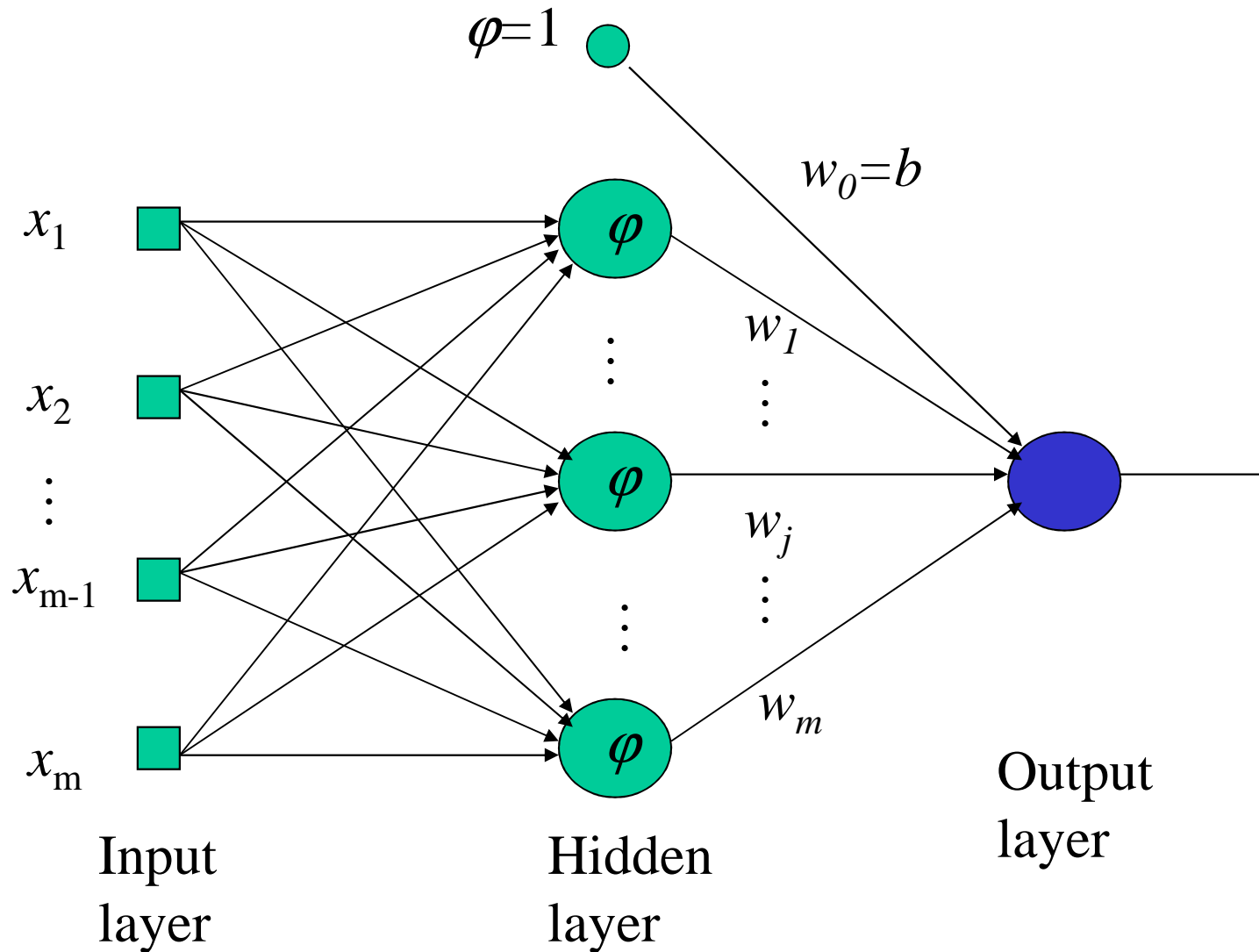


Radial Basis Function



Radial Basis Function

$$F(\mathbf{x}) = \sum_{i=1}^m w_i \varphi_i(\mathbf{x})$$
$$\varphi_i(\mathbf{x}) = G(\|\mathbf{x} - \mathbf{t}_i\|)$$

