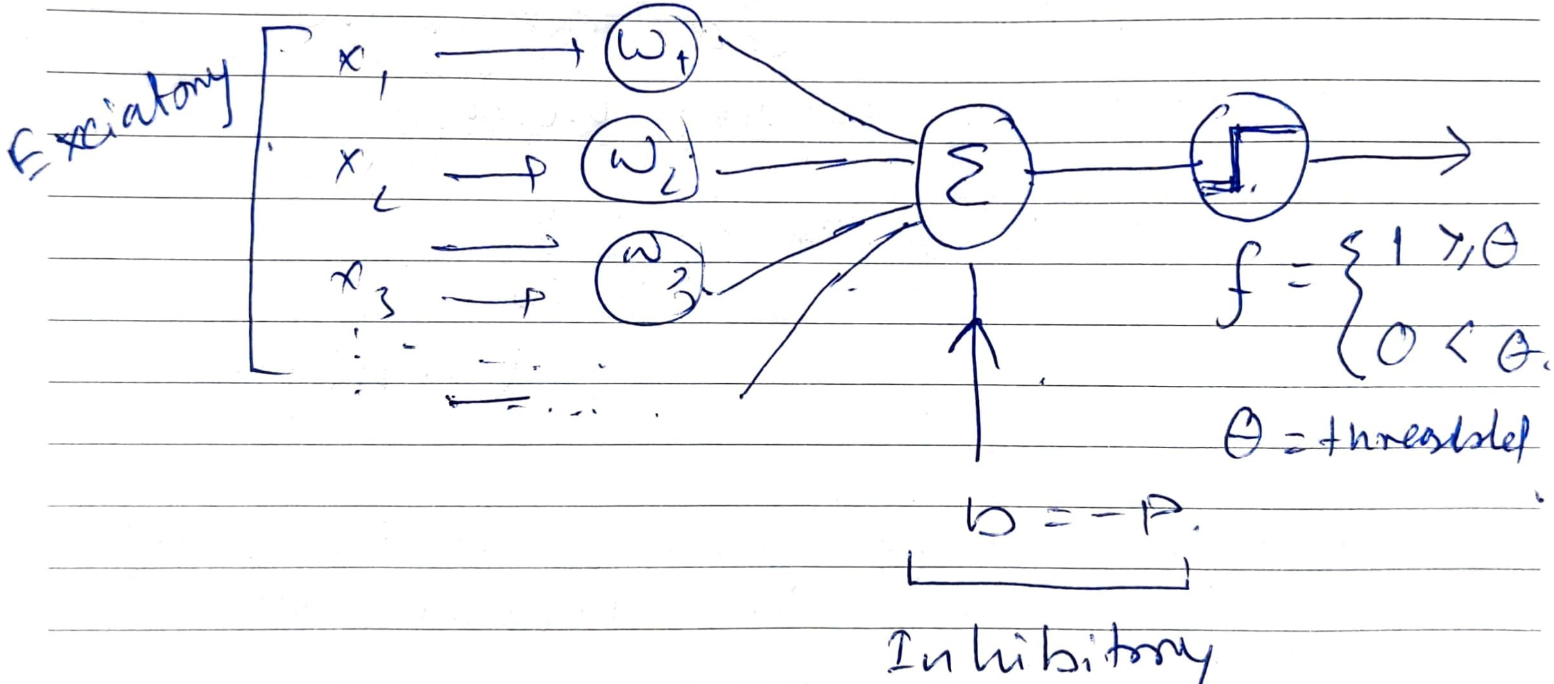
Deep Learning:

① McCulloch - Pitts Neuron [1943]



* No perticular training algorithm.

② Hebbian Learning [1949]

"Neurons fire together, wire together"

