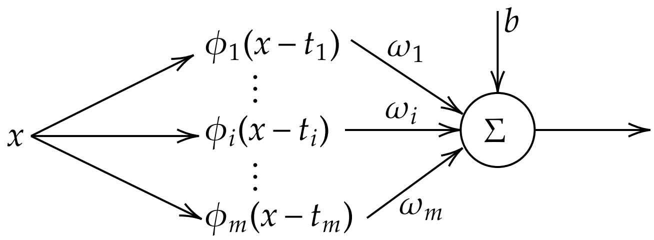
1. 网络结构及输入输出的关系

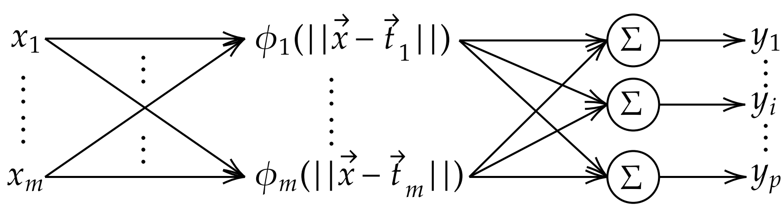
1.1. 单输入、单输出



$b=0$ 时，$y=\sum_{i=1}^{m}\omega_i\phi_i(x-t_i)$

其中 ϕ_i 为径向基函数，如 $\phi(r)=e^{-\frac{r^2}{\sigma^2}}$ 或 $\phi(r)=\frac{1}{\sqrt{r^2+c^2}}$

1.2. 多输入、多输出

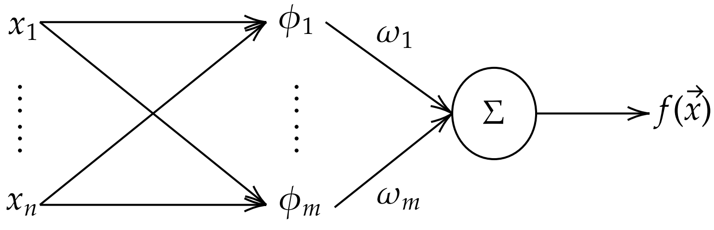


输入与输出之间的关系：$y_i=\sum_{k=1}^{m}\omega_{ik}(\phi_k(||\vec{x}-\vec{t}_k||)\quad i=1,2,\cdots,p\quad \vec{x}\in R^n$

2. RBF 用于函数逼近

已知 N 个 n 维样本 $\left\{\vec{x}^p\in R^n\middle|p=1,\cdots,N\right\}$，对应 N 个实数 $\left\{d^i\middle|i=1,\cdots,N\right\}$

逼近函数 $f:R^n\rightarrow R^1\quad f\left(\vec{x}^i\right)=d^i\ \ i=1,2,\cdots,N$



$f(\vec{x})=\sum_{k=1}^{m}\omega_k\phi_k(||\vec{x}-\vec{t}_k||)$

对于 N 个 n 维的输入，可以得到 $[\Phi]_{N\times M}$ —— 核矩阵

$$\begin{bmatrix} \phi_1^1 & \phi_2^1 & \cdots & \phi_m^1 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1^i & \phi_2^i & \cdots & \phi_m^i \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1^N & \phi_2^N & \cdots & \phi_m^N \end{bmatrix} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_i \\ \vdots \\ \omega_m \end{bmatrix} = \begin{bmatrix} d^1 \\ \vdots \\ d^i \\ \vdots \\ d^m \end{bmatrix} \text{ 其中 } \phi_i^j = \phi\left(\left|\left|\vec{x}^j - \vec{t}_i\right|\right|\right)$$

$[\Phi]\vec{\omega} = \vec{d} \quad \Rightarrow \quad \vec{\omega} = [\Phi]^{-1}\vec{d}$，其中 $[\Phi]^{-1} = \left([\Phi]^\top[\Phi]\right)^{-1}[\Phi]^\top\vec{d}$

3. 用 RBF 进行分类

分属于 2 类的 N 个训练样本 $\left\{\vec{x}^1, \cdots, \vec{x}^N\right\}, \vec{x}^p \in R^n, p = 1, 2, \cdots, N$ 在 R^n 线性不可分

m 个 RBF 函数 $\vec{\phi}(\vec{x}) = (\phi_1(\vec{x}), \phi_2(\vec{x}), \cdots, \phi_m(\vec{x}))^\top$

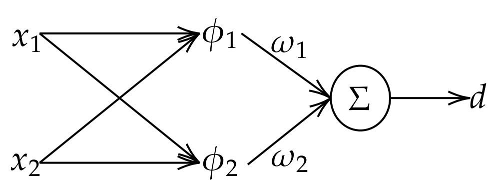
如果 $\vec{x}$ 经过 $\vec{\phi}$ 的非线性映射，形成了一个线性可分的分布，即

$$\vec{\omega}^\top\vec{\phi}(\vec{x}) \geqslant 0 \quad \vec{x} \in C_1$$
$$\vec{\omega}^\top\vec{\phi}(\vec{x}) < 0 \quad \vec{x} \in C_2$$