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

$\Phi_s(F)$: Standard error term

$\Phi_c(F)$: Regularizing term

Radial Basis Function

define

$$\mathbf{d} = [d_1, d_2, \dots, d_N]^T$$

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

$$\mathbf{w} = [w_1, w_2, \dots, w_m]^T$$

Radial Basis Function

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

Minimizing the following function

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

We have

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

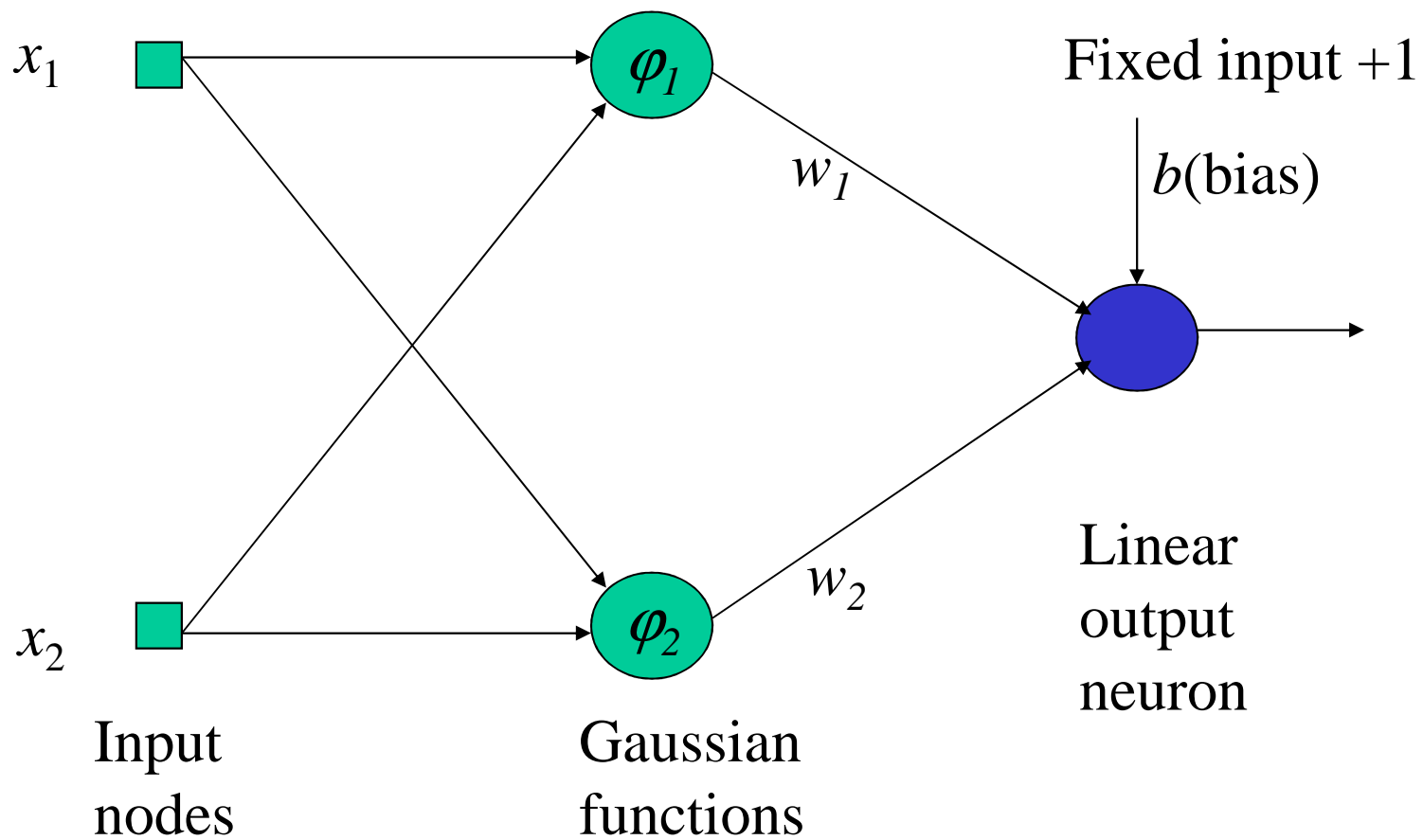
or

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

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

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

$$t_1 = [1,1]$$

$$t_2 = [0,0]$$

$$y(\mathbf{x}) = \sum_{i=1}^2 w_i G(\|\mathbf{x} - \mathbf{t}_i\|) + b$$

$$G(\|\mathbf{x} - \mathbf{t}_i\|) = \exp(-\|\mathbf{x} - \mathbf{t}_i\|^2)$$

$$y(\mathbf{x}_j) = d_j$$

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

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

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

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

Self-organized Selection of Centers

Algoritmo

