RADIAL BASIS FUNCTION NETWORKS

– Radial Basis Functions:

$x \in \mathbb{R}^n \quad \mu \in \mathbb{R}^n \qquad r = \|x - \mu\|$ (Euclidean distance)

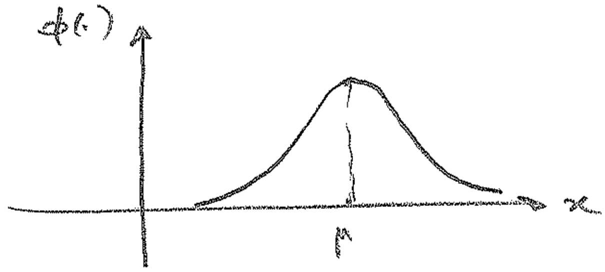
1°) $\phi : \mathbb{R} \to \mathbb{R}$, $\phi(r)$: only depends on RADIAL DISTANCE

2°) $\lim_{r \to \infty} \phi(r) = 0$ (Localization)

(– Typically one assumes $\phi(0) = 1$ (normalization and decreasing functions, although this is not necessary)

– Typical basis function: Gaussian

$$\phi(\|x-\mu\|) = e^{-\frac{\|x-\mu\|^2}{\sigma^2}}$$



σ: determines the "sharpness" of the gaussian curve.

$\sigma \to 0$, $\phi(\cdot) \to \delta(\cdot)$ (impulse)

RADIAL BASIS FUNCTION NETWORKS

- Radial Basis Functions:

$x \in \mathbb{R}^n \quad \mu \in \mathbb{R}^n \qquad r = \|x-\mu\|$ (Euclidean distance)

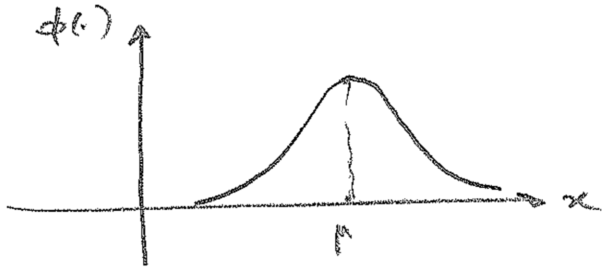
1°) $\phi: \mathbb{R} \to \mathbb{R}$, $\phi(r)$: only depends on RADIAL DISTANCE

2°) $\lim_{r \to \infty} \phi(r) = 0$ (Localization)

(- Typically one assumes $\phi(0) = 1$ (normalization and decreasing functions, although this is not necessary)

- Typical basis function: Gaussian

$$\phi(\|x-\mu\|) = e^{-\frac{\|x-\mu\|^2}{\sigma^2}}$$



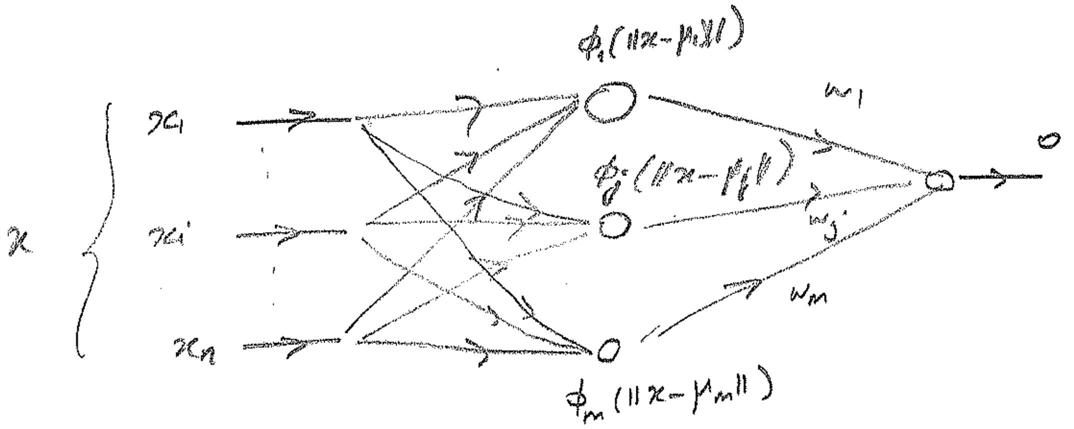
σ: determines the "sharpness" of the gaussian curve.

$\sigma \to 0$, $\phi(\cdot) \to \delta(\cdot)$ (impulse)

— Could be used in function approximation/interpolation.

— Typical configuration.

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}$$



$$\Rightarrow \boxed{o = \sum_{j=1}^{m} w_j \, \phi_j(\|x-\mu_j\|)}$$

in case $\phi_j(\|x-\mu_j\|) = e^{-\frac{\|x-\mu_j\|^2}{\sigma^2}}$

$$\boxed{o = \sum_{j=1}^{m} w_j \, e^{-\frac{\|x-\mu_j\|^2}{\sigma^2}}}$$

if $x \sim \mu_j \Rightarrow e^{-\frac{\|x-\mu_j\|^2}{\sigma^2}}$ is relatively big

if x is far away from $\mu_j \Rightarrow e^{-\frac{\|x-\mu_j\|^2}{\sigma^2}}$ is relatively small

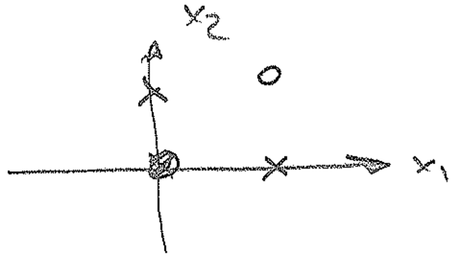
Hence each radial basis function is LOCALIZED AROUND μ_j (center ≡ mean) point.

