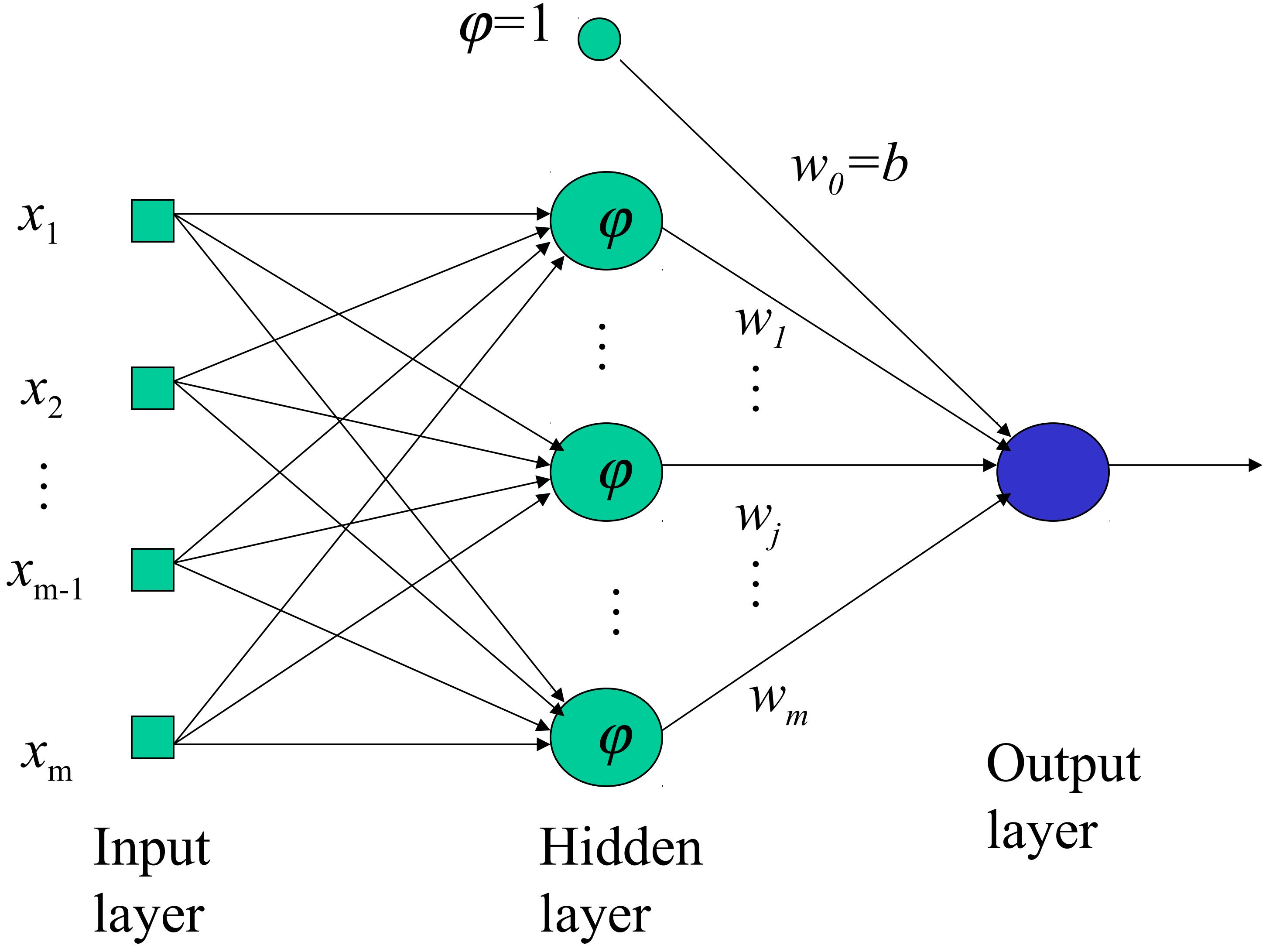
Radial Basis Function



Radial Basis Function

$$F(\mathbf{x}) = \sum_{i=1}^{m} w_i \varphi_i(\mathbf{x})$$

$$\varphi_i(\mathbf{x}) = G(\|\mathbf{x} - \mathbf{t}_i\|)$$

$$\begin{aligned}\boldsymbol{\Phi}(F) &= \boldsymbol{\Phi}_s(F) + \lambda \boldsymbol{\Phi}_c(F) \\ &= \sum_{i=1}^{m} \left(d_i - \sum_{j=1}^{m} w_j G(\|\mathbf{x}_i - \mathbf{t}_i\|) \right)^2 + \lambda \|\mathbf{D}F\|^2\end{aligned}$$

