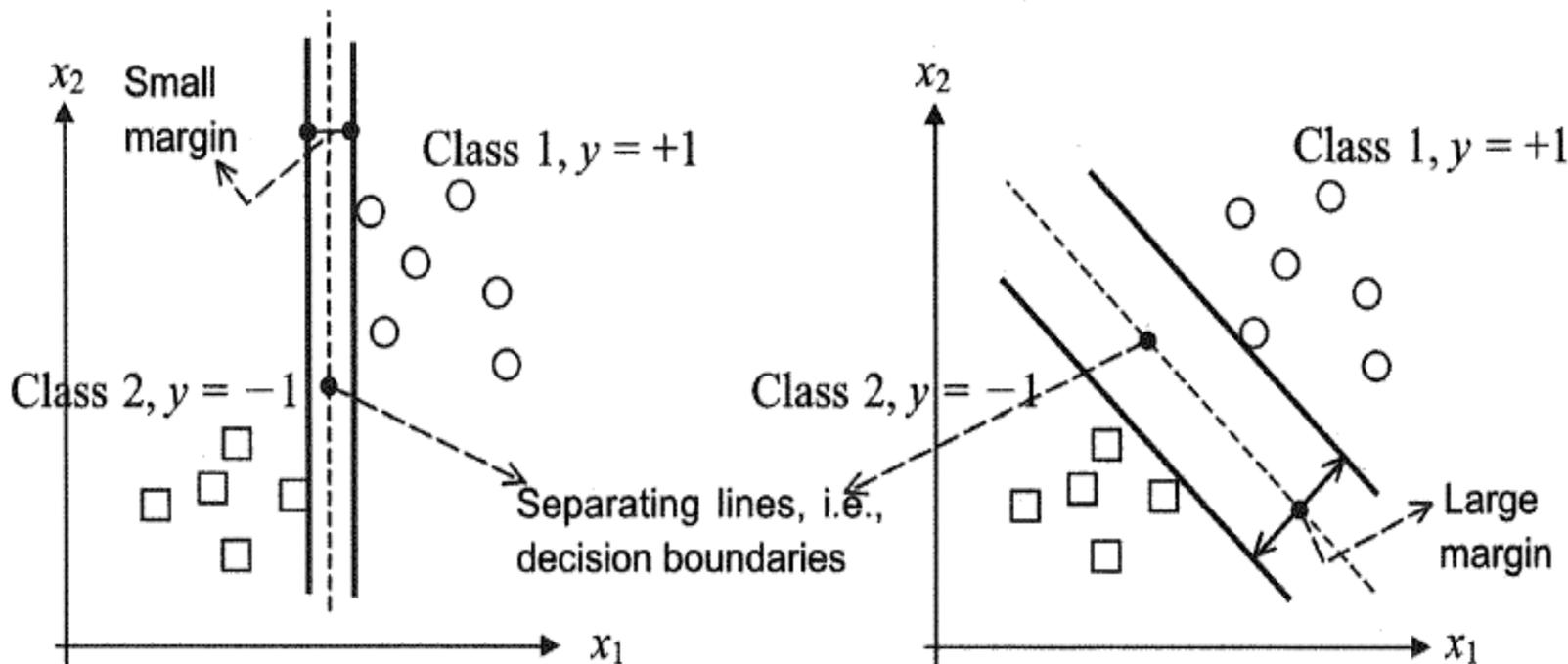




支持向量机(SVM)网络



- 最优线性分界面（二分类问题）
 - 对线性可分集，总能找到使样本正确划分的分界面，而且有无穷多个，哪个是最优的？
 - 一种最优的分界准则（从对样本及参数的鲁棒性看）是使两类模式向量分开的间隔最大。





- 最优线性分界面的确定

- 两分类的线性判别函数的一般表达式为: $g(x) = w^T x + b$ 。
- 方程 $g(x) = 0$ 定义了一个超平面 H , 它把两类训练样本完全分开。设 N 个样本为: $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$, $x_i \in R^n$, $y_i \in \{+1, -1\}$, 则有分类规则:

$$w^T x_i + b \geq 0, \quad y_i = +1$$

$$w^T x_i + b < 0, \quad y_i = -1$$

- 由于训练集线性可分, 改变权向量的摸, 总可改写分类规则为:

$$w^T x_i + b \geq \delta > 0 \quad \longrightarrow \quad w^T x_i + b \geq 1, \quad y_i = +1$$

$$w^T x_i + b \leq -\delta < 0 \quad \longrightarrow \quad w^T x_i + b \leq -1, \quad y_i = -1$$

