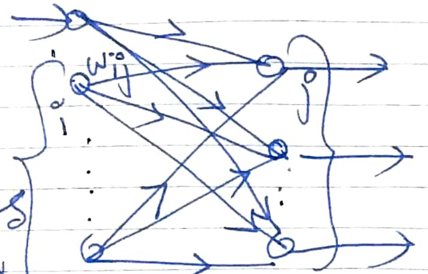
Appointments

Neural N/w.

Perceptron

bias 1



$e=(t-O)$

C_i

o/p layer.

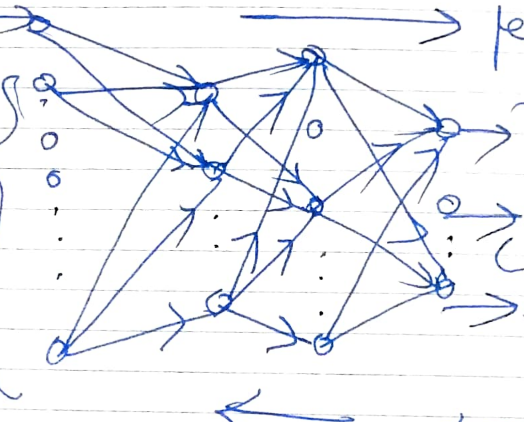
I/p features vector

dimension d.

MLP (Multi layer Perceptron)

bias 1

feed forward n/w for classification



Input vectors.

d-dimension

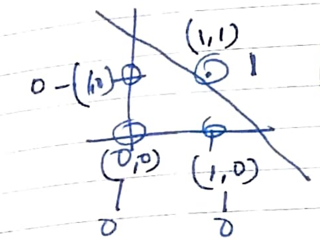
learning — backpropagation N/w

Appointments

linearly Separable - perceptron

non-linearly Separable — MLP.

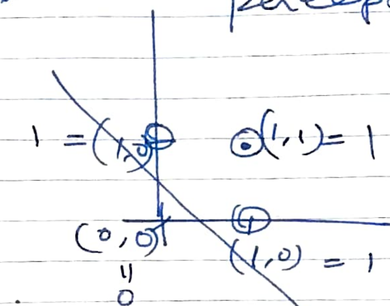
perceptron.



(AND)

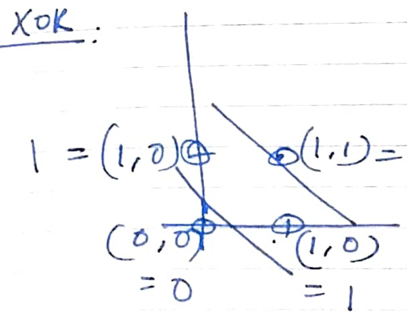
linearly separable.

(OR)



linearly separable

XOR.

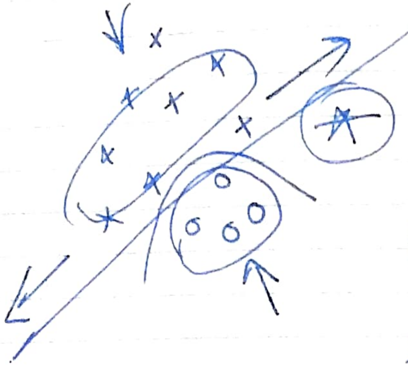


MLP.

non-linearly separable.

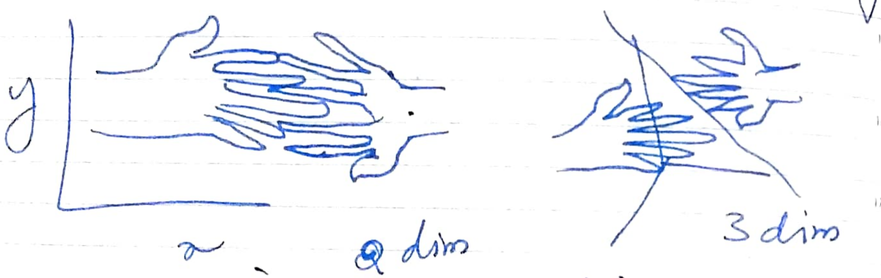
RBF

Non linearly separable problem



* non linearly separable problem transformable to linearly separable problems.