$\boldsymbol{\Phi}_s(F)$: Standard error term

$\boldsymbol{\Phi}_c(F)$: Regularizing term

Radial Basis Function

define

$$\mathbf{d} = \left[d_1, d_2, \ldots, d_N\right]^T$$

$$\mathbf{G} = \begin{bmatrix} G(\mathbf{x}_1, \mathbf{t}_1) & G(\mathbf{x}_1, \mathbf{t}_2) & \ldots & G(\mathbf{x}_1, \mathbf{t}_m) \\ G(\mathbf{x}_2, \mathbf{t}_2) & G(\mathbf{x}_2, \mathbf{t}_2) & \ldots & G(\mathbf{x}_2, \mathbf{t}_m) \\ \ldots & \ldots & \ldots & \ldots \\ G(\mathbf{x}_N, \mathbf{t}_1) & G(\mathbf{x}_N, \mathbf{t}_2) & \ldots & G(\mathbf{x}_N, \mathbf{t}_m) \end{bmatrix}$$

$$\mathbf{w} = \left[w_1, w_2, \ldots, w_m\right]^T$$

Radial Basis Function

$$\mathbf{G}_0 = \begin{bmatrix} G(\mathbf{t}_1, \mathbf{t}_1) & G(\mathbf{t}_1, \mathbf{t}_2) & \ldots & G(\mathbf{t}_1, \mathbf{t}_m) \\ G(\mathbf{t}_2, \mathbf{t}_2) & G(\mathbf{t}_2, \mathbf{t}_2) & \ldots & G(\mathbf{t}_2, \mathbf{t}_m) \\ \ldots & \ldots & \ldots & \ldots \\ G(\mathbf{t}_N, \mathbf{t}_1) & G(\mathbf{t}_N, \mathbf{t}_2) & \ldots & G(\mathbf{t}_N, \mathbf{t}_m) \end{bmatrix}$$

Minimizing the following function

$$\boldsymbol{\Phi}(F) = \boldsymbol{\Phi}_s(F) + \lambda \boldsymbol{\Phi}_c(F)$$

$$= \sum_{i=1}^{m} \left(d_i - \sum_{j=1}^{m} w_j G(\|\mathbf{x}_i - \mathbf{t}_i\|) \right)^2 + \lambda \|\mathbf{D}F\|^2$$

We have

$$\left(\mathbf{G}^T \mathbf{G} + \lambda \mathbf{G}_0\right)\mathbf{w} = \mathbf{G}^T \mathbf{d}$$