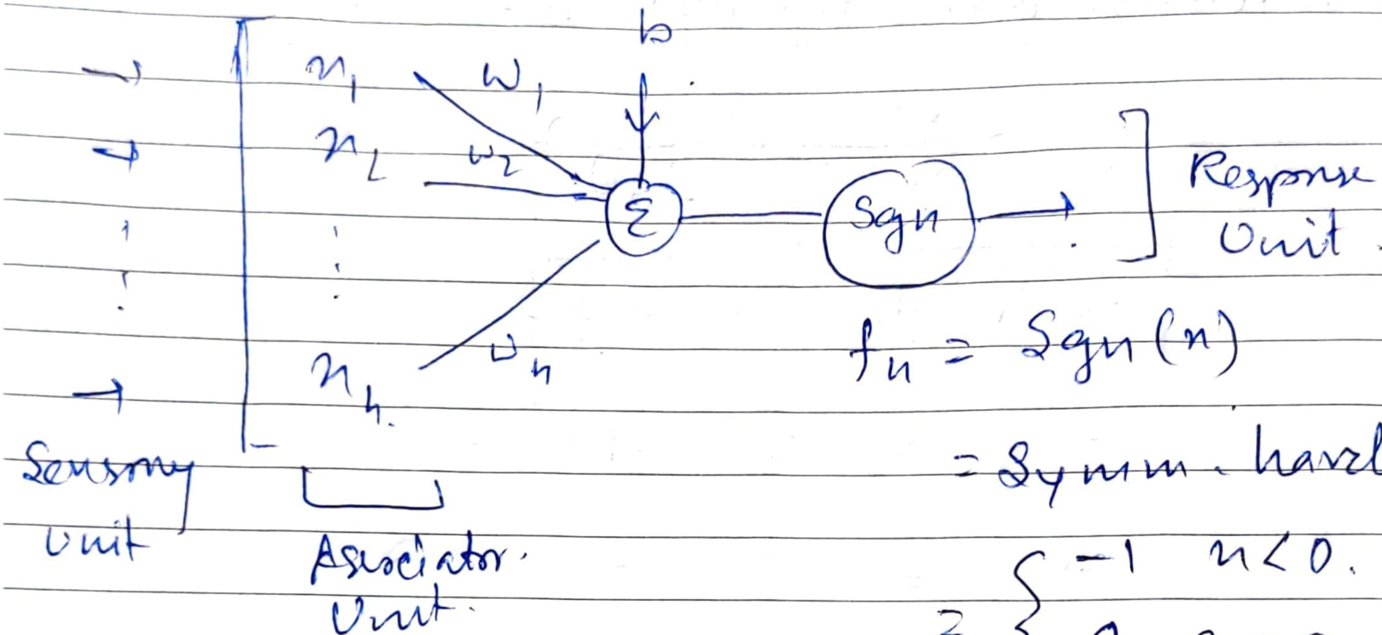
X	Y	ΔW.
+	+	+
−	−	+
+	−	−
−	+	−

Weight & bias updation:

$$w_i(new) = w_i(old) + x_i \cdot y$$

$$b_i(new) = b(old) + y.$$

③ - Rosenblatt's Perceptron : [1958]



$$f_n = \text{Sgn}(n)$$

$$= \text{Symm. hardlim}$$

$$= \begin{cases} -1 & n < 0 \\ 0 & n = 0 \\ 1 & n > 0 \end{cases}$$

Perceptron Learning Theorem

$$w_i(\text{new}) = w_i(\text{old}) + \alpha\,[t - y]\,x_i$$

$$t = \begin{cases} +1 & x \in C_1 \\ -1 & x \in C_2 \end{cases}$$

* C_1 & C_2 are two class for binary classification

If solu^n exist this learning will generate correct response for all training pattern within finite number of steps.

(4) Adaptive Linear Neuron (Adaline)

* Consists of single linear unit.

<u>linear unit</u>: Unit with Linear activation function

ie. $net = \sum x_i w_i + b$

* Can be trained by Delta / Widrow-Hoff rule / LMS Learning Algo.
* Activation (generally): Bypolar Step fun. / Sigmoid / Hyperbolic Tan

<u>Delta / Widrow-Hoff Rule</u>:

$$\Delta W = \alpha (t - y_{in}) x_i$$

Weight & Bias Updation:

$$w_i(new) = w_i(old) + \alpha (t - y_{in}) x_i$$

$$b(new) = b(old) + \alpha (t - y_{in})$$

for non linear activation functn.,

$$\Delta W_{ji} = \alpha (t_j - y_j) g'(n_j) x_i \quad n_j = net$$

5. Least Mean Square:

Loss (L) = Mean Sq Err.

ie $L(w,b) = E[t(k) - y(k)]^2$

$= E\, e(k)^2 = E[e(k)^T \cdot e(k)]$

$\nabla \hat{L}(w,b) = \nabla e(k)^2$

* Approx LMS without expectation of e.

