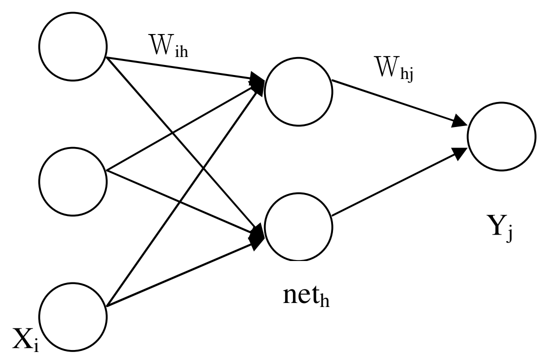
Process computation steps

BPN 使用步驟說明



1. 計算 Input -->Hidden layer

$$net_h = \sum_i W_{ih} \bullet X_i - \Theta_h$$

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

2. 計算 Hidden -->Output layer

$$net_j = \sum_i W_{hj} \bullet H_h - \Theta_j$$

$$Y_j = f(net_h) = \frac{1}{1+e^{-net_j}}$$

3. 計算 total error &difference for correction

$$\delta_j = Y_j(1-Y_j)(T_j - Y_j)$$

$$\delta_h = H_h(1-H_h)\sum_i W_{hj}\ \delta_j$$

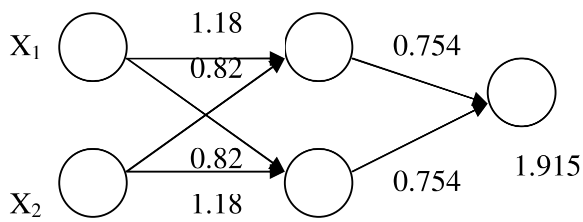
4. $\Delta W_{hj} = \eta\ \delta_j\ H_h$ $\quad \Delta\Theta_j = -\eta\ \delta_j$

$\Delta W_{ih} = \eta\ \delta_h\ X_i$ $\quad \Delta\Theta_h = -\eta\ \delta_h$

5. update weights

$W_{hj} = W_{hj} + \Delta W_{hj}$ ，$W_{ih} = W_{ih} + \Delta W_{ih}$

$\Delta W_{12} = \eta\ \delta_1 X_1$ =(10)(-0.018)(-1)=0.18
$\Delta W_{21} = \eta\ \delta_1 X_2$ =(10)(-0.018)(-1)=0.18
$\Delta \Theta_1 = -\eta\ \delta_1$ = -(10)(-0.018)=0.18
說明 p-415　　第一次修正後
討論



net_h

1. # of Hidden nodes ↑則 converge 慢，但可達最小誤差值
2. 一般 # of Hidden nodes=(Input nodes + Output nodes)/2
   或 Hidden nodes=(Input nodes + Output nodes)$^{1/2}$
3. Hidden 以 1-2 個 layer 最恰當
   因為 1-2 個 layer 以足夠反應現況
   太多 layer 會太複雜
4. η 約為 05.-1.0 為叫好的 learning rate

The gradient steepest descent method

Recall: $net_j^n = \sum_j WijA_i^{n-1} - \theta_j$

We want the computed output and expected output getting close to ∅

E=(1/2) $\sum_j (Tj - Aj)^2$ => ΔW_{ij}=- η $\dfrac{\partial E}{\partial Wij}$

Therefore, we want to obtain $\dfrac{\partial E}{\partial Wij}$ so that we can update weights to improve the network

$$\frac{\partial E}{\partial Wij} = \left(\frac{\partial E}{\partial net_j^n}\right)\left(\frac{\partial net_j^n}{\partial Wij}\right)$$

$$= \left(\frac{\partial E}{\partial A_j^n}\right)\left(\frac{\partial A_j^n}{\partial net_j^n}\right)\left(\frac{\partial net_j^n}{\partial Wij}\right)$$

(3) (2) (1)

For (1)

$$\frac{\partial net_j^n}{\partial Wij} = \frac{\partial(\sum_k WkjA_k^{n-1} - \theta j)}{\partial Wij} = A_i^{n-1}$$