* By increasing the dimensionality



non linearly separable — linearly separable.

RBF

* imposes nonlinearly ~~separable~~ transformation ~~of the~~ of the feature vectors
* ↑ the dimensionality

$P \rightarrow M$

$$\phi(X) = [\phi_1(x), \phi_2(x), \phi_3(x) \ldots \phi_M(x)]^t$$

→ real values.

feature vector function. → ~~of~~ of p dimension.

$M > p$. — increases the dimensionality.

What is radial basis function?

RBF has receptor

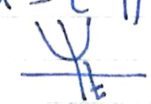
$r = \|t - x\|$

$\|x - t\|$

Various RBF

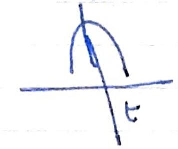
Multiquadrics.

$\phi(r) = (r^2 + c^2)^{1/2} \quad c > 0$



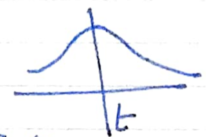
Inverse Multiquadrics.

$\phi(r) = \dfrac{1}{(r^2 + c^2)^{1/2}} \quad c > 0$



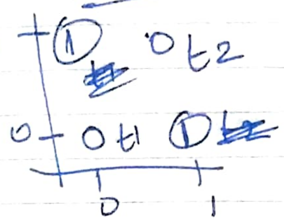
* Gaussian function.

$\phi(r) = \exp\left[-\dfrac{r^2}{2\sigma^2}\right] \quad \sigma > 0$



XOR

RBF



$\phi(x) = e^{-\|x-t\|^2}$

RBF non-linearly separable transformed into linearly separable by ↑ ~~the~~ dimensionality

		$\phi_1(x)$	$\phi_2(x)$
0	0	1	.1
0	1	.4	.4
1	0	.4	.4
1	1	.1	1

$\phi(x) = e^{-\|x-t\|^2}$

$\phi_1(x) = e^{-\|x-t_1\|^2}$

$\phi_2(x) = e^{-\|x-t_2\|^2}$

$\phi_1(0,0) = e^{-\|(0,0)-(0,0)\|^2} = 1$

$\boxed{\text{assume } 2\sigma^2 = 1}$

$\phi_1(0,1) = .4$

$\phi(x) = e^{\left(-\frac{r^2}{2\sigma^2}\right)}$ for $\sigma > 0$.

07 Sunday

$\phi_1(1,0) = .4$

$e^{-\left(\sqrt{(0-0)^2+(0-1)^2}\right)^2}$

$\phi_1(1,1) = e^{-2} = .13 = .1$

$e^{-1} = .367 = .4$

$\phi_2(0,0) = e^{-\|(1,1)-(0,0)\|^2} = e^{-2} = 0.1$

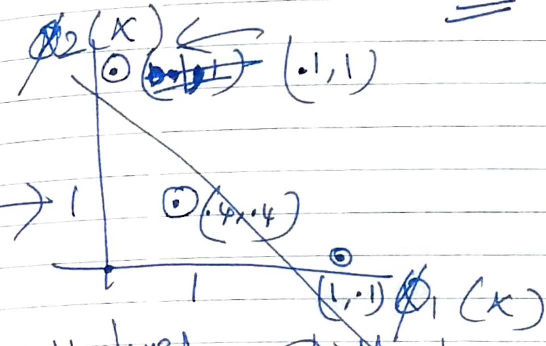
$\phi_2(0,1) = e^{-1} = .4$ ← $= e^{-\|(1,1)-(0,1)\|^2}$

