UNSUPERVISED LEARNING OF CLUSTERS :

This learning is based upon clustering of I/P data.

NO a priori knowledge is available about an i/p's membership in a particular class.

Rather,

gradually detected characteristics & training history will be used to assist the N/w in defining

- classes &
- Possible boundaries b/w them.

Such an unsupervised classification is called "CLUSTERING"

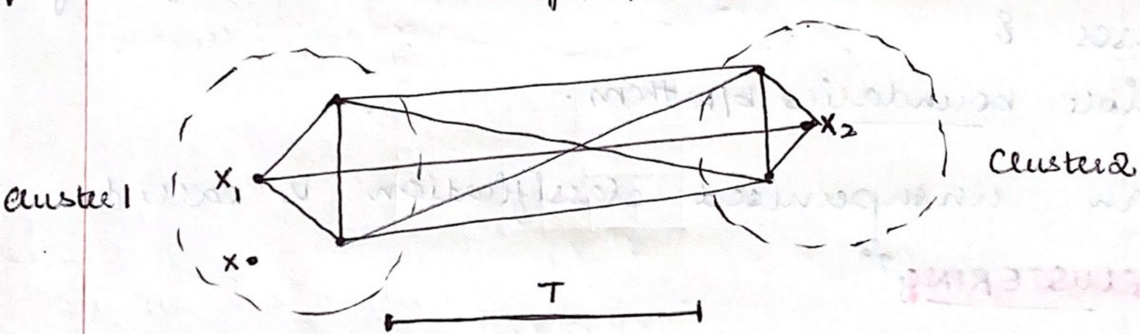
The objective of clustering neural N/w, is to categorize or cluster data.

Clustering should be followed by labelling data/clusters with appropriate category names.
This process is termed as "CALIBRATION"

Clustering is understood to be →
"grouping of similar objects & separating of dissimilar ones."

The clustering technique presented below assumes that the no. of classes is known 'a priori'

The pattern set $\{X_1, X_2 --- X_N\}$ is submitted to the input. to determine decision function required to identify possible clusters.



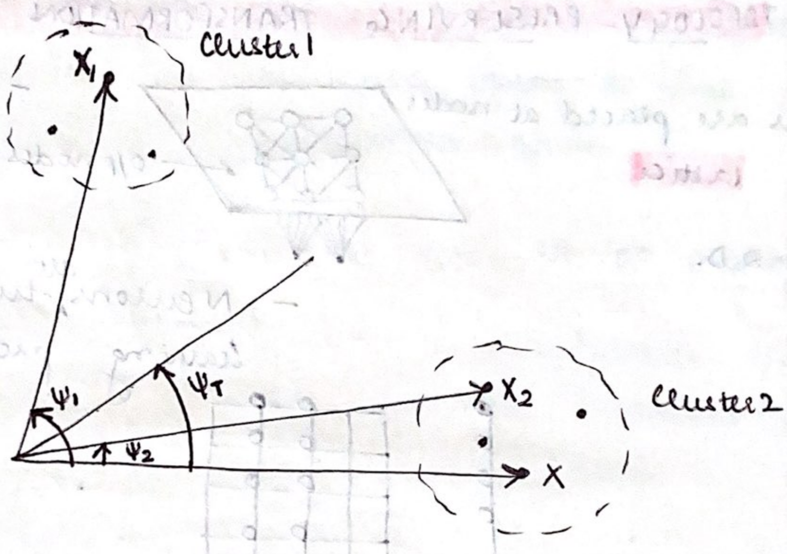
The Euclidean distance is,

$$\|X - X_i\| = \sqrt{(X - X_i)^t (X - X_i)}$$

The rule of similarity is simple:

"Smaller the distance, closer the patterns"