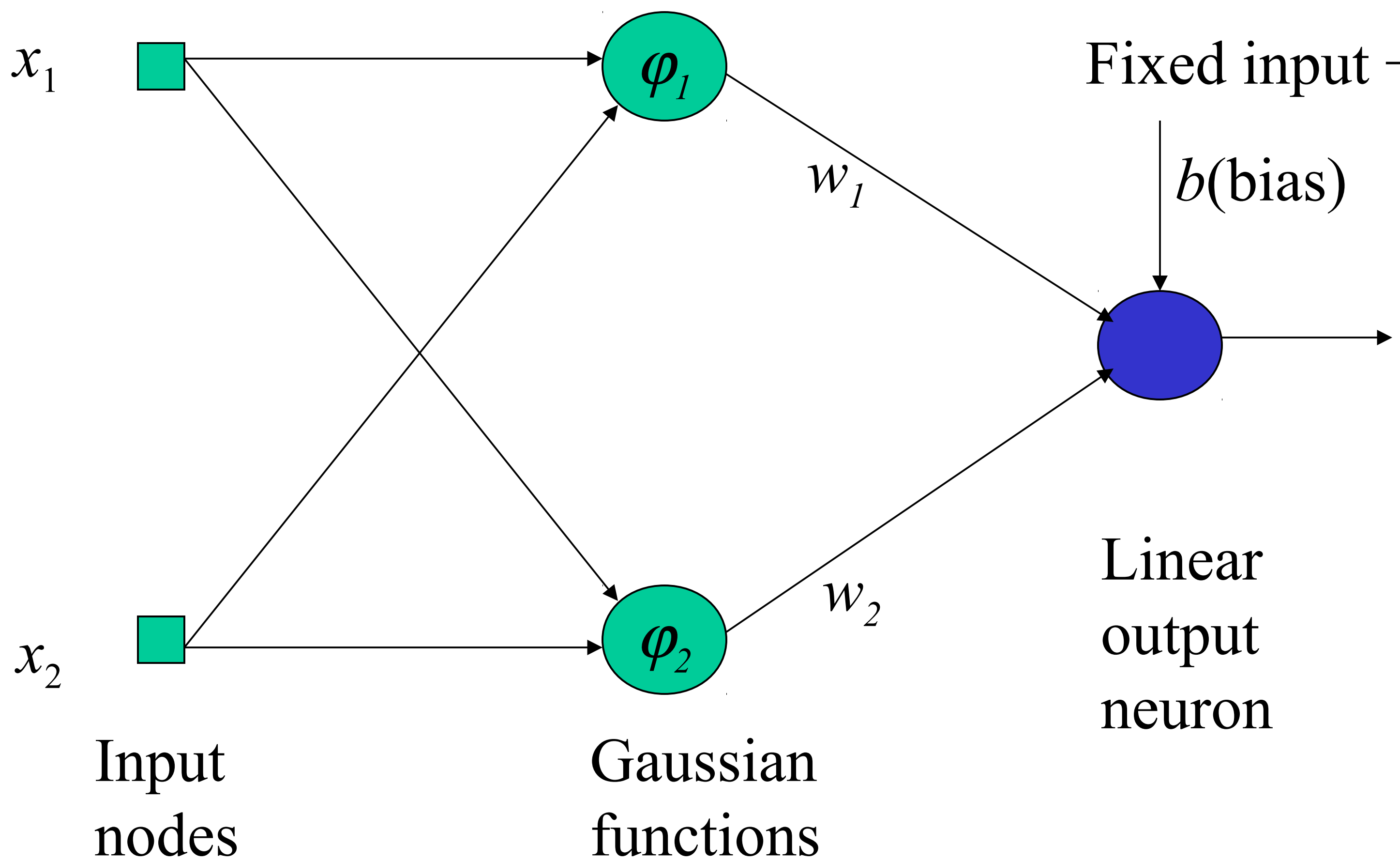
or

$$\mathbf{w} = \mathbf{G}^+ \mathbf{d}, \lambda = 0$$

$$\mathbf{G}^+ = \left(\mathbf{G}^T \mathbf{G}\right)^{-1} \mathbf{G}$$

Radial Basis Function

XOR Problem



Radial Basis Function

XOR Problem

Data point j	input pattern $\mathbf{x}_j$	Desired output d_j
1	(1, 1)	0
2	(0, 1)	1
3	(0, 0)	0
4	(1, 0)	1

Radial Basis Function

XOR Problem

$$t_1 = [1,1]$$
$$t_2 = [0,0]$$
$$y(\mathbf{x}) = \sum_{i=1}^{2} w_i G(\|\mathbf{x} - \mathbf{t}_i\|) + b$$
$$G(\|\mathbf{x} - \mathbf{t}_i\|) = \exp\left(-\|\mathbf{x} - \mathbf{t}_i\|^2\right)$$
$$y(\mathbf{x}_j) = d_j$$

Radial Basis Function

XOR Problem

$$\mathbf{G} = \begin{bmatrix} 1 & 0.1353 & 1 \\ 0.3678 & 0.3678 & 1 \\ 0.1353 & 1 & 1 \\ 0.3678 & 0.3678 & 1 \end{bmatrix}$$

$$\mathbf{d} = [0,1,0,1]$$

$$\mathbf{w} = [w, w, b]$$

Radial Basis Function

XOR Problem

$$\mathbf{w} = \mathbf{G}^{+}\mathbf{d} = \left(\mathbf{G}^{T}\mathbf{G}\right)^{-1}\mathbf{G}^{T}\mathbf{d}$$

$$\mathbf{G}^{+} = \begin{bmatrix} 1.8292 & -1.2509 & 0.6727 & -1.2509 \\ 0.6727 & -1.2509 & 1.8292 & -1.2509 \\ -0.9202 & 1.4202 & -0.9202 & 1.4202 \end{bmatrix}$$