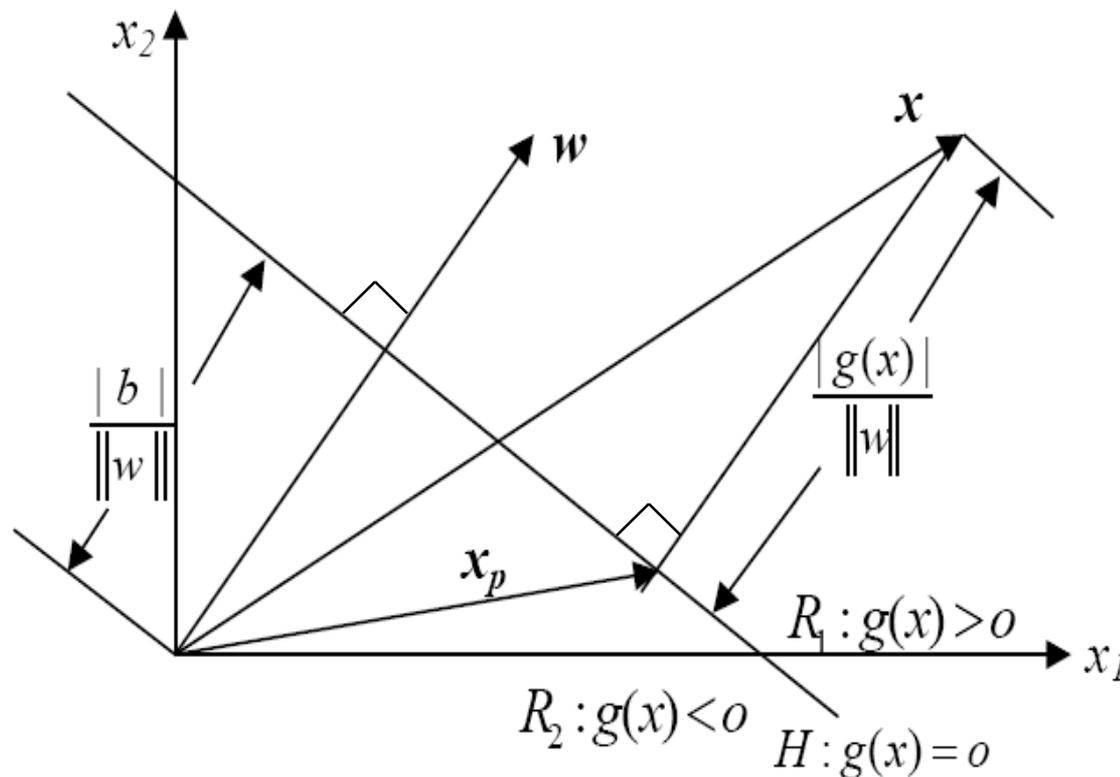
- 进一步合并有紧凑式:

$$y_i(w^T x_i + b) \geq 1 \quad \text{同样还有: } |g(x_i)| \geq 1$$



- 最优线性分界面的确定（续）

- $g(x)$ 可以看成是从 x 到超平面的距离的一种度量，见图。
把 x 表示成： $x = x_p + r w / \|w\|$ ，其中 r 是 x 到 H 的垂直距离，
则有： $g(x) = w^T (x_p + r w / \|w\|) + b = r \|w\|$ ，即： $r = g(x) / \|w\|$





- 最优线性分界面的确定（续）

- **间隔**：离分界面 H 最近的样本点(即使 $|g(x)|=1$ 的样本点)与分界面的距离，它是 $g(x)/\|w\|=1/\|w\|$ 。这样，两类模式间隔的距离为 $2/\|w\|$ 。

- **最优分界面**：为使两类间隔最大，应使 $\|w\|$ 最小，等价于使 $\|w\|^2$ 最小。所以，最优分界面应满足：

$$\min_w \frac{1}{2} w^T w \quad \text{和} \quad y_i(w^T x_i + b) \geq 1$$

- **支持向量**：距离最优分界面最近的位于间隔边界上的那些样本向量，也就是使得等号 $y_i(w^T x_i + b) = 1$ 或 $|g(x_i)| = 1$ 成立的那些样本向量。



- 最优分界面的求解

- 用Lagrange乘子法最小化代价函数： $\Psi(w) = \frac{1}{2}(w^T w)$ 。

- 构造Lagrange函数：

$$L(w, b, \alpha) = \frac{1}{2}(w^T w) - \sum_{i=1}^N \alpha_i [y_i (w^T x_i + b) - 1]$$