Another rule is COSINE of ANGLE b/w X & X_i :



$$\cos\psi = \frac{X^t X_i}{\|X\| \cdot \|X_i\|}$$

for $\cos\psi_2 < \cos\psi_1$ pattern X is more similar to X_2 than X_1

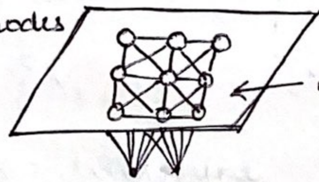
KOHONEN SELF ORGANISED FEATURE MAP ALGO:

Kohonen SOM:

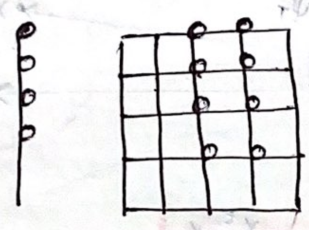
- Based on competitive learning (unsupervised)
- Only 1 O/P neuron activated a a time.
- This neuron with maximum response is the Winner.

TOPOLOGY-PRESERVING TRANSFORMATION

- neurons are placed at nodes of lattice



- location of neurons is tuned to be ordered.
- 1D or 2D.
- Neurons are tuned by competitive learning process.



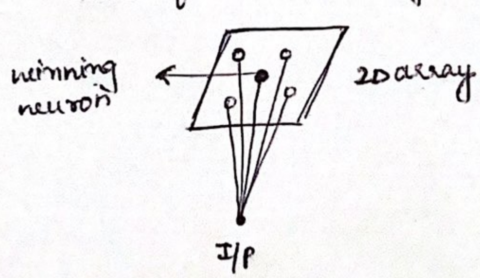
SOM as NN Model:

COMPUTATIONAL MAP: Basic building blocks in info-processing infrastructure of nervous system.

HUMAN BRAIN: Organised in such a way that diff. sensory i/p are represented by ordered computational maps.

BASIC MAPPING MODEL:

- It captures essential features of human brain.
- Remains tracable
- capable of dimensionality reduction.
- class of vector coding.



KOHONEN MODEL

FORMATION OF SOM

- competition (largest value of func^n is se
- winner of competition
- cooperation
- synaptic adaptation.
- adjust weights.

SVM

Classification, Features & Decision regions:

PATTERN: It is a quantitative description of an obj, event etc.
The classification may involve:

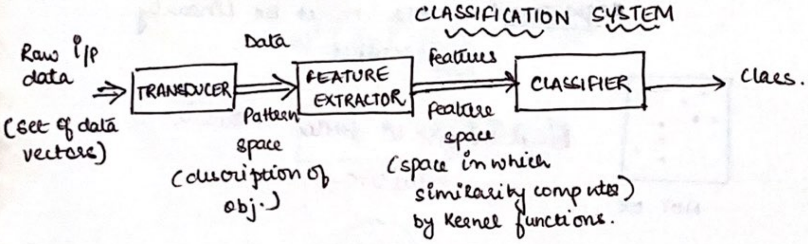
↳ SPATIAL PATTERNS
eg. picture
videos, etc.

↳ TEMPORAL PATTERNS // ordered seq. of data appearing in time.
eg. speech signals v/s time
electrocardiogram, etc.

GOAL of pattern classification:
Assign an obj., event etc to one of predefined classes. (category)

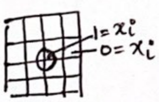
Problem with classificaⁿ: To discriminate i/p data within obj population via search for invariant (constant) attributes among members of population.

CLASSIFICATION SYSTEM



Let, I/p component be denoted by X vector
ie $X = [x_1 \; x_2 \; -- \; x_n]^t$

classification is obtained by decision function $i_0(X)$



where $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ — 1 if cell contains portion of obj; 0 if not

DISCRIMINANT FUNCTIONS:

Consider Patterns (input vectors)
$x_1 \; x_2 \; --- \; x_p$

where $p \to$ size of pattern set.

No. of categories = R

During classification, the classifier takes decision based on DISCRIMINANT FUNCTION ie $g(x)$

there can be R discriminant functions. ie.