$$\mathbf{w} = \begin{bmatrix} -2.5018 \\ -2.5018 \\ +2.8404 \end{bmatrix}$$

Self-organized Selection of Centers

Algoritmo

Passo 1: Escolher K centros iniciais de grupos que podem ser arbitrariamente escolhidos do conjunto de objetos $\mathbf{s} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ a ser classificados (n é o numero de objetos). Deste modo, determina-se um conjunto de centros cuja quantidade k é fornecida como parâmetro. Seja então o conjunto inicial de centros, vetores

$$\mathbf{z}(0) = \{\mathbf{z}_1(1), \mathbf{z}_2(1), ..., \mathbf{z}_K(1)\}$$

K-médias

Passo 2: distribuir os objetos $\mathbf{x} \in \mathbf{s}$ entre os K grupos $S_j(t)$

assumindo o critério:

$$\mathbf{x} \in S_j(t) \quad \text{se } \|\mathbf{x} - \mathbf{z}_j(t)\| < \|\mathbf{x} - \mathbf{z}_i(t)\|,$$

$$\forall i, j = 1, 2, ..., K \text{ e } j \neq i$$

K-médias

Passo 3: Como os centros foram arbitrariamente determinados, calcula-se os novos centros como as medias de cada grupo. Assim, o conjunto $\mathbf{z}(t)$ assumirá novos valores dados por

$$\mathbf{z}_j(t+1) = \frac{1}{N_j} \sum_{\mathbf{x} \in S_j(t)} \mathbf{x}$$

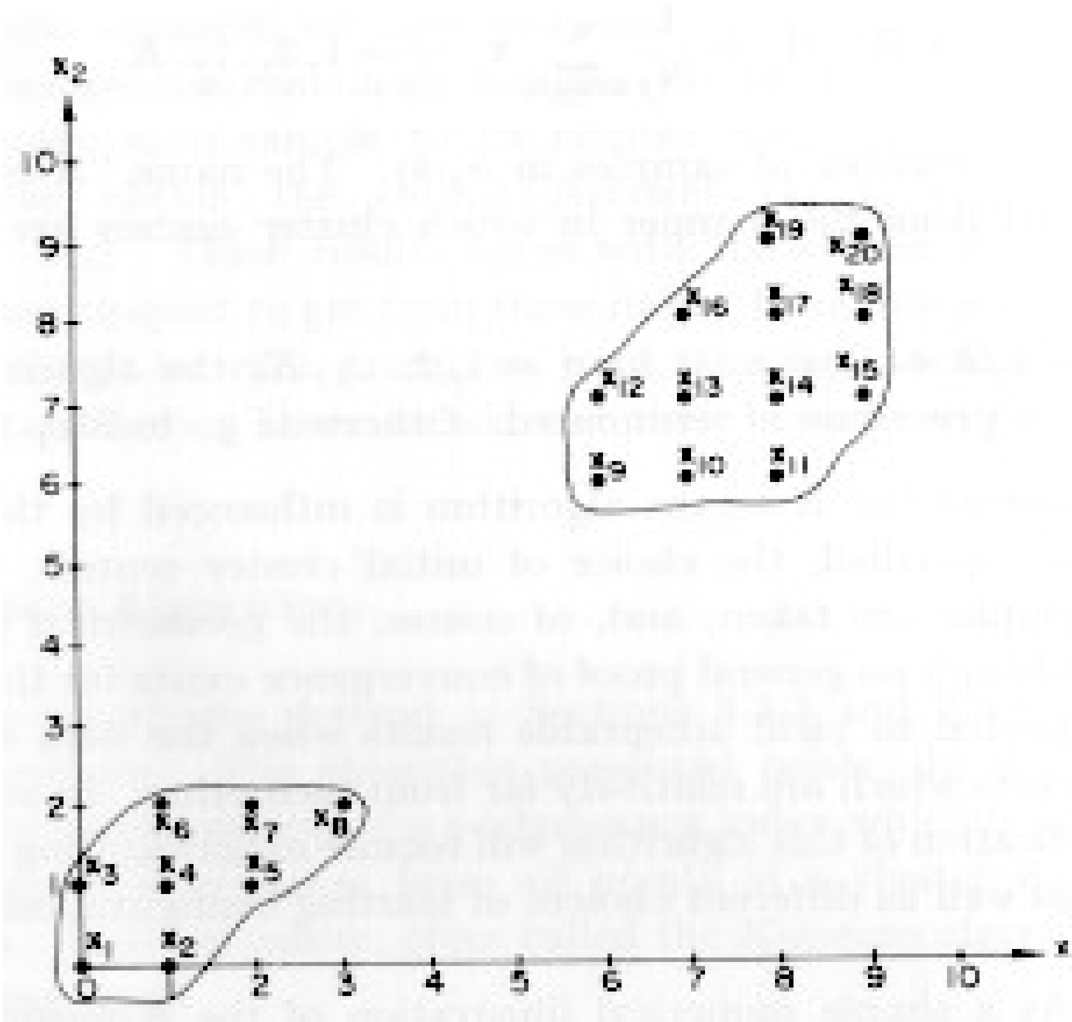
Onde N_j é o numero de amostras em $S_j(t)$