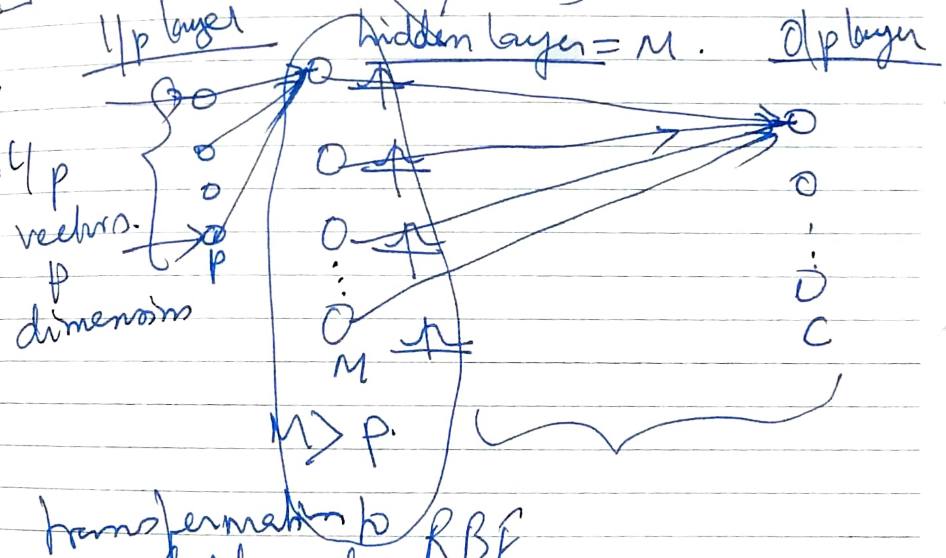
$\left(\sqrt{(1-0)^2+(1-1)^2}\right)^2$

$= \left(\sqrt{1^2+0^2}\right)^2 = \sqrt{1}^2 = 1$

$\phi_2(1,0) = e^{-1} = .4$

$\phi_2(1,1) = e^{-0} = 1$





transformation to hidden space RBF

FEBRUARY MARCH APRIL

2 level of training

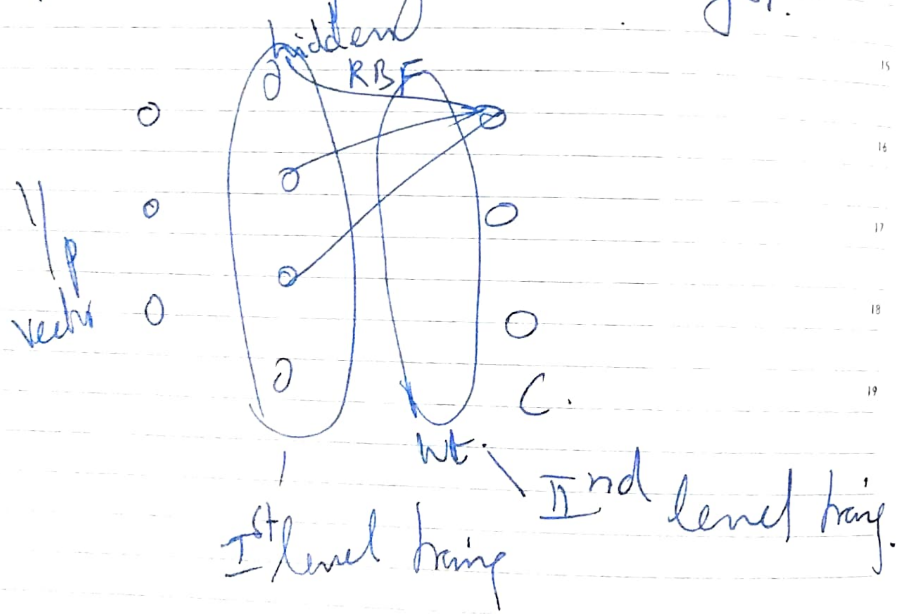
Ist level training - traing of hidden node - RBF.

$$\phi(x) = e^{-\left(\frac{r^2}{2\sigma^2}\right)} \rightarrow o\ t.$$

$$r = \| t - x \|$$

σ - Spade of RBF

t, σ - traing of hidden layer.



MLP.

I/p → hidden → O/p.

- ↓ err,

update weight

I/p. ← O/P ←

RBF

I/p - hidden (M).

Ist level → traing of hidden layer node.

RBF, t, σ.

N no. of training vectors.

$x_1, x_2 \ldots x_N$.

↓

each vector of dim p.

↓

M.

t_i = i = 1,..M. hidden node.

σ_i = i = 1,..M.

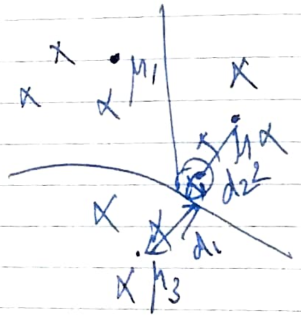
(1).

~~N of~~

cluster N no. of training samples into M no. of clusters.

~~Af~~ N — M.

N >> M.



$d_1 > d_2$, x_i $\xrightarrow{}$ $X_2 - X_1$

After so many ~~cluster~~ iterations.

cluster centres converges.

↓

Cluster centers becomes the receptors (t) of RBF node in hidden layer

$\frac{t_i}{\sigma_i}$ p no. of nearest receptors.

nei

$$\sigma_j = \sqrt{\sum_{i=1}^{p} (t_j - t_i)^2 \cdot \frac{1}{p}}$$

RBF — hidden

O t_i σ_i — $\phi_1(X)$

O t_j σ_j — $\phi_2(X)$

O — $\phi_M(X)$

w_{ij} — for classification.

M ↑. — becomes linearly separability.

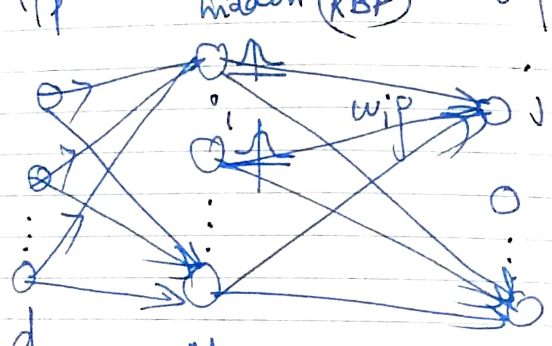
Least mean square error tech.

~~X~~ . $\underline{N}$.

$X \begin{bmatrix} - & - & - & - \\ - & - & & \end{bmatrix}$

$N \times \underline{d}$ — d - dim of feature vector.

I/p   hidden (RBF)   O/p layer.



$M > d$.   c.

↓

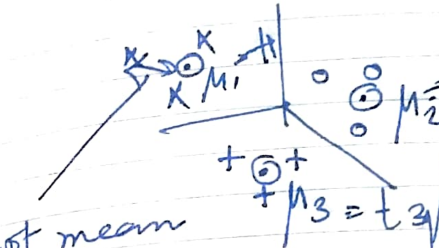
linear combination

$$\phi_i(x) = e^{-\left(\frac{\|x - t_i\|^2}{2\sigma^2}\right)}$$

$x \rightarrow c_i$

N . samples.

M. RBF - clusters.

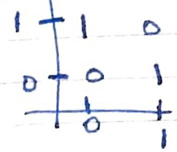


root mean square.

$$\cancel{\sigma_i = \frac{1}{p}\sum_{i=1}^{p} (t_i - t_k)^2}$$

$$\boxed{\sigma_i = \sqrt{\frac{1}{P}\sum_{k=1}^{P} (t_i - t_k)^2} \qquad \phi_i(x) = e^{-\frac{\|x - t_i\|^2}{2\sigma^2}}}$$

XOR



$P = 2$

$\phi_1 \rightarrow t_1 = (0,0) - \sigma_1 = 1 \quad \phi_1(x) = e^{-\frac{\|x - t_1\|^2}{2\sigma_1^2}}$

$\phi_2 \rightarrow t_2 = (0,1) - \sigma_2 = 1 = \phi_2(x) = e^{-\frac{\|x - t_2\|^2}{2\sigma_2^2}}$

$\phi_3 \rightarrow t_3 = (1,0) - \sigma_3 = 1$

$t_4 \rightarrow t_4 = (1,1) - \sigma_4 = 1 = \phi_4(x) = e^{-\frac{\|x - t_4\|^2}{2\sigma_4^2}}$

Appointments

I/p feature	ϕ_1	ϕ_2	ϕ_3	ϕ_4
0 0	1	0.6	0.6	0.4
0 1	0.6	1.0	0.4	0.6
1 0	~~0.6~~	~~0.4~~	1	.6
1 1	.4	.6	.6	1
	2		1	

(0,0) (0,1)

$$\phi_2(X) = e^{-\frac{\|x-t_2\|^2}{2\sigma_2}} \quad \sqrt{(0-0)^2+(0-1)^2}$$

$$= \sqrt{1^2} = 1$$

$$= e^{-(1/2)} = e^{-.5} = .6$$

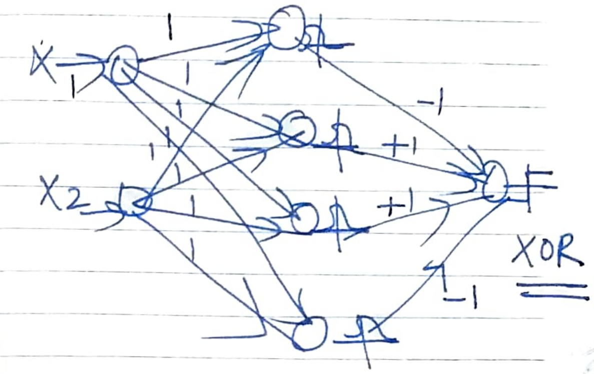
weight.	-1	$+1$	$+1$	-1

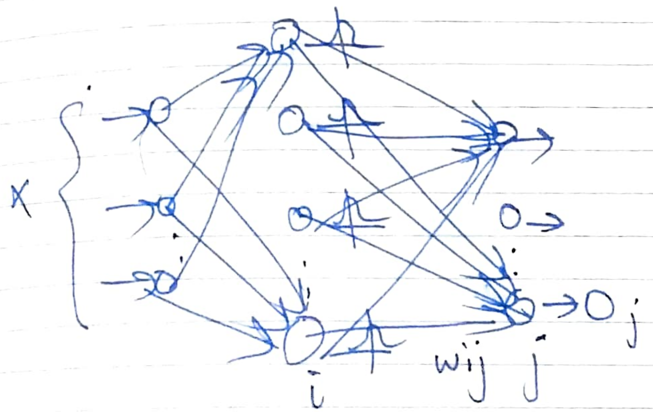
$.6 + .6 - 1 - .4$

$1.2 - 1 - .4 = -.2$

Appointments

$\Sigma w_i \phi_i$	O/p.
$-.2$	0
0.2	1
0.2	1
-0.2	0





$$O_j = \sum_{i=1}^{M} w_{ij} \cdot \phi_i(x)$$

if $x \in \omega_j$

$$\sum_{i=1}^{M} w_{ij}\phi_i(x) > 0 = +1$$

if $x \notin \omega_j$

$$\sum_{i=1}^{M} w_{ij}\phi_i(x) \le 0 = -1$$

Sample X_k $k = 1, 2 \ldots N$ samples.

$\phi_i(X_k) = \phi_{ik}$.

$\sum_{i=1}^{M} w_{ij} \phi_{ik} = +1$ if $X_k \in \omega_k$

$= 0$ if $X_k \notin \omega_k$

$$\begin{bmatrix} \phi_{11} & \phi_{21} & \phi_{31} & \cdots & \phi_{M1} \\ \phi_{12} & \phi_{22} & \phi_{32} & & \phi_{M2} \\ \\ \phi_{1N} & \phi_{2N} & & & \phi_{MN} \end{bmatrix} \begin{bmatrix} w_{1j} \\ w_{2j} \\ \vdots \\ w_{Nj} \end{bmatrix} = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{Nj} \end{bmatrix}$$

$b_{ij} = 1$ if $x_i \in w_j$

$= 0$ if $x_i \notin w_j$

$$\boxed{\phi W_j = b_j}$$

$e = \phi W_j - b_j$

$J(w_j) = \| \phi w_j - b_j \|^2$.

$\nabla J(w_j) = 2\phi^t (\phi w_j - b_j)$

$= 0$

$$w_j = \underbrace{(\phi^t \phi)^{-1} \cdot \phi^t}_{= \text{pseudo inverse}} \cdot b_j$$

$w_j = \phi^+ b_j$

Adv	RBF	MLP
①	training faster	slower.
②	interpret → the meaning in the hidden layer	not interpretable
		no. of hidden layers + no. of hidden nodes in each hidden layer
③	classifier slower.	→ faster →