$g_1(x) \; .... \; g_R(x)$ we need to compare them

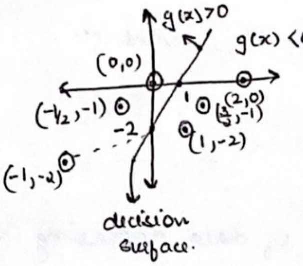
if $g_i(x) > g_j(x)$ for $i, j = 1, 2 --- R$

thus, within region $\mathcal{X}_i$, the i^{th} function will have max value.

decision surface equation is :

$$g_i(x) - g_j(x) = 0.$$

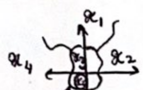
Eg.



class 1 : $[0\ 0]^t$, $[-.5\ -1]^t$, $[-1\ -2]^t$
class 2 : $[2\ 0]^t$, $[1.5\ -1]^t$, $[1\ -2]^t$

$g(x) = -2x_1 + x_2 + 2$.

$R = \underline{\underline{2}}$.

→ 

Region denoted by $\mathcal{R}_i$ → decision region

They are seperated by → decision surface

LINEAR CLASSIFIER :

linear models for classification seperate i/p vectors into classes using linear (HYPERPLANE) decision boundaries.

↓

// decision surface.

Requirement : data must be linearly seperable.

OK　　　NOT OK

Bigger margin is better & this is intuitive.

LINEAR DISCRIMINANT FUNCTION :

$R = 2$

A linear combination of components of x → actual weight or pattern weight?

(ie) $g(x) = w^T x + w^0$　　$w = x_1 - x_2$

any priority to a vector

where , w → weight vector　　x → i/p vector.

w_0 → threshold weight · limit value

2 regions $\mathcal{R}_1$ & $\mathcal{R}_2$ if $g(x) = w^T x + w_0 > 0$　$x \in \mathcal{R}_1$

else $g(x) = w^T x + w_0 < 0$　$x \in \mathcal{R}_2$

if $g(x) = 0$ // on decision surface & can belong to any region/class.

↓

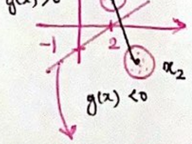
ie → if x_1 & x_2 are 2 pts on decision hyperplane

$w^T x_1 = w^T x^2 = 0$

$w^T(x_1 - x_2) = 0$

ref: pg 108.

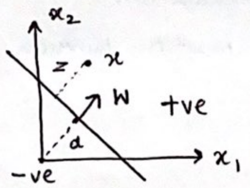
$\underline{w_0} = \frac{1}{2}\left(\|x_2\|^2 - \|x_1\|^2\right)$



$\underline{g(x)} = w^t x + \frac{1}{2}\left(\|x_2\|^2 - \|x_1\|^2\right).$

$\sqrt{x_2^2 + x_1^2}$

HYPERPLANE GEOMETRY :



where → $g(x)$ is measure of distance from hyperplane to x

① Training data

Its SIGN marks on which side of hyperplane x is.

③ d is dist. from origin to hyperplane.

we know, $g(x) = w^T x + w_0$

& on decision surface $g(x) = 0$

$$d = \frac{w_0}{\|w\|}$$

ie. NORMALIZED VALUE
ie. it is vector in same direcⁿ as point but with length = 1.

x can be expressed as its projection on H (x_p) + dist. to H times unit vector in that direction

$$z \times \frac{W}{\|W\|}$$

ie.

$$x = x_p + z\frac{W}{\|W\|}$$

$$\therefore \quad g(x) = w^T(x) + w_0$$

$$= w^T\left(x_p + z\frac{W}{\|W\|}\right) + w_0$$

$$= \underbrace{w^T x_p + w_0}_{g_p(x)=0 \text{ as on hyperplane.}} + z\frac{w^T W}{\|W\|}$$

$$= z\frac{w^T w}{\|w\|}$$

$$= z\|w\|$$

$$\therefore \quad z = \frac{g(x)}{\|w\|}$$

PERCEPTRON ALGORITHM :

It was invented by Frank Rosenblatt (1962)

It is an iterative algo.