K-médias

Passo 4: Testar a convergência da nova classificação com classificação anterior, isto é, se $\mathbf{z}(t) = \mathbf{z}(t+1)$. Se a igualdade não ocorrer, o passo 2 devera ser repetido até que ocorra.

K-médias

Exemplo



K-médias

Exemplo,

Passo 1: Assume $K = 2$ e escolhe $\mathbf{z}_1(1) = \mathbf{x}_1 = (0, 0)$,

$\mathbf{z}_2(1) = \mathbf{x}_2 = (1, 0)$;

Passo 2:

$\|\mathbf{x}_1 - \mathbf{z}_1(1)\| < \|\mathbf{x}_1 - \mathbf{z}_i(1)\|$ e $\|\mathbf{x}_3 - \mathbf{z}_1(1)\| < \|\mathbf{x}_3 - \mathbf{z}_i(1)\|$,

$i = 2$, então, $S_1(1) = \{\mathbf{x}_1, \mathbf{x}_3\}$. Restos padrões são mais

próximos a $\mathbf{z}_2(1)$, então, $S_2(1) = \{\mathbf{x}_2, \mathbf{x}_4, \mathbf{x}_5, \ldots, \mathbf{x}_{20}\}$

K-médias

Passo 3: Atualiza os centros de cluster

$$\mathbf{z}_1(2) = \frac{1}{N_1} \sum_{\mathbf{x} \in S_1(1)} \mathbf{x} = \frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_3) = \begin{pmatrix} 0.0 \\ 0.5 \end{pmatrix}$$

$$\mathbf{z}_2(2) = \frac{1}{N_2} \sum_{\mathbf{x} \in S_2(1)} \mathbf{x} = \frac{1}{18}(\mathbf{x}_2 + \mathbf{x}_4 + \ldots + \mathbf{x}_{20}) = \begin{pmatrix} 5.67 \\ 5.33 \end{pmatrix}$$

Passo 4: $\mathbf{z}_\mathrm{j}(2) \neq \mathbf{z}_\mathrm{j}(1)$, $j = 1, 2$, então, retorna para o

Passo 2.

K-médias

Passo 2: Com os novos centros de cluster, obtemos que

$||\mathbf{x}_j - \mathbf{z}_1(2)|| < ||\mathbf{x}_j - \mathbf{z}_2(2)||$, para $j = 1, 2, ..., 8$ e

$||\mathbf{x}_j - \mathbf{z}_2(2)|| < ||\mathbf{x}_j - \mathbf{z}_2(2)||$, para $j = 9, 10, ..., 20$,

então, $S_1(2) = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_8\}$ e $S_2(2) = \{\mathbf{x}_9, \mathbf{x}_{10}, ..., \mathbf{x}_{20}\}$.

K-médias

Passo 3: Atualiza os centros de cluster

$$\mathbf{z}_1(3) = \frac{1}{N_1} \sum_{\mathbf{x} \in S_1(2)} \mathbf{x} = \frac{1}{8}(\mathbf{x}_1 + \mathbf{x}_2 + \ldots + \mathbf{x}_8) = \begin{pmatrix} 1.25 \\ 1.13 \end{pmatrix}$$

$$\mathbf{z}_2(3) = \frac{1}{N_2} \sum_{\mathbf{x} \in S_2(2)} \mathbf{x} = \frac{1}{12}(\mathbf{x}_9 + \mathbf{x}_{10} + \ldots + \mathbf{x}_{20}) = \begin{pmatrix} 7.67 \\ 7.33 \end{pmatrix}$$

Passo 4: $\mathbf{z}_j(3) \neq \mathbf{z}_j(2)$, $j = 1, 2$, então, retorna para o

Passo 2.