分界面 $\vec{\omega}^\top\vec{\phi}(\vec{x}) = 0$

例：用 RBF 求解 XOR 问题

解：变量为 x_1, x_2, d，d 是 x_1 和 x_2 的异或结果



x_1	x_2	d
0	0	0
0	1	1
1	0	1
1	1	0

取 $\phi_1(\vec{x}) = e^{-||\vec{x}-\vec{t}_k||^2}, \phi_2(\vec{x}) = e^{-||\vec{x}-\vec{t}_2||^2}, \vec{t}_1 = (1,1)^\top, \vec{t}_2 = (0,0)^\top$

x_1	x_2	ϕ_1	ϕ_2	d
0	0	e^{-2}	e^0	0
0	1	e^{-1}	e^{-1}	1
1	0	e^{-1}	e^{-1}	1
1	1	e^0	e^{-2}	0

$$\begin{bmatrix} e^{-2} & 1 \\ e^{-1} & e^{-1} \\ e^{-1} & e^{-1} \\ 1 & e^{-2} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} = \left([\phi]^\top [\phi]\right)^{-1} [\phi]^\top \vec{d}$$

x 空间线性不可分 $\xrightarrow{\text{映射}}$ ϕ 空间线性可分

4. RBF 常用的学习方法

常用 Gauss-RBF 核函数时，需要确定 ① δ ② 中心值 $\vec{t}_k$ ③ $\vec{\omega}^i$

4.1. 由样本中选出固定的中心值 $\vec{t}_k$

$\delta = \dfrac{d_{max}}{\sqrt{k}}$ k：中心的个数 d_{max}：中心间最大距离

$$\phi_k(||\vec{x} - \vec{t}_k||) = \exp\left(-\frac{||\vec{x} - \vec{t}_k||^2}{\delta^2}\right) = \exp\left(-\frac{k}{d_{max}^2} ||\vec{x} - \vec{t}_k||^2\right)$$

4.2. 聚类法

- 随机选取 m 个聚类中心 $\vec{t}_k, k = 1, 2, \cdots, m$，样本集 $\left\{\vec{x}^p, p = 1, 2, \cdots, N\right\}$
- $\mathrm{J} = \sum_{k=1}^{m} \sum_{\vec{x}^p \in C_{\vec{t}_k}} \left|\left|\vec{x}^p - \vec{t}_k\right|\right|^2 \to \min$

  $\vec{x}^p \in C_{\vec{t}_k}$ 表示 $\vec{x}^p$ 离中心最近，$\vec{C}_{\vec{t}_k}$ 表示以 $\vec{t}_k$ 为中心的类
- 求样本中心，$\vec{t}_k = \dfrac{1}{N_k} \sum_{\vec{x}^p \in C_{\vec{t}_k}} \vec{x}^p$，$N_k$ 为 $C_{\vec{t}_k}$ 中样本的数目
- 重复二三两步，直至 $\mathrm{J} \to \min$
- $\sigma_k^2 = \sum_{\vec{x}^p \in C_{\vec{t}_k}} \dfrac{\left|\left|\vec{x}^p - \vec{t}_k\right|\right|}{N_k}$

4.3. 自组织法

- 随机初始化 $\vec{t}_k(0), k = 1, 2, \cdots, m$
- 训练样本，对每个中心求距离，确定归属
- 学习 $\begin{cases} \vec{t}_k(t) = \vec{t}_k(t-1) + \eta\left(\vec{x}^p(t) - \vec{t}_k(t)\right) & \vec{x}^p \in C_{\vec{t}_k} \\ \vec{t}_k(t) = \vec{t}_k(t-1) & \vec{x}^p \notin C_{\vec{t}_k} \end{cases}$
- 重复二三两步，直至 $\vec{t}_k(t) = \vec{t}_k(t-1)$
- $\delta_k^2 = \sum_{\vec{x} \in C_{\vec{t}_k}} \frac{||\vec{x} - \vec{t}_k||^2}{N_k}$

4.4. 梯度法

目标函数 $J(\vec{t}, \vec{d}) = \frac{1}{2}\sum_{p=1}^{N}\left(d^p - y^p\right)^2 \to \min \quad y^p = \sum_{k=1}^{m} \omega_k \phi_k\left(\left|\left|\vec{x}^p - \vec{t}_k\right|\right|\right)$

$$\frac{\partial J}{\partial \vec{t}_k} = \sum_{k=1}^{N} \omega_k \left(d^p - y^p\right) \exp\left(-\frac{\left|\left|\vec{x}^p - \vec{t}_k\right|\right|^2}{\delta_k^2}\right) \frac{2\left(\vec{x}^p - \vec{t}_k\right)}{\delta_k^2}$$

$$\frac{\partial J}{\partial \delta_k} = -\sum_{p=1}^{N} \omega_k \left(d^p - y^p\right) \exp\left(-\frac{\left|\left|\vec{x}^p - \vec{t}_k\right|\right|^2}{\delta_k^2}\right) \frac{2\left|\left|\vec{x}^p - \vec{t}_k\right|\right|}{\delta_k^2}$$

$$\Delta \vec{t}_k = -\eta_t \frac{\partial J}{\partial \vec{t}_k} \qquad \Delta \delta_k = -\eta_\delta \frac{\partial J}{\partial \delta_k} \qquad \eta_t, \eta_\delta > 0$$

4.5. $\vec{\omega}$ 的计算

1. $\vec{\omega} = [\phi]^{-1}\vec{d}$，其中 $[\phi]^{-1} = \left([\phi]^\top [\phi]\right)^{-1} [\phi]^\top$

2. 梯度法

$$J(\vec{t}, \vec{d}) = \frac{1}{2}\sum_{p=1}^{m}\left(d^p - y^p\right)^2, y^p = \sum_{k=1}^{m} \omega_k \phi_k\left(\left|\left|\vec{x}^p - \vec{t}_k\right|\right|\right)$$

$$\frac{\partial J}{\partial \vec{\omega}_k} = \sum_{p=1}^{m}\left(d^p - y^p\right) \cdot \left(-\frac{\partial y^p}{\partial \omega_k}\right) = -\sum_{p=1}^{m}\left(d^p - y^p\right) \sum_{k=1}^{m} \phi_k\left(\left|\left|\vec{x}^p - \vec{t}_k\right|\right|\right)$$