→ Strategy :
It is to start with a random guess at the weights (w), and then iteratively change the weights to move the hyperplane in a direction that lowers the classification errors.

→ Problem :
How to compute unknown parameters $w_1, w_2 \cdots w_0$?

→ Assumption :
Two classes $\mathcal{X}_1$ & $\mathcal{X}_2$ are linearly seperable, ie. there exists hyperplane $\hat{w}$, such that

$$\hat{w}^T x > 0 \qquad x \in \mathcal{X}_1$$

$$\hat{w}^T x < 0 \qquad x \in \mathcal{X}_2$$

UNIT:2

ARTIFICIAL NEURAL N/W

Artificial Neural N/W: It is a Cellular N/W that is able to acquire, store and utilize experiential knowledge.

Neural N/W: It can be defined as interconnection of neurons

FEEDFORWARD N/W

Let us consider:

m neurons

n inputs

It output & input vectors are:

$$O = [o_1\ o_2\ ---\ o_m]^t$$

$$X = [x_1\ x_2\ ---\ x_n]^t$$

The Activation value is given as

$$net_i = \sum_{j=1}^{n} w_{ij}\, x_j$$

where $w_{ij} \rightarrow$ weight.

$$O_i = f(w^t x)$$

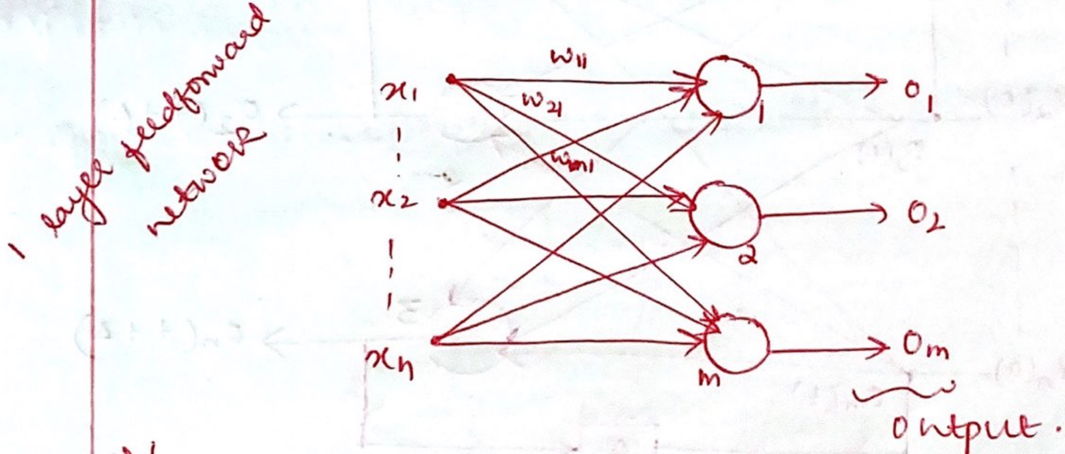
The weight vector is defined as follows:

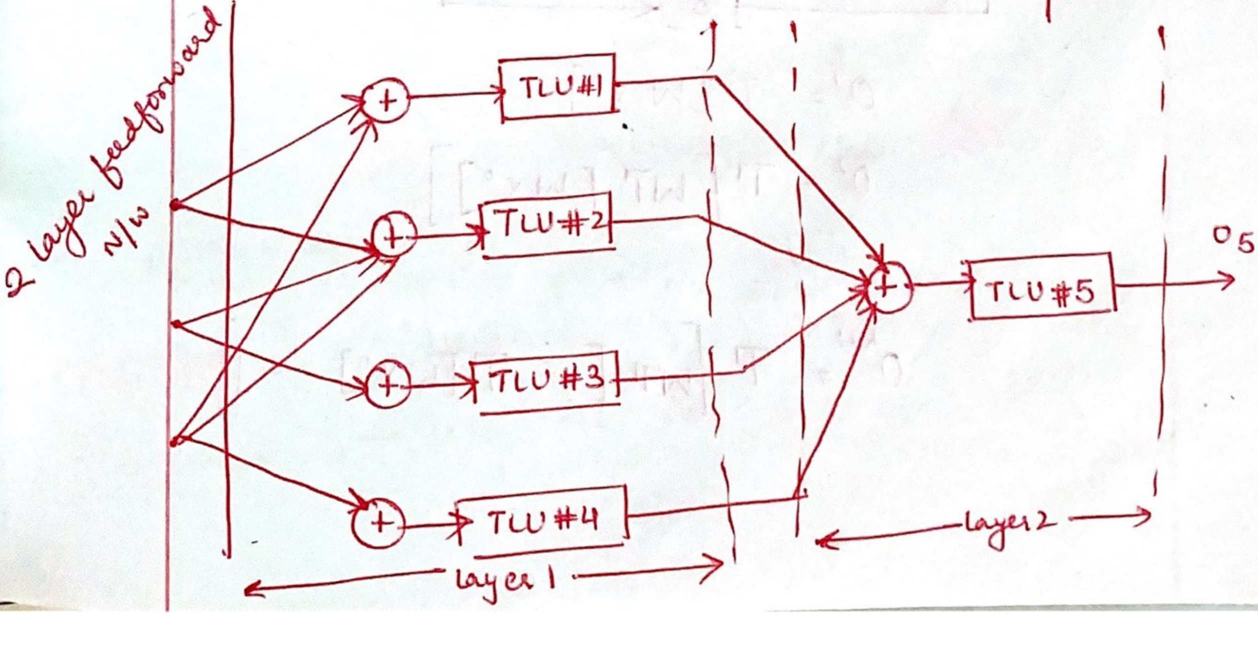
$$w_i \triangleq [w_{i1}\ w_{i2}\ ---\ w_{in}]^t$$

Introducing nonlinear matrix operator (Γ)

$$O = \Gamma[Wx]$$

where W (weight matrix) $\triangleq \begin{bmatrix} w_{11} & - & - & - \\ - & - & - & \\ - & - & - & \\ & & & w_{mn} \end{bmatrix}$