K-médias

Passo 2: Está gerando mesmo resultado como na iteração anterior: $\mathbf{z}_1(4) = \mathbf{z}_1(3)$ e $\mathbf{z}_2(4) = \mathbf{z}_2(3)$

Passo 3: Atualiza os centros de cluster (mesmo resultado)

Passo 4: O algoritmo converge e gera os seguinte centros de cluster:

$$\mathbf{z}_1 = \begin{pmatrix} 1.25 \\ 1.13 \end{pmatrix} \qquad \mathbf{z}_2 = \begin{pmatrix} 7.67 \\ 7.33 \end{pmatrix}$$

K-Média Nebuloso

O algoritmo de *k*-media resulta uma classificação com limite que determinam se um determinado pertence ou não a uma determinada classe. No entanto, algumas aplicações é desejável que tais limites sejam substituídos por grau de pertinência.

K-Média Nebuloso

O grau de pertinência do objeto *r* a um grupo *j* é dado por:

$$P\left(C_j \mid \mathbf{x}_r\right) = \frac{\left(1/d_{rj}\right)^{1/(b-1)}}{\sum_{i=1}^{K}\left(1/d_{ri}\right)^{1/(b-1)}} \qquad d_{rj} = \left\|\mathbf{x}_r - \mathbf{z}_j\right\|$$

O parâmetro *b* ajusta a mistura entre os grupos. Se *b* for 0, por exemplo, não haverá misturas e o algoritmo é um *k*-medias convencional. No entanto, conforme *b* aumenta a fronteira entre os grupos fica mais nebulosa.

K-Média Nebuloso

A seguinte condição deve ser satisfeita:

$$\sum_{j=1}^{K} P\left(C_j \mid \mathbf{x}_r\right) = 1$$

K-Média Nebuloso

Assim, na inicialização, o grau de pertinência de cada cluster deve ser determinado. Esta função de pertinência indicara uma probabilidade de um objeto pertencer a um determinado grupo. A cada iteração do algoritmo *k*-media nebuloso, os valores da função se alteram, pois os centros **z** também se alterarão conforme:

$$\mathbf{z}_j(k) = \frac{\sum_{r=1}^{n} \left[P\left(C_j \mid \mathbf{x}_r\right)\right]^b \mathbf{x}_r}{\sum_{r=1}^{n} \left[P\left(C_j \mid \mathbf{x}_r\right)\right]^b}$$

Radial Basis Function

Interpolation Problem

Given a set of N different points $\{\mathbf{x}_i \in R^m | i = 1, 2, \ldots, N\}$ and a Set of N real numbers $\{d_i \in R^1 | i = 1, 2, \ldots, N\}$, find a function $F: R^m \rightarrow R^1$ that satisfies the interpolation condition:

$$F(\mathbf{x}_i) = d_i$$

Redial Basis Function

Interpolation Problem

Specifically, we want the function F to be represented by the following form:

$$F(\mathbf{x}) = \sum_{i=1}^{m} w_i \varphi_i(\mathbf{x})$$

$$\varphi_i(\mathbf{x}) = G(\|\mathbf{x} - \mathbf{x}_i\|)$$

$\varphi_i(\mathbf{x})$: Redial basis functions

$\mathbf{x}_i$: Centers of redial basis function

The most used redial basis function – Gaussian function:

$$\varphi(r) = \exp\left(-\frac{r^2}{2\delta^2}\right)$$

Redial Basis Function

$$\Phi_s(F) = \frac{1}{2}\sum_{i=1}^{N}(d_i - F(\mathbf{x}_i))^2 = \frac{1}{2}\sum_{i=1}^{N}\left(d_i - \sum_{j=1}^{N} w_j G(\|\mathbf{x}, \mathbf{x}_i\|)\right)^2$$

$$\Phi_c(F) = \frac{1}{2}\|\mathbf{D}F\|^2$$

$\Phi_s(F)$: Standard error term

$\Phi_c(F)$: Regularizing term

Redial Basis Function

Minimize the following function:

$$\Phi(F) = \Phi_s(F) + \lambda\Phi_c(F)$$
$$= \sum_{i=1}^{N}\left(d_i - F(\mathbf{x}_i)\right)^2 + \lambda\|\mathbf{D}F\|^2$$