其中 $\alpha_i \geq 0$ 为Lagrange乘子，达到极值的必要条件为：

必要条件1： $\frac{\partial L(w, b, \alpha)}{\partial w} = 0$ 即： $w - \sum_{i=1}^N \alpha_i y_i x_i = 0$

必要条件2： $\frac{\partial L(w, b, \alpha)}{\partial b} = 0$ 即： $\sum_{i=1}^N \alpha_i y_i = 0$

- 从最优化理论的KTT条件得出解必须满足：

$$\alpha_i [y_i (w^T x_i + b) - 1] = 0 \quad i = 1, 2, \dots, N$$

- 从必要条件1看到，只有 $\alpha_i > 0$ 的样本对权起作用，而此时必有 $y_i (w^T x_i + b) - 1 = 0$ ，即相应的样本是支持向量。故解向量 w 是由支持向量构建的，它们决定分类结果。



- 最优分界面的求解（续）

- 根据Lagrange优化方法，用对偶定理求乘子 α_i 最优解。

- 展开Lagrange函数有：

$$L(w, b, \alpha) = \frac{1}{2} (w^T w) - \sum_{i=1}^N \alpha_i y_i w^T x_i - b \sum_{i=1}^N \alpha_i y_i + \sum_{i=1}^N \alpha_i$$

- 将 $w = \sum_{i=1}^N \alpha_i y_i x_i$ 和 $\sum_{i=1}^N \alpha_i y_i = 0$ 代入上式，则有：

$$Q(\alpha) = L(w, b, \alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i^T x_j$$

- 求上式的最大值，可得最优解 α_i^* ，则最优权向量为：

$$w^* = \sum_{i=1}^{N_s} \alpha_i^* y_i x_i \quad (N_s \text{ 是支持向量的个数})$$

- 最优偏置可选用一个支持向量样本求得： $b^* = y_i - w^{*T} x_i$ 。

- 最优分界面是：

$$w^{*T} x + b^* = 0$$



- 线性不可分问题

- 向高维空间（特征空间）映射

- 模式可分性的Cover定理

将复杂的模式分类问题非线性地投射到高维空间将比投射到低维空间更可能是线性可分的。

- 基本原理

通过某种非线性映射 $z = \Phi(x) = [\varphi_1(x), \varphi_2(x), \dots, \varphi_m(x)]^T$ 将样本映射到一个高维空间（特征空间），在这个高维空间中构造最优分类超平面：

$$w^T \Phi(x) + b = 0$$

在特征空间用线性可分的结果，即 $w = \sum_{i=1}^N \alpha_i y_i \Phi(x_i)$ 代入上式得

$$\sum_{i=1}^N \alpha_i y_i \Phi(x_i)^T \Phi(x) + b = 0$$

这样在映射到高维空间也只须进行内积运算，这是可以用原空间的函数实现的。根据泛函理论，只要核函数 $K(x_i, x) = \Phi(x_i)^T \Phi(x)$ 满足Mercer条件，它就对应某个变换空间中的内积。