$$\frac{\partial \hat{L}(w,b)}{\partial w} = \frac{\partial e(k)^2}{\partial w}$$

$$= 2e(k) \frac{\partial e(k)}{\partial w}$$

03 Sunday

$$\frac{\partial \hat{L}(w,b)}{\partial b} = \frac{\partial e(k)^2}{\partial b} = 2e(k) \cdot \frac{\partial e(k)}{\partial b}$$

$$\frac{\partial e(k)}{\partial w} = \frac{\partial}{\partial w}[t(k) - y(k)]$$

$$y(k) = \sigma\big[\underbrace{w^T x(k) + b}_{z}\big]$$

$$\frac{\partial}{\partial w}\Big[t(k) - \sigma\big[w^T x(k) + b\big]\Big]$$

$$= -\frac{\partial}{\partial w}\,\sigma.\big[w^T x(k) + b\big]$$

$$= -\frac{\partial \sigma(z)}{\partial \cancel{w}\, z}\cdot\frac{\partial}{\partial w}\big[w^T x(k) + b\big]$$

$$= -\frac{\partial \sigma(z)}{\partial \cancel{w}\, z}\, x(k) = -\sigma'(z)\, x(k)$$

$$\frac{\partial e(k)}{\partial b} = \frac{\partial}{\partial b}\big[t(k) - \sigma\big(w^T x(k) + b\big)\big]$$

$$= -\frac{\partial \sigma(z)}{\partial \cancel{b}\, z}\,\frac{\partial}{\partial b}\big[w^T x(k) + b\big]$$

$$\frac{\partial \sigma}{\partial \cancel{b}} = -\frac{\partial \sigma(z)}{\partial z} = -\sigma'(z)$$

A/c Steepest Decent,

$$x_{k+1} = x_k - \alpha \nabla L(\omega, b)$$

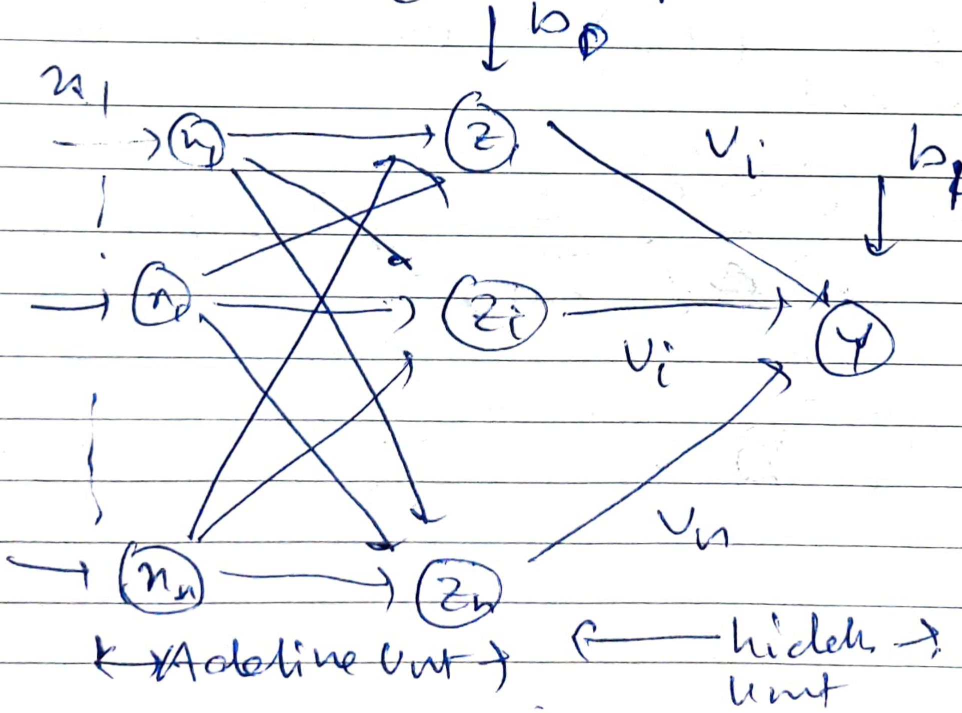
$$\omega(k+1) = \omega_{(k)} + 2\alpha\, e(k) \cdot \sigma'(z) \cdot x(k)$$

$$= \omega_{(k)} + \eta\, e(k) \cdot \sigma'(z) \cdot x(k)$$

$$b(k+1) = b_{(k)} + 2\alpha\, e(k) \cdot \sigma'(z)$$

$$= b(k) + \eta\, e(k) \cdot \sigma'(z)$$

(6) Madaline [Multiple Adeline]



Learning: if $t \neq y$; $t = +1$

$$b_j(\text{new}) = b_j(\text{old}) + \alpha(1 - z_{ij})$$

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha(1 - z_{ij})x_i$$

if, $t \neq y$; $t = -1$

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha(-1 - z_{ij})x_i$$

$$b_j(\text{new}) = b_j(\text{old}) + \alpha(-1 - z_{ij})$$

if $t = y$ no updatn req.