λ: regularizing parameter

Redial Basis Function

Frechet differential:

$$d\Phi(F,h) = \left[\frac{d}{d\beta} \Phi(F + \beta h) \right]_{\beta=0}$$

In elementary calculus, the tangent to a curve is a stright line that gives the best approximation of the curve in the neighborhood of the point of tangency. Thus the Frechet differential of a functional may be interpreted as the local best linear approximation. A necessary condition for the function $F(\mathbf{x})$ to be a relative extremum of the functional $\Phi(F)$ is that its Frechet differential must be zero, i.e.:

$$d\Phi(F,h) = d\Phi_s(F,h) + \lambda d\Phi_c(F,h) = 0$$

Redial Basis Function

Frechet differential of the standard error term:

$$d\Phi_s(F,h) = \left[\frac{1}{2}\frac{d}{d\beta}\sum_{i=1}^{N}\left[d_i - F(\mathbf{x}_i) - \beta h(\mathbf{x}_i)\right]^2\right]_{\beta=0}$$

$$= -\sum_{i=1}^{N}\left[d_i - F(\mathbf{x}_i) - \beta h(\mathbf{x}_i)\right] h(\mathbf{x}_i)\,|_{\beta=0}$$

$$= -\sum_{i=1}^{N}\left[d_i - F(\mathbf{x}_i)\right] h(\mathbf{x}_i)$$

$$= -\left(h, \sum_{i=1}^{N}\left[d_i - F\right]\delta_{\mathbf{x}_i}\right)_H$$

Inner product of two functions in H space.

Redial Basis Function

Frechet differential of the regularizing term:

$$d\Phi_c(F,h) = \frac{1}{2}\frac{d}{d\beta}\int_{R^m}\left(\mathbf{D}\left(F+\beta h\right)\right)^2 d\mathbf{x}\,|_{\beta=0}$$

$$= \int_{R^m}\left(\mathbf{D}\left(F+\beta h\right)\right)\mathbf{D}h\,d\mathbf{x}\,|_{\beta=0}$$

$$= \int_{R^m}\mathbf{D}F\mathbf{D}h\,dx$$

$$= \left(\mathbf{D}h, \mathbf{D}F\right)_H$$

Redial Basis Function

Green's identity

$$\int_{R^m} u(\mathbf{x})\mathbf{D}v(\mathbf{x})d\mathbf{x} = \int_{R^m} v(\mathbf{x})\tilde{\mathbf{D}}u(\mathbf{x})d\mathbf{x}$$

$\tilde{\mathbf{D}}$: adjoint operator of $\mathbf{D}$

$u(\mathbf{x})$ and $v(\mathbf{x})$ are differentiable functions.

Redial Basis Function

Frechet differential of the regularizing term:

$$d\Phi_c(F,h) = (\mathbf{D}h, \mathbf{D}F)_H$$

$$= \int_{R^m} h(\mathbf{x})\, \tilde{\mathbf{D}}\, \mathbf{D}F d\mathbf{x}$$

$$= \left(h, \tilde{\mathbf{D}}\, \mathbf{D}F \right)_H$$

Redial Basis Function

Frechet differential of the whole function:

$$d\Phi(F,h) = \left(h, \left[\tilde{\mathbf{D}}\,\mathbf{D}F - \frac{1}{\lambda}\sum_{i=1}^{N}(d_i - F)\delta_{\mathbf{x}_i} \right] \right)_H$$

The Frechet differential is zero for every $h(\mathbf{x})$ in H space if and only if the following condition holds:

$$\tilde{\mathbf{D}}\,\mathbf{D}F - \frac{1}{\lambda}\sum_{i=1}^{N}(d_i - F)\delta_{\mathbf{x}_i} = 0$$

or

$$\tilde{\mathbf{D}}\,\mathbf{D}F_\lambda(\mathbf{x}) = \frac{1}{\lambda}\sum_{i=1}^{N}(d_i - F(\mathbf{x}_i))\delta(\mathbf{x} - \mathbf{x}_i)$$