Passo 1: Escolher K centros iniciais de grupos que podem

ser arbitrariamente escolhidos do conjunto de objetos

$\mathbf{s} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ a ser classificados (n é o número 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

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

K-médias

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

assumindo o critério:

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

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

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

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

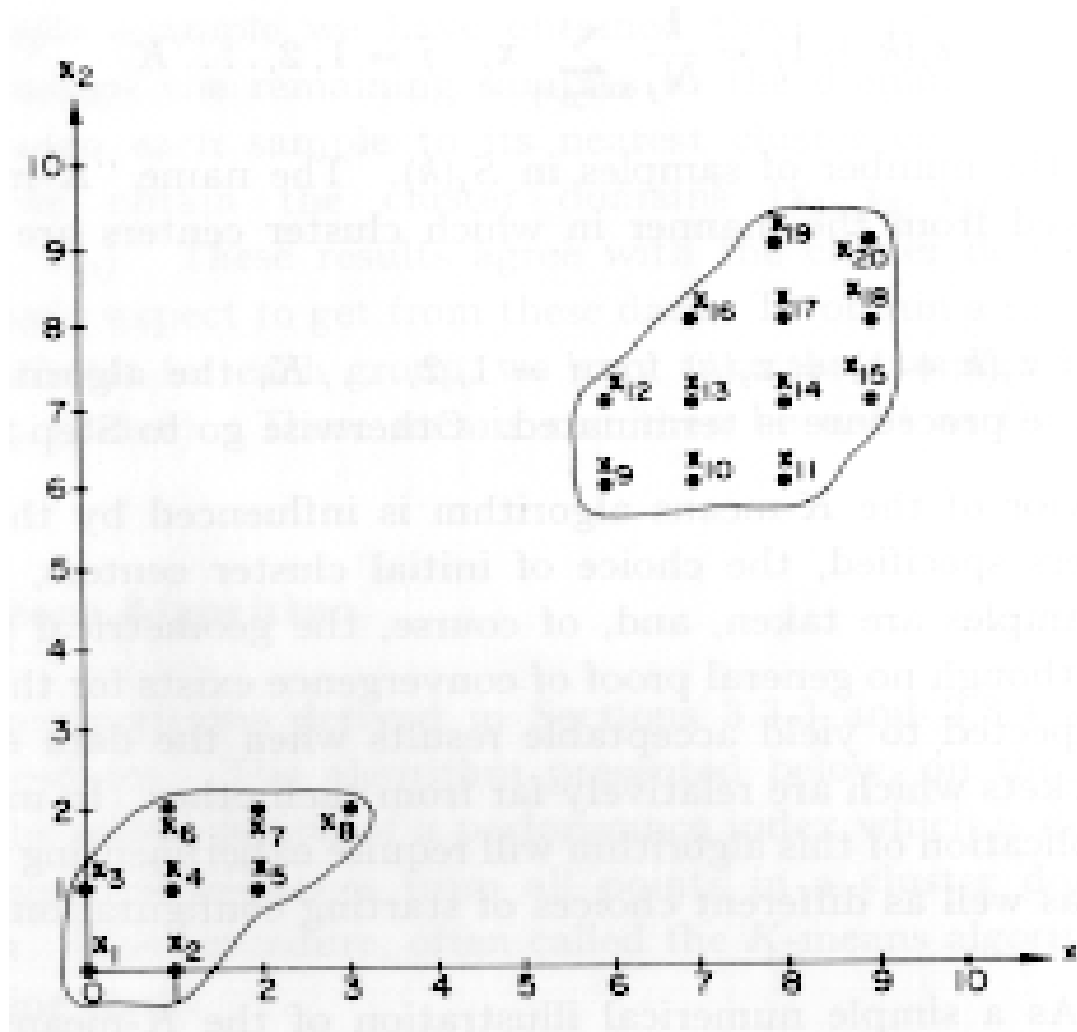
Onde N_j é o numero de amostras em $S_j(t)$

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 deverá 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:

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

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

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

K-médias

Passo 3: Atualiza os centros de cluster

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

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

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

Passo 2.

