-Artificial Neural Network-
Chapter 4 Adaline & Madaline

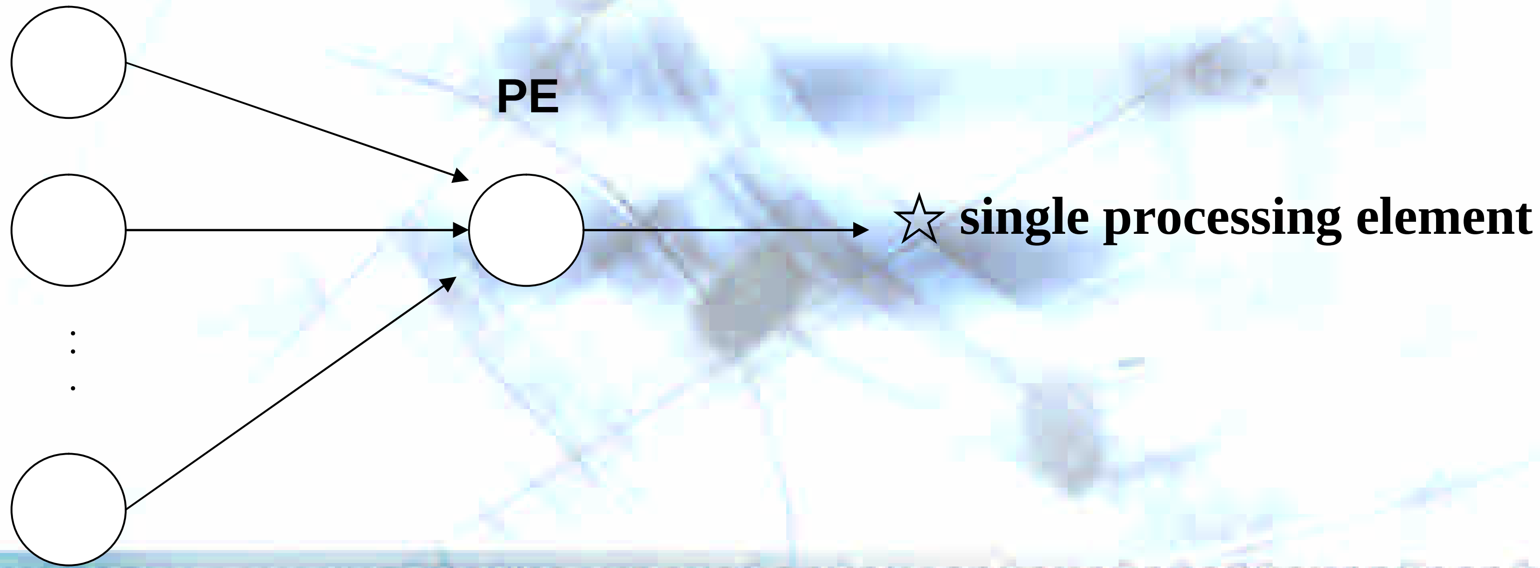
朝陽科技大學
資訊管理系
李麗華 教授

Outline

- ADALINE
- MADALINE
- Least-Square Learning Rule
- The proof of Least-Square Learning Rule

ADALINE (1/3)

- ADALINE: (Adaptive Linear Neuron) 1959 by Bernard Widrow



ADALINE (2/3)

- Method :① The value in each unit must +1 or –1
  (perceptron 爲1)
  - net = $\sum X_i W_i$

$$X_0 = 1 \therefore \text{net} = W_0 + W_1 X_1 + W_2 X_2 + \cdots + W_n X_n$$

$$Y = \begin{cases} 1 & \text{net} \geq 0 \\ & \text{if} \\ \boxed{-1} & \text{net} < 0 \end{cases}$$

different from perception's transfer function

ADALINE (3/3)

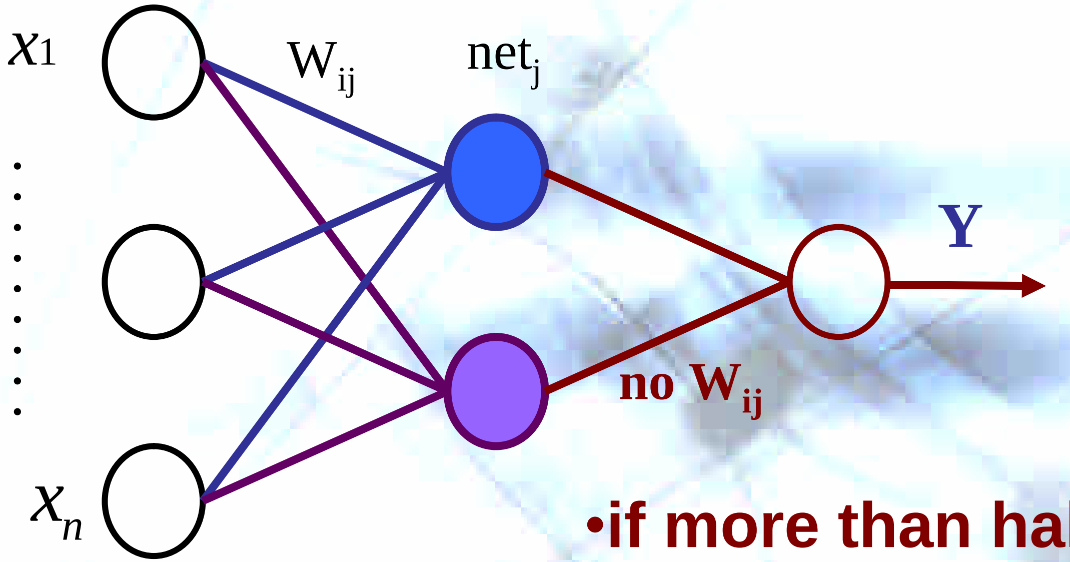
$\Delta W_i = \eta$ (T-Y) X_i,   T = expected output