Green's function $G(\mathbf{x},\boldsymbol{\xi})$

satisfies the following partial differential equation everywhere except at the point $\mathbf{x}=\boldsymbol{\xi}$, where it has a singularity:

$$\mathbf{L}G(\mathbf{x},\boldsymbol{\xi})=0$$

i.e.:

$$\mathbf{L}G(\mathbf{x},\boldsymbol{\xi})=\delta(\mathbf{x}-\boldsymbol{\xi})$$

Let $\varphi(\boldsymbol{\xi})$ a continuous or piecewise continuous function

Then the function $F(\mathbf{x})=\int\limits_{R^m}G(\mathbf{x},\boldsymbol{\xi})\varphi(\boldsymbol{\xi})d\boldsymbol{\xi}$

is a solution of the following equation

$$\mathbf{L}F(\mathbf{x})=\varphi(\mathbf{x})$$

$$\mathbf{L}F(\mathbf{x}) = \mathbf{L}\int_{R^m} G(\mathbf{x},\boldsymbol{\xi})\varphi(\boldsymbol{\xi})\,d\boldsymbol{\xi}$$

$$= \int_{R^m} \mathbf{L}G(\mathbf{x},\boldsymbol{\xi})\varphi(\boldsymbol{\xi})\,d\boldsymbol{\xi}$$

$$= \int_{R^m} \delta(\mathbf{x}-\boldsymbol{\xi})\varphi(\boldsymbol{\xi})\,d\boldsymbol{\xi}$$

Sifty property

$$\int_{R^m} \delta(\mathbf{x}-\boldsymbol{\xi})\varphi(\boldsymbol{\xi})\,d\boldsymbol{\xi} = \varphi(\mathbf{x})$$

Solution to the regularization problem:

Let

$$\mathbf{L} = \tilde{D} D$$

$$\varphi(\boldsymbol{\xi}) = \frac{1}{\lambda} \sum_{i=1}^{N} \left(d_i - F(\mathbf{x}_i)\right) \delta(\boldsymbol{\xi} - \mathbf{x}_i)$$

Solution to the regularization problem:

$$F_\lambda(\mathbf{x}) = \int_{R^m} G(\mathbf{x}, \boldsymbol{\xi}) \left\{ \frac{1}{\lambda} \sum_{i=1}^{N} [d_i - F(x_i)] \delta(\boldsymbol{\xi} - \mathbf{x}_i) \right\} d\boldsymbol{\xi}$$

$$= \frac{1}{\lambda} \sum_{i=1}^{N} [d_i - F(x_i)] \int_{R^m} G(\mathbf{x}, \boldsymbol{\xi}) \delta(\boldsymbol{\xi} - \mathbf{x}_i) d\boldsymbol{\xi}$$

$$= \frac{1}{\lambda} \sum_{i=1}^{N} [d_i - F(x_i)] G(\mathbf{x}, \mathbf{x}_i)$$

So we have

$$w_i = \frac{1}{\lambda}\left[d_i - F\left(\mathbf{x}_i\right)\right]$$

$$F_\lambda\left(\mathbf{x}\right) = \sum_{i=1}^{m} w_i G\left(\mathbf{x}, \mathbf{x}_i\right)$$

Introducing the following definition:

$$\mathbf{F}_\lambda = \left[F_\lambda(\mathbf{x}_1), F_\lambda(\mathbf{x}_2), \ldots, F_\lambda(\mathbf{x}_N)\right]^T$$

$$\mathbf{d} = \left[d_1, d_2, \ldots, d_N\right]^T$$

$$\mathbf{G} = \begin{bmatrix} G(\mathbf{x}_1, \mathbf{x}_1) & G(\mathbf{x}_1, \mathbf{x}_2) & \ldots & G(\mathbf{x}_1, \mathbf{x}_N) \\ G(\mathbf{x}_2, \mathbf{x}_2) & G(\mathbf{x}_2, \mathbf{x}_2) & \ldots & G(\mathbf{x}_2, \mathbf{x}_N) \\ \ldots & \ldots & \ldots & \ldots \\ G(\mathbf{x}_N, \mathbf{x}_1) & G(\mathbf{x}_N, \mathbf{x}_2) & \ldots & G(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}$$

$$\mathbf{w} = \left[w_1, w_2, \ldots, w_N\right]^T$$

We have

$$\mathbf{w} = \frac{1}{\lambda}\left[\mathbf{d} - \mathbf{F}_\lambda\right]$$

$$\mathbf{F}_\lambda = \mathbf{G}\mathbf{w}$$

or

$$\left(\mathbf{G} + \lambda\mathbf{I}\right)\mathbf{w} = \mathbf{d}$$

Finally

$$\mathbf{w} = \left(\mathbf{G} + \lambda\mathbf{I}\right)^{-1}\mathbf{d}$$