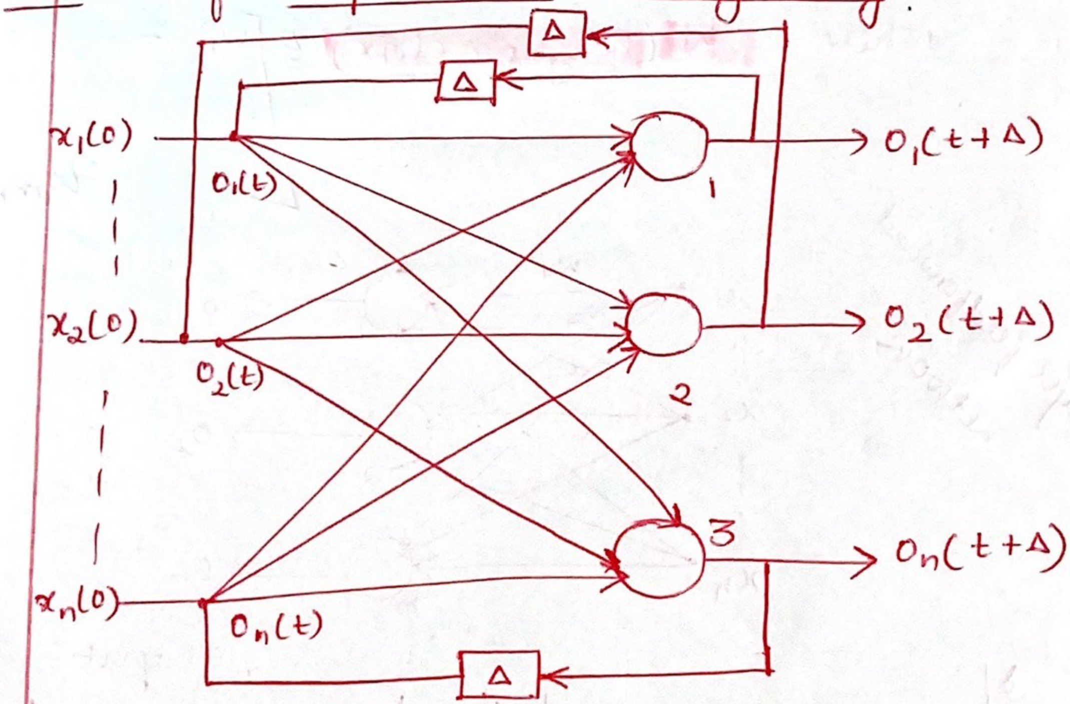




FEEDBACK N/w

In this N/w neurons output are connected to their input.

The essence of closing loop is to enable control of output O_i through O_j.



$$O^1 = T[W x^0]$$

$$O^2 = T[W T[W x^0]]$$

$$\vdots$$

$$O^{K+1} = T[W T[\cdots T[W x^0] \cdots]]$$

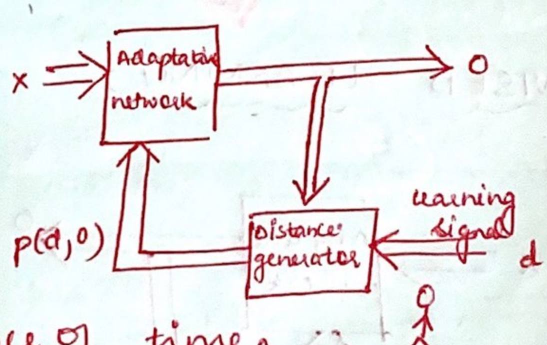
SUPERVISED & UNSUPERVISED LEARNING

BATCH LEARNING: • It takes place when N/w weights are adjusted in single training step.
It is also called **Recording**.

• Learning with feedback either from **teacher** or **environment** is typical for neural N/w.

• Such learning is called **INCREMENTAL** & usually performed in steps.

SUPERVISED LEARNING: (CONVENTIONAL LEARNING)



• At each instance of time, whenever i/p is applied, the 'd' of system is provided by teacher.

• $\rho[d, o]$ is difference b/w actual & desired response, serves as an error measure.

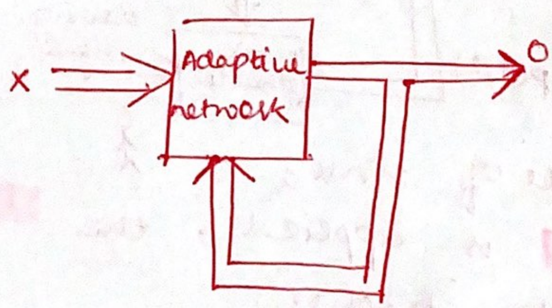
- In these,

responses are known, thus, the error can be used to modify weights so that error decreases.

- A set of i/p & o/p patterns called a TRAINING SET is required for this learning mode.

- Typically, this learning → REWARDS accurate classifica'
PUNISHES inaccurate responses.