$W_i = W_i + \Delta W_i$

- ADALINE can solve only linear problem(the limitation)

MADALINE

- MADALINE： It is composed of many ADALINE （Multilayer Adaline.）



•if more than half of $net_j \geq 0$, then output ＋1,otherwise, output －1

•After the second layer, the majority vote is used.

Least-Square Learning Rule (1/6)

- $X_j = (x_0, x_1, \cdots, x_n)^t$, (i.e. $X_j = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}$) $1 \le j \le L$

字母小寫 ***j*** **:代表第 *j* 組input pattern**

t **:代表向量轉置 (transpose)**

L **: 代表input pattern數量**

- $W = (w_0, w_1, \cdots, w_n)^t$, (i.e. $W = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{pmatrix}$)

$$Net_j = W^t X_j = \sum_{i=0}^{n} w_i x_i = w_0 x_0 + w_1 x_1 + \ldots + w_n x_n$$

Least-Square Learning Rule (2/6)

- By applying the least-square learning rule the weights is ：

$$RW^* = P$$

$$W^* = R^{-1}P \quad \text{where} \begin{cases} \text{R: Correlation Matrix} \\ R = \dfrac{R'}{L}, R' = R_1' + R_2' + \ldots + R_L' = \sum_{j=1}^{L} X_j X_j^t \\ P^t = \dfrac{T_j X_j^t}{L} \end{cases}$$

Least-Square Learning Rule (3/6)

$$\text{Example}: X_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}_{T_1=1} \quad X_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}_{T_2=1} \quad X_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}_{T_3=-1}$$

	X_1	X_2	X_3	T_j
X_1	1	1	0	1
X_2	1	0	1	1
X_3	1	1	1	-1

Least-Square Learning Rule (4/6)

- *Sol.* 先算R

$$\left.\begin{aligned}
R_1' &= \begin{pmatrix}1\\1\\0\end{pmatrix}(110) = \begin{pmatrix}1&1&0\\1&1&0\\0&0&0\end{pmatrix}\\
R_2' &= \begin{pmatrix}1\\0\\1\end{pmatrix}(101) = \begin{pmatrix}1&0&1\\0&0&0\\1&0&1\end{pmatrix}\\
R_3' &= \begin{pmatrix}1\\1\\1\end{pmatrix}(111) = \begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix}
\end{aligned}\right\} R = \frac{1}{3}\begin{pmatrix}3&2&2\\2&2&1\\2&1&2\end{pmatrix} \Rightarrow \begin{pmatrix}1&2/3&2/3\\2/3&2/3&1/3\\2/3&1/3&2/3\end{pmatrix}$$

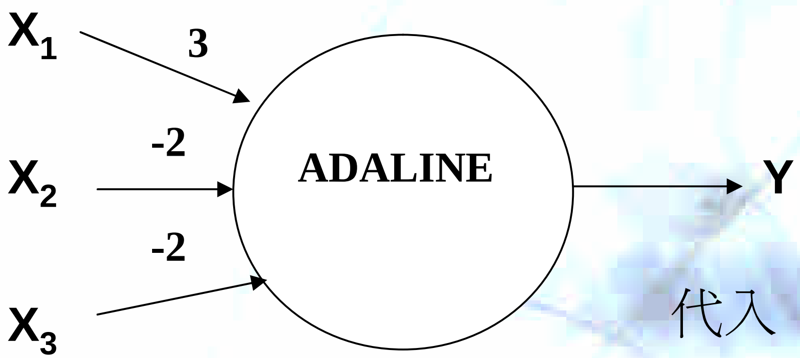
Least-Square Learning Rule (5/6)

$$\left.\begin{aligned} P_1^t &= 1\cdot(1,1,0) = (110) \\ P_2^t &= 1\cdot(1,0,1) = (101) \\ P_3^t &= -1\cdot(1,1,1) = (-1-1-1) \end{aligned}\right\} P^t = \frac{1}{3}(100) = \left(\frac{1}{3},0,0\right)$$

$$R\cdot W^* = P \Rightarrow \begin{pmatrix} 1 & 2/3 & 2/3 \\ 2/3 & 2/3 & 1/3 \\ 2/3 & 1/3 & 2/3 \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} 3W_1 + 2W_2 + 2W_3 = 1 \\ 2W_1 + 2W_2 + W_3 = 0 \\ 2W_1 + W_2 + 2W_3 = 0 \end{cases} \Rightarrow \begin{matrix} W_1 = 3 \\ W_2 = -2 \\ W_1 = -2 \end{matrix}$$