K-médias

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

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

$$\|\mathbf{x}_j - \mathbf{z}_2(2)\| < \|\mathbf{x}_j - \mathbf{z}_1(2)\|, \text{ para } j = 9, 10, \dots, 20,$$

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

K-médias

Passo 3: Atualiza os centros de cluster

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

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

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:

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

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:

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

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:

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

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 \mathbf{z} também se alterarão conforme:

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

Radial Basis Function

Interpolation Problem

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

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

Radial Basis Function

Interpolation Problem

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

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

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

\mathbf{x}_i : Centers of radial basis function

The most used radial basis function – Gaussian function:

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

Radial 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

Radial Basis Function

Minimize the following function:

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

λ : regularizing parameter

Radial 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 straight 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.:

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

Radial Basis Function

Frechet differential of the standard error term:

$$\begin{aligned}d\Phi_s(F, h) &= \left[\frac{1}{2} \frac{d}{d\beta} \sum_{i=1}^N [d_i - F(\mathbf{x}_i) - \beta h(\mathbf{x}_i)]^2 \right]_{\beta=0} \\&= - \sum_{i=1}^N [d_i - F(\mathbf{x}_i) - \beta h(\mathbf{x}_i)] h(\mathbf{x}_i) \Big|_{\beta=0} \\&= - \sum_{i=1}^N [d_i - F(\mathbf{x}_i)] h(\mathbf{x}_i) \\&= - \left(h, \sum_{i=1}^N [d_i - F] \delta_{\mathbf{x}_i} \right)_H\end{aligned}$$

Inner product of two functions in H space.

Radial Basis Function

Frechet differential of the regularizing term:

$$\begin{aligned}d\Phi_c(F, h) &= \frac{1}{2} \frac{d}{d\beta} \int_{R^m} (\mathbf{D}(F + \beta h))^2 d\mathbf{x} \Big|_{\beta=0} \\ &= \int_{R^m} (\mathbf{D}(F + \beta h)) \mathbf{D} h d\mathbf{x} \Big|_{\beta=0} \\ &= \int_{R^m} \mathbf{D} F \mathbf{D} h dx \\ &= (\mathbf{D} h, \mathbf{D} F)_H\end{aligned}$$

Radial 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.

Radial Basis Function

Frechet differential of the regularizing term:

$$\begin{aligned}d\Phi_c(F, h) &= (\mathbf{D}h, \mathbf{D}F)_H \\ &= \int_{R^m} h(\mathbf{x}) \tilde{\mathbf{D}} \mathbf{D}F d\mathbf{x} \\ &= \left(h, \tilde{\mathbf{D}} \mathbf{D}F \right)_H\end{aligned}$$

Radial 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 $G(\mathbf{x}, \xi)$

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

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

i.e.:

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

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

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

is a solution of the following equation

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

$$\begin{aligned}\mathbf{L}F(\mathbf{x}) &= \mathbf{L} \int_{R^m} G(\mathbf{x}, \xi) \varphi(\xi) d\xi \\ &= \int_{R^m} \mathbf{L}G(\mathbf{x}, \xi) \varphi(\xi) d\xi \\ &= \int_{R^m} \delta(\mathbf{x} - \xi) \varphi(\xi) d\xi\end{aligned}$$

Sifty property

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

Solution to the regularization problem:

Let

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

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

Solution to the regularization problem:

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

So we have

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

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

Introducing the following definition:

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

$$\mathbf{d} = [d_1, d_2, \dots, d_N]^T$$

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

$$\mathbf{w} = [w_1, w_2, \dots, w_N]^T$$

We have

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

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

or

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

Finally

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