UNSUPERVISED LEARNING:



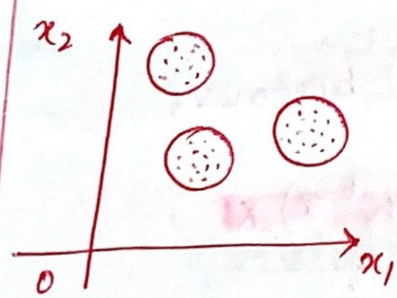
It is sometimes called 'learning without teacher'

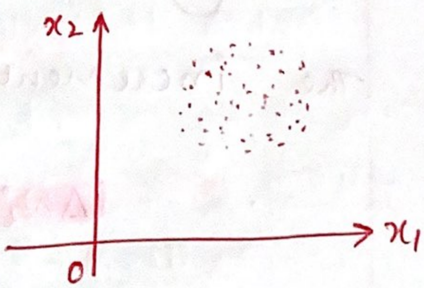
• In this the desired response is not known thus,

explicit error info cannot be used to improve N/W behavior.

• As no info is available as to correctness or incorrectness of response ∴ learning is based on observations.

2D patterns

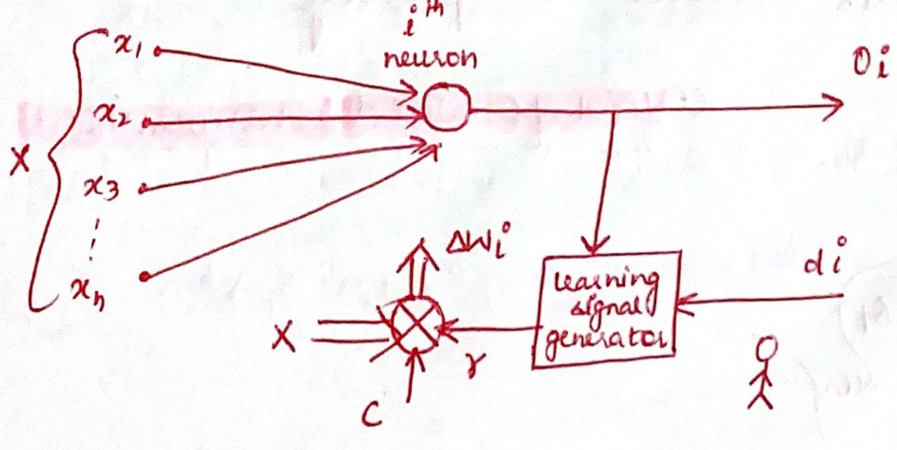




clustered no clusters

*

weight learning rules for supervised learning :



PERCEPTRON LEARNING RULE:

In this, the learning signal is the difference b/w the desired and actual neuron's response.

learning is supervised.

$$r \triangleq d_i - o_i$$

$r \rightarrow$ learning signal

$d_i \rightarrow$ desired o/p

$o_i \rightarrow$ o/p after performance

$$o_i = \operatorname{sgn}(w^t x)$$

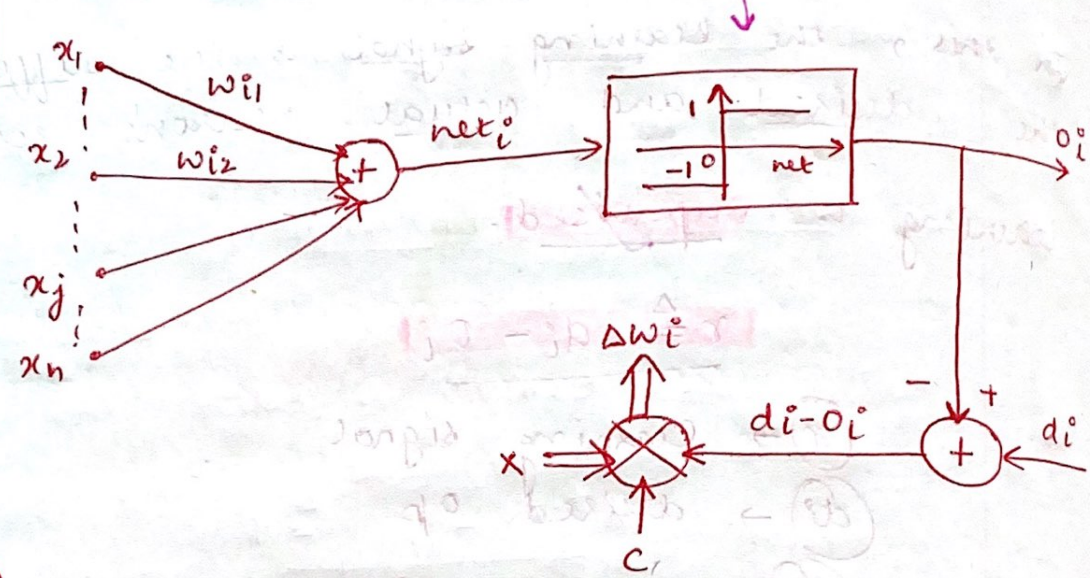
$$\Delta w_i = c[d_i - \operatorname{sgn}(w_i^t x)]\, x$$

