LMS Algorithm.

(1) The performance index for ADLINE algorithm is MSE.

$$F(x) = E[e^2]$$
$$= E[(t-a)^2]$$
$$= E[(t - x^T z)^2] \quad \text{——— (1)}$$

because, we know

$$a = W^T p + b$$

which can be written as:-

$$\underline{a = x^T z.}$$

eqn.(1) can be written as:-

$$F(x) = E\left[t^2 - 2t x^T z + (x^T z \cdot z^T x)\right]$$
$$= E[t^2] - 2x^T E[tz] + x^T E[zz^T] x. \quad \text{——— (2)}$$

This can be further written as follows:

$$F(x) = c - 2x^T h + x^T R x \quad \text{——— (3)}$$

where,

$$c = E[t^2],$$
$$h = E[tz],$$
$$R = E[zz^T].$$

h gives cross-correlation between the i/p vector and associated target, R is the i/p correlation matrix.

~~Now~~ the general form of quadratic function is

$$F(x) = \frac{1}{2} x^T A x + d^T x + C \quad \text{--- (4)}$$

Now, if we compare eqn.(3) with (4) then we have the following:—

$$F(x) = C + d^T x + \frac{1}{2} x^T A x \quad \text{--- (5)}$$

So, $d = -2h$ and ~~B~~ $A = 2R$.

A is a hessian matrix and if the eigenvalues of the hessian matrix are positive then we ~~have~~ ~~our~~ surely get global minimum.

$$\underset{(\text{correlation})}{R} \begin{cases} +ve \rightarrow \text{unique global optimum/minimum value.} \\ -ve \rightarrow \text{no minimum/weak minimum.} \end{cases}$$

calculation of the ~~derivative~~ gradient of eqn.(5)

$$\nabla F(x) = \nabla\left(c + d^T x + \frac{1}{2} x^T A x\right)$$

$$= d + AX$$

$$= -2h + 2RX \quad \text{--- (6)}$$

Now to find the stationary pt. $F(x) = 0$

hence

$$-2h + 2RX = 0$$

$$\Rightarrow \boxed{X = R^{-1} h.} \quad \text{--- (7)}$$

→ The goal is to locate the minimum point and it is only possible if performance index is known.

→ We can find minimum point directly from eqn. 7 ~~too~~ if the h and R value is possible to calculate.

→ If it is not possible then gradient ~~of~~ calculation of eqn. 6 and may be steepest descent algorithm ~~may~~ ~~to~~ help us to find the minimum. In real life problem it is not desirable or convenient to calculate h and R. For this reason we will use an approximate steepest descent algorithm.

→ MSE : $F(x) = \left(t(k) - a(k)\right)^2$

$$= e^2(k)$$

$e(k)$ is the ~~kth~~ error observed in the k^{th} iteration.

Gradient in each iteration :

$$\nabla F(x) = \nabla e^2(k)$$

→ As the weight and bias are the two parameter which adjusted in each iteration. The error calculation ~~is~~ depends on the weight and bias. So, the derivatives ~~adjust~~ weight and bias is calculated in terms of weight and bias to find the ~~two~~ unique weight and bias matrix which provides minimum error.

→ Hence,

$$\nabla e^2(K) = \frac{\partial e^2(K)}{\partial w_j}$$

$$= 2e(K) \frac{\partial e(K)}{\partial w_j}$$

and

$$\nabla e^2(K) = \frac{\partial e^2(K)}{\partial b}$$

$$= 2e(K) \frac{\partial e(K)}{\partial b}.$$

Now,

$$\frac{\partial e(K)}{\partial w_j} = \frac{\partial}{\partial w_j}\left[(t(K) - a(K))\right]$$

$$= \frac{\partial}{\partial w_j}\left[t(K) - (w^T p(K) + b)\right]$$

$$= -p_j(K)$$

Similarly,

$$\frac{\partial e(K)}{\partial b} = \frac{\partial}{\partial b}\left[t(K) - a(K)\right]$$

$$= \frac{\partial}{\partial b}\left[t(K) - (w^T p(K) + b)\right]$$

$$= -1$$

So,

$$\nabla F(x) = \nabla e^2(k)$$
$$= -2e(k)z(k) \quad (13)$$

Now, according to steepest descent algorithm

$$x_{k+1} = x_k - \alpha \nabla F(x) \quad (14)$$

if we substitute eqn. (13) into (14) then

$$x_{k+1} = x_k + 2\alpha e(k)z(k) \quad (15)$$

eqn. (15) can be written as :–

$$w(k+1) = w_k + 2\alpha e(k)p(k)$$

and

$$b(k+1) = b_k + 2\alpha e(k)$$

$p_1 = [1, 1]$ $\quad t_1 = [-1, -1]$

$p_2 = [1, 2]$ $\quad t_2 = [-1, -1]$

$p_3 = [2, -1]$ $\quad t_3 = [-1, 1]$

$p_4 = [2, 0]$ $\quad t_4 = [-1, 1]$

$p_5 = [-1, 2]$ $\quad t_5 = [1, -1]$

$p_6 = [-2, 1]$ $\quad t_6 = [1, -1]$

$p_7 = [-1, -1]$ $\quad t_7 = [1, 1]$

$p_8 = [-2, -2]$ $\quad t_8 = [1, 1]$

$w = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$\alpha = 0.4$

$tf =$ Linear

$\not\equiv$

log/7 Sigmoidi

Iteration :- 1

$$a(0) = \text{linear}(W(0)\,p(0) + b(0))$$

$$= \text{linear}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$e(0) = t(0) - a(0) = \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -3 \\ -3 \end{bmatrix}$$

$$W(1) = W(0) + 2\alpha\, e(0)\, p^T(0)$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 2(0.04) \begin{bmatrix} -3 \\ -3 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.76 & -0.24 \\ -0.24 & 0.76 \end{bmatrix}$$

$$b(1) = b(0) + 2\alpha\, e(0)$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2(0.04) \begin{bmatrix} -3 \\ -3 \end{bmatrix}$$

$$= \begin{bmatrix} 0.76 \\ 0.76 \end{bmatrix}$$

Iteration-2

$$a(1) = \text{linear}(W(1)\,p(1) + b(1))$$

$$= \text{linear}\left(\begin{bmatrix} 0.76 & -0.24 \\ -0.24 & 0.76 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0.76 \\ 0.76 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 1.04 \\ 2.04 \end{bmatrix}$$

$$e(1) = t(1) - a(1) = \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} 1.04 \\ 2.04 \end{bmatrix} = \begin{bmatrix} -2.04 \\ -3.04 \end{bmatrix}$$

$$W(2) = W(1) + 2\alpha\, e(1)\, p^T(1)$$

$$= \begin{bmatrix} 0.76 & -0.24 \\ -0.24 & 0.76 \end{bmatrix} + 2.(0.04) \begin{bmatrix} -2.04 \\ -3.04 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5968 & -0.5664 \\ -0.4832 & 0.2736 \end{bmatrix}$$

$$b(2) = b(1) + 2\alpha e(1)$$

$$= \begin{bmatrix} 0.76 \\ 0.76 \end{bmatrix} + 2(0.04)\begin{bmatrix} -2.04 \\ -3.04 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5968 \\ 0.5968 \end{bmatrix}$$