- 线性不可分问题（续）

- 向高维空间（特征空间）映射

- 只要找到适当的核函数 $K(x_i, x_j)$ 就可实现某个非线性变换后的线性分类，类似线性可分情形有：

- 对 α_i 求以下函数的极大值

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

满足约束条件 $\sum_{i=1}^N \alpha_i y_i = 0$ 和 $\alpha_i \geq 0$ 。

- 设最优解为 α_i^* ，则最优分界面可写为：

$$\sum_{i=1}^N \alpha_i^* y_i K(x_i, x) + b^* = 0$$



- 线性不可分问题（续）

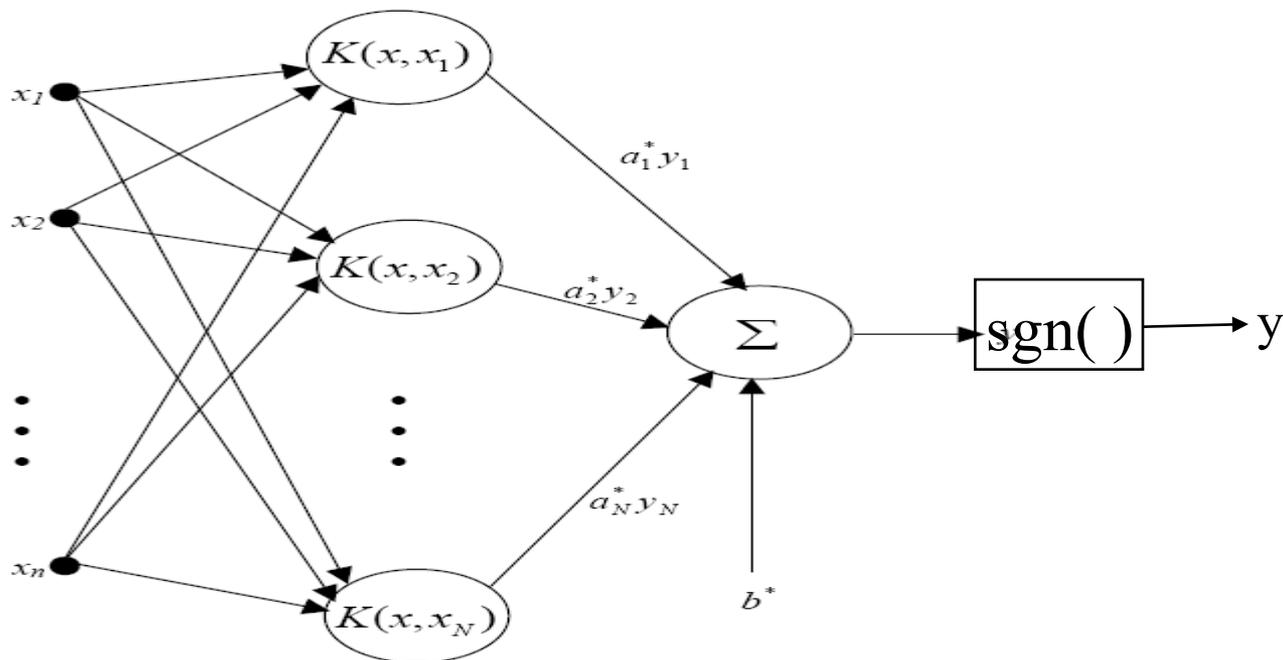
- 向高维空间（特征空间）映射

- 不同的核函数将形成不同的算法，常用的有：

- 多项式核函数： $K(x_i, x) = (x^T x_i + 1)^d$

- 径向基函数： $K(x_i, x) = \exp(-\frac{\|x - x_i\|}{\sigma^2})$

- 支持向量机的结构图





- 线性不可分问题（续）（映射后也不能保证线性可分）
 - 增加松弛项，使分界面在训练集上平均分类误差最小。
 - 原问题为：

寻找权向和偏置的最优值，使得它们满足约束条件

$$y_i(w^T x_i + b) \geq 1 - \varepsilon_i \text{ 和 } \varepsilon_i \geq 0, \quad i = 1, 2, \dots, N$$

其中松弛变量 $\varepsilon_i = 0$ 时， x_i 是支持向量

松弛变量 $0 < \varepsilon_i \leq 1$ ， x_i 落入间隔区，在分界面的正确一侧

松弛变量 $1 < \varepsilon_i$ 时， x_i 落入分界面的错误一侧

此时，使得权向量和松弛变量最小化代价函数为

$$\Psi(w, \varepsilon) = \frac{1}{2} w^T w + C \sum_{i=1}^N \varepsilon_i$$

其中 $C > 0$ 是个常数，由使用者选定控制对错分样本的惩罚程度



- 对偶问题为:

寻找Lagrange乘子最大化目标函数

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i^T x_j$$

满足约束条件

$$\sum_{i=1}^N \alpha_i y_i = 0 \text{ 和 } 0 \leq \alpha_i \leq C \text{ (此项与前面线性可分情况结果不同)}$$



• 支持向量机的设计算法

– 在能够进行变换 $\Phi(x)$ 的情况下:

(1) 在约束条件 $\sum_{i=1}^N \alpha_i y_i = 0$ 和 $\alpha_i \geq 0$ (或 $0 \leq \alpha_i \leq C$) 下求函数

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \Phi^T(x_i) \Phi(x_j) \text{ 极大值点 } \alpha^* ;$$

(2) 计算最优权值和偏置值: $w^* = \sum_{i=1}^{N_s} \alpha_i^* y_i \Phi(x_i)$, $b^* = y_i - w^{*T} x_i$;

(3) 支持向量机的最优分界面为: $\sum_{i=1}^N \alpha_i^* y_i \Phi^T(x_i) \Phi(x) + b^* = 0$ 。

– 在选择核函数 $K