Least-Square Learning Rule (6/6)

- Verify the net:



代入（1,1,0） $net=3X_1-2X_2-2X_3=1$ Y=1 ok

代入（1,0,1） $net=3X_1-2X_2-2X_3=1$ Y=1 ok

代入（1,1,1） $net=3X_1-2X_2-2X_3=-1$ Y=-1 ok

$<\varepsilon_k^2>=0$ **this is minimum ==> best solution**

(*)同學回家應了解反矩陣的計算方法

Proof of Least Square Learning Rule(1/3)

- We use Least Mean Square Error to ensure the minimum total error. As long as the total error approaches zero, the best solution is found. Therefore, we are looking for the minimum of $\langle \varepsilon_k^2 \rangle$.
- Proof:

$$\textbf{mean} \rightarrow <\varepsilon^2> = \frac{1}{L}\sum_{k=1}^{L}\varepsilon_k^{\ 2} = \frac{1}{L}\sum_{k=1}^{L}(T_k - Y_k)^2 = \frac{1}{L}\sum_{k=1}^{L}(T_k^{\ 2} - 2T_kY_k + Y_k^{\ 2})$$

$$= \frac{1}{L}\sum_{k=1}^{L}T_k^{\ 2} - \frac{2}{L}\sum_{k=1}^{L}T_kY_k + \frac{1}{L}\sum_{k=1}^{L}Y_k^{\ 2} \textbf{let} < T_{\mathbf{k}}^{\mathbf{2}} > \textbf{represents mean}$$

$$= < T_{\mathbf{k}}^{\mathbf{2}} > - \frac{2}{L}\sum_{k=1}^{L}T_kY_k + \frac{1}{L}\cdot[W^t(\sum_{k=1}^{L}X_kX_k^{\ t})W]$$

Proof of Least Square Learning Rule(2/3)

$$ps :==> \sum_{k=1}^{1} Y_k^{\ 2} = \sum_{k=1}^{L} (\sum_{i=1}^{n} w_i x_{ik})^2 = \sum_{k=1}^{L} (W^t X_k)^2 = \sum_{k=1}^{L} (W^t X_k)(X_k^{\ t} W)$$

$$= W^t (\sum_{k=1}^{L} X_k \cdot X_k^{\ t}) \cdot W$$

承上頁

$$=< T_k^2 > - \frac{2}{L} \sum_{k=1}^{L} T_k Y_k + W^t [\frac{1}{L} (\sum_{k=1}^{L} X_k X^t)] W$$

$$=< T_k^2 > - \frac{2}{L} \sum_{k=1}^{L} T_k (W^t X_k) + W^t \langle X_k X_k^{\ t} \rangle W$$

$$=< T_k^2 > -2[\frac{1}{L} \sum_{k=1}^{L} (T_k X_k^{\ t}) W] + W^t \langle X_k X_k^{\ t} \rangle W$$

$$=< T_k^2 > -2 \langle T_k X_k^{\ t} \rangle W + W^t \langle X_k X_k^{\ t} \rangle W$$

令此項爲**R**

Proof of Least Square Learning Rule(3/3)

$*$ **Let** $R_k = X_k X_k{}^t$ **, i.e.,** R_k **is a n $\times$ n matrix, also called Correlation Matrix.**

$*$ **Let** $\mathbf{R'} = \mathbf{R_1} + \mathbf{R_2} + \ldots + \mathbf{R_K} + \ldots + \mathbf{R_L} = \sum_{k=1}^{L} T_k X_k{}^t$

$*$ **Let** $\mathbf{R} = \mathbf{R'}/\mathbf{L}$ **(i.e. mean of** R' **)** $==> \mathrm{R} =< X_k X_k^t >$

承上頁

$$=< T_k{}^2 > -2\langle T_k X_k{}^t \rangle W + W^t R W$$

★ **Find W* such that** $< \varepsilon_k^2 >$ **is minimal**

$$\frac{\partial \langle \varepsilon_k{}^2 \rangle}{\partial W} = [\langle T_k{}^2 \rangle - 2\langle T_k X_k{}^t \rangle W + W^t R W]'$$

$$= -2 < T_k X_k > +2RW = 2RW - 2P \qquad \textbf{Let P} = \langle T_k X_k{}^t \rangle$$

if $\frac{\partial \langle \varepsilon_k{}^2 \rangle}{\partial W} = 0 \Rightarrow \mathbf{2}RW^* - \mathbf{2}P = 0$ 即 $W^* = R^{-1}P$ 或 $RW^* = P$