For (2)

$$\frac{\partial A_j^n}{\partial net_j^n} = \frac{\partial f(net_j^n)}{\partial net_j^n} = f'(net_j^n)$$

For (3): when n is the output layer

$$\frac{\partial E}{\partial A_j^n} = \frac{\partial[1/2\sum_k (T_k - A_k^n)^2]}{\partial A_j^n} = -(Tj - A_j^n)$$

when n is the hidden layer

$$\frac{\partial E}{\partial A_j^n} = \sum_k \left(\frac{\partial E}{\partial net_k^{n+1}}\right)\left(\frac{\partial net_k^{n+1}}{\partial A_j^n}\right) = -\sum_k \delta_k^{n+1} W_{jk}$$

From (1)(2)(3) we have two types of values:

When n is output layer

$$\frac{\partial E}{\partial Wij} = -(Tj - A_j^n)\, f'(net_j^n)\, A_i^{n-1} \qquad (代入(B))$$

or $= -\delta_j^n A_i^{n-1}$ (代入(A))

we get $\delta_j^n = (T_j - A_i^{n-1}) f'(net_j^n)$

When n is hidden layer

$$\frac{\partial E}{\partial Wij} = -[\sum_k \delta_k^{n+1} W_{jk}]\, f'(net_j^n)\, A_i^{n-1} \qquad (代入(B))$$

or $= -\delta_j^n A_i^{n-1}$ (代入(A))

we get $\delta_j^n = [\sum_k \delta_k^{n+1} W_{jk}] f'(net_j^n)$

$$\left.\begin{aligned}
&\Rightarrow \frac{\partial E}{\partial Wij} = - -\delta_j^n A_i^{n-1} \\
&\Rightarrow \Delta W_{ij} = +\eta\ \delta_j^n A_i^{n-1} \\
&\Rightarrow \Theta = -\eta\ -\delta_j^n
\end{aligned}\right\}
\qquad
\begin{aligned}
&W_{ij} = W_{ij} + \Delta W_{ij} \\
&\Theta_j = \Theta_j + \Delta \Theta_j
\end{aligned}$$

(*) $f(net_j^n) = \dfrac{1}{1+e^{-net_j}} = (1+e^{-netj})^{-1}$

$f'(net_j^n) == [(1+e^{-netj})^{-1}]' = [-(1+e^{-netj})^{-2}][-(e^{-netj})]$

$= \dfrac{e^{-netj}}{(1+e^{-net_j})^2} = \dfrac{e^{-netj}}{(1+e^{-net_j})} \cdot \dfrac{1}{1+e^{-net_j}}$

$= f(net_j)(1-f(net_j))$

$$\delta_j^n = \begin{cases} (T_j - Y_j)Y_j(1-Y_j) & \text{if n is output layer} \\ [\sum_k \delta_j^{n+1} W_{ik}] \bullet H_j(1-H_j) & \text{if n is hidden layer} \end{cases}$$

Learning computation

1. $\text{net}_h = \sum_i W_{ih} \bullet X_i - \theta_h$ Compute value of the hidden layer

   $H_h = f(\text{net}_h) = \dfrac{1}{1+e^{-net_h}}$

2. $\text{net}_j = \sum_h W_{hj} \bullet H_h - \theta_j$ Compute value of the output layer

   $Y_j = f(\text{net}_j) = \dfrac{1}{1+e^{-net_j}}$

3. $\delta_j = Y_j(1 - Y_j)(T_j - Y_j)$ Compute the value difference for correction

   $\delta_h = H_h(1 - H_h)\sum_j W_{hj}\delta_j$

4. $\Delta W_{hj} = \eta \ \delta_j H_h$ $\Delta \Theta_j = -\eta \ \delta_j$ Compute the value to be updated

   $\Delta W_{ih} = \eta \ \delta_h X_i$ $\Delta \Theta_h = -\eta \ \delta_h$

5. $W_{hj} = W_{hj} + \Delta W_{hj}$ $\Theta_j = \Theta_j + \Delta \Theta_j$

   $W_{ih} = W_{ih} + \Delta W_{ih}$ $\Theta_h = \Theta_h + \Delta \Theta_h$