This rule is applicable only for BINARY NEURONS.

Under this rule, weights are adjusted if and only if o_i is incorrect.

$+$ when $d_i = 1$ & $\operatorname{sgn}(w^t x) = -1$

$-$ when $d_i = -1$ & $\operatorname{sgn}(w^t x) = 1$



Pg 65 see.

DELTA LEARNING RULE :

It is oney valid for continious activation functions. It is supervised training mode.

$$r \stackrel{\Delta}{=} [d_i - f(w_i^t x)] f'(w_i^t x)$$

The learning signal is called delta.

The error is defined as,

$$E \stackrel{\Delta}{=} \frac{1}{2}(d_i - o_i)^2$$

ie $E = \frac{1}{2}[d_i - f(w_i^t x)]^2$

The error gradient vector value is

$$\nabla E = -(d_i - o_i) f'(w_i^t x) x$$

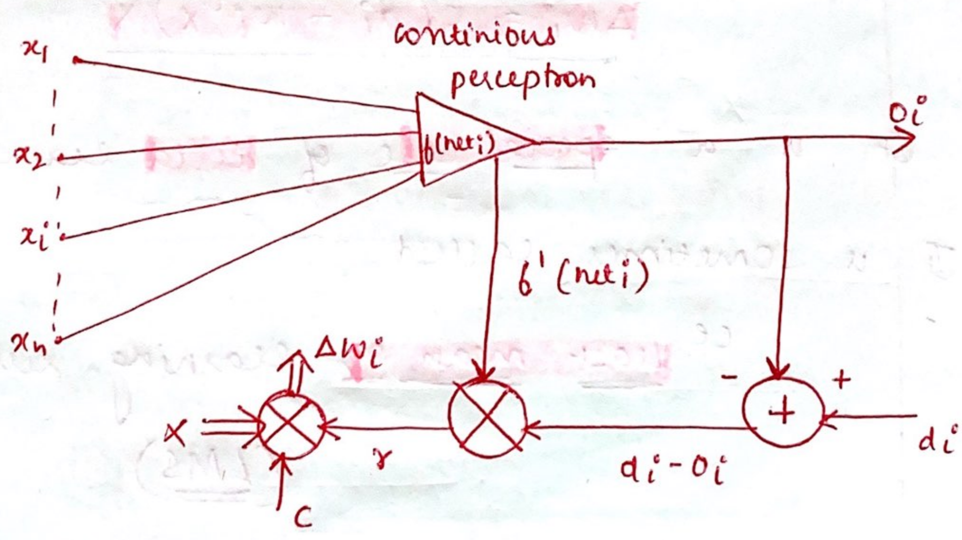
η is positive constant

and $\underset{\eta}{c}\,r\,X$

$$\Delta w_i = \eta (d_i - o_i) f'(net_i) X$$

$f'(net) = \frac{1}{2}[1 - o^2]$

It is also called CONTINIOUS PERCEPTRON TRAINING RULE.



Pg 68 see.