac{\partial c_m(t)}{\partial Y_{b_k}(t)} \frac{\partial Y_{b_k}(t)}{\partial b_k(t)} \frac{\partial b_k(t)}{\partial Y_{aj}(t)} \\
&\qquad \frac{\partial Y_{aj}(t)}{\partial a_j(t)} \frac{\partial a_j(t)}{\partial w_{jp}^{(h_1)}(t)} \\
&= w_{jp}^{(h_1)}(t) - \eta(d_m(t) - y_m(t))(-1)(1)(w_{mk}^{(h_3)}(t))[1 - \tanh^2(b_k(t))] \\
&\qquad (w_{kj}^{(h_2)}(t))(\frac{1}{1+e^{-a_j(t)}})(\frac{e^{-a_j(t)}}{1+e^{-a_j(t)}})(u_p(t)) \\
&= w_{jp}^{(h_1)}(t) + \eta(d_m(t) - y_m(t))(w_{mk}^{(h_3)}(t))[1 - \tanh^2(b_k(t))] \\
&\qquad (w_{kj}^{(h_2)}(t))(\frac{1}{1+e^{-a_j(t)}})(\frac{e^{-a_j(t)}}{1+e^{-a_j(t)}})(u_p(t))
\end{aligned}
$$

(f) Update rule for the biases of the first hidden neurons:

$$
\begin{aligned}
d_j^{(h_1)}(t+1) &= d_j^{(h_1)}(t) + \Delta d_j(t) \\
&= d_j^{(h_1)}(t) - \eta \frac{\partial E_m(t)}{\partial d_j^{(h_1)}(t)} \\
&= d_j^{(h_1)}(t) - \eta \frac{\partial E_m(t)}{\partial e_j(t)} \frac{\partial e_m(t)}{\partial y_m(t)} \frac{\partial y_m(t)}{\partial c_m(t)} \frac{\partial c_m(t)}{\partial Y_{b_k}(t)} \frac{\partial Y_{b_k}(t)}{\partial b_k(t)} \frac{\partial b_k(t)}{\partial Y_{a_j}(t)} \\
&\qquad \frac{\partial Y_{aj}(t)}{\partial a_j(t)} \frac{\partial a_j(t)}{\partial d_j^{(h_1)}(t)} \\
&= d_j^{(h_1)}(t) - \eta(d_m(t) - y_m(t))(-1)(1)(w_{mk}^{(h_3)}(t))[1 - \tanh^2(b_k(t))] \\
&\qquad (w_{kj}^{(h_2)}(t))(\frac{1}{1+e^{-a_j(t)}})(\frac{e^{-a_j(t)}}{1+e^{-a_j(t)}})(1) \\
&= d_j^{(h_1)}(t) + \eta(d_m(t) - y_m(t))(w_{mk}^{(h_3)}(t))[1 - \tanh^2(b_k(t))] \\
&\qquad (w_{kj}^{(h_2)}(t))(\frac{1}{1+e^{-a_j(t)}})(\frac{e^{-a_j(t)}}{1+e^{-a_j(t)}})
\end{aligned}
$$