⑦ Back propagation:

Feed Forward:

$$a^{m+1} = f^{m+1}(w^{m+1} a^m + b^{m+1})$$

A/c to LMS,

$$w^m_{ij}(k+1) = w^m_{i,j}(k) - \alpha \frac{\partial \hat{F}}{\partial w^m_{i,j}}$$

$$b_i^m(k+1) = b_i(k) - \alpha \frac{\partial \hat{L}}{\partial b_i^m}$$

$$\frac{\partial \hat{L}}{\partial w_{ij}^m} = \frac{\partial \hat{L}}{\partial z_i^m} \cdot \frac{\partial z_i^m}{\partial w_{ij}^m}$$

$$\frac{\partial \hat{L}}{\partial b_i^m} = \frac{\partial \hat{L}}{\partial z_i^m} \cdot \frac{\partial z_i^m}{\partial b_{i}^m}$$

where, $z_i^m = \sum_{j=1}^{k} w_{i,j}^m a_j^{m-1} + b_i^m$

k = number of neuron in layer m

$$\frac{\partial z_i^m}{\partial w_{ij}} = a_j^{m-1} \quad ; \quad \frac{\partial z_i^m}{\partial b_i^m} = 1$$

$$\delta_i^m = \frac{\partial \hat{L}}{\partial z_i^m}$$

$$\therefore \boxed{\begin{aligned} w_{ij}^m(k+1) &= w_{ij}(k) - \alpha \delta_i^m a_j^{m-1} \\ b_i^m(k+1) &= b_i(k) - \alpha \delta_i^m \end{aligned}}$$

$$s^m = \frac{\partial \hat{L}}{\partial z^m} = \begin{bmatrix} \frac{\partial \hat{L}}{\partial z_1^m} \\ \vdots \\ \frac{\partial \hat{L}}{\partial z_k^m} \end{bmatrix}$$

$$s^m = \frac{\partial \hat{L}}{\partial z^m} = \frac{\partial \hat{L}}{\partial z^{m+1}} \frac{\partial z^{m+1}}{\partial z^m} \rightarrow \underline{\text{back propagate}}$$

$$\frac{\partial z^{m+1}}{\partial z^m} = \begin{bmatrix} \frac{\partial z_1^{m+1}}{\partial z_1^m} & \cdots & \frac{\partial z_1^{m+1}}{\partial z_k^m} \\ \vdots & \cdots & \vdots \\ \frac{\partial z^{m+1}}{\partial z_k^m} & \cdots & \frac{\partial z^{m+1}}{\partial z_k} \end{bmatrix}$$

$$\frac{\partial z_i^{m+1}}{\partial z_j^m} = \frac{\partial}{\partial z_j^m} \left[\sum_{i=1}^{k} w_{i\,l}^{m+1} a_l^m + b_l^{m+1} \right]$$

$$= w_{i,j}^{m+1} \frac{a_j^m}{\partial z_j^m}$$ [for $l = j$; other values are 0]

$$= w_{ij}^{m+1} \frac{\partial f^{m}(z_j^{m})}{\partial z_j^{m}} = w_{ij}^{m+1} \dot{f}^{m}(z_j^{m})$$

$$\frac{\partial z^{m+1}}{\partial z^{m}} = W^{m+1} \dot{F}^{m}(z^{m})$$

$$\dot{F}^{m}(z^{m}) = \begin{bmatrix} \dot{f}^{m}(z_1^{m}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \dot{f}^{m}(z_k^{m}) \end{bmatrix}$$

$$s^{m} = \left(\frac{\partial z^{m+1}}{\partial z_m}\right)^{T} \frac{\partial \hat{L}}{\partial z^{m+1}} = W^{m+1} \dot{F}^{m}(z^{m}) \frac{\partial \hat{L}}{\partial z^{m+1}}$$

10 Sunday $\boxed{s^{m} = W^{m+1} \dot{F}^{m}(z^{m})\, s^{m+1}}$

$$s^{M} \to s^{M-1} \to \cdots \to s^{1}.$$

$$s^{M} = \frac{\partial \hat{L}}{\partial z_i^{M}} = \frac{\partial}{\partial z_i^{M}} \sum_{j=1}^{k} (t_j - a_j^{M})^2 = -2(t_i - a_i^{M}) \frac{\partial a_i^{M}}{\partial z_i^{M}}$$

[for $i = j$]

$$\frac{\partial a_i^M}{\partial z_i^M} = \dot{f}^M(z_i^M)$$

$$\therefore s^M = -2(t_i - a_i^M)\,\dot{f}^M(z_i^M)$$

$$\boxed{s^M = -2\dot{F}^M(z^M)\,(t - a^M)}$$