— This could be used in function approximation
— pattern recognition as well.

EXAMPLE: CONSIDER EXOR PROBLEM:

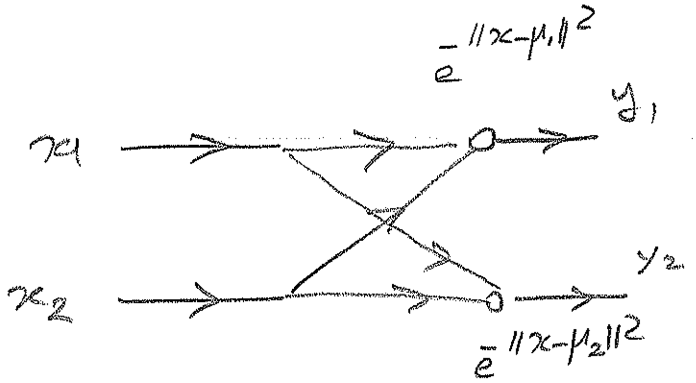
$$O = x_1 \oplus x_2$$

x_1	x_2	O
0	0	1
0	1	0
1	0	0
1	1	1



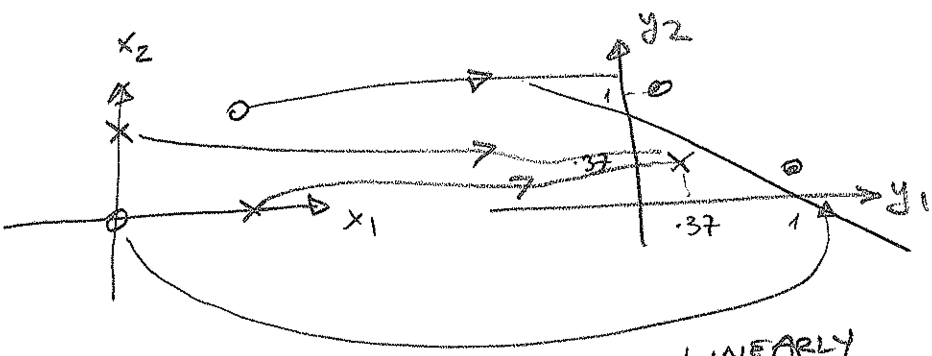
NOT LINEARLY SEPARABLE

Consider the following RBF Network



$\mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\mu_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\Rightarrow$ y_1 & y_2 are LOCALIZED on $\underline{\underline{C_1}}$

x_1	x_2	y_1	y_2
0	0	1	.135
0	1	.37	.37
1	0	.37	.37
1	1	.135	1



LINEARLY SEPARABLE!

INTERPOLATION:

Let the sample points be given as $(x_1, b_1), \ldots, (x_N, b_N)$ i.e. $\longrightarrow \boxed{y(\cdot)} \longrightarrow b$ $\boxed{y(x_i) = b_i}$

Use $\underline{\underline{N}}$ basis functions, centered at sample points $\underset{\sim}{x_i}$

$$F(x) = \sum_{i=1}^{N} w_i \, \phi(\|x - x_i\|)$$

Define $g_{ij} = \phi(\|x_j - x_i\|)$

$$w_1 g_{11} + w_2 g_{12} + \cdots + w_N g_{1N} = b_1$$

$$\vdots$$

$$w_1 g_{N1} + w_2 g_{N2} + \cdots + g_{NN} g_{NN} = b_N$$

$$\underbrace{\begin{bmatrix} g_{11} & \cdots g_{1i} & \cdots g_{1N} \\ g_{i1} \cdots & g_{ii} \cdots & g_{iN} \\ g_{N1} \cdots & g_{Ni} \cdots & g_{NN} \end{bmatrix}}_{\text{KNOWN}} \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_i \\ \vdots \\ w_N \end{bmatrix}}_{\substack{\text{UNKNOWN} \\ \text{WEIGHTS}}} = \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_i \\ \vdots \\ b_N \end{bmatrix}}_{\text{KNOWN}}$$

$$G \qquad\qquad w \qquad\qquad b$$

$$\boxed{Gw = b}$$

FACT: If $x_i \neq x_j \Rightarrow$ G is nonsingular for a large class of RBF (including gaussian)

$$\Rightarrow \quad \boxed{w = G^{-1} b}$$

TRAINING

Training set, as given before

$$\mathcal{T}_r = \{ (x_1, b_1), \dots, (x_N, b_N) \}$$

$$E(n) = \frac{1}{2} \| g(x_n) - b_n \|^2$$

$$g(x) = \sum_{i=1}^{N} w_i \, \phi(\|x - \mu_i\|)$$

$$E_{cum} = \sum_{n=1}^{N} E(n)$$

Update the weights

$$\boxed{w_j \leftarrow w_j - \eta \frac{\partial E(n)}{\partial w_j}}$$

– Computation of gradient $\frac{\partial E(n)}{\partial w_j}$ is not difficult, given the form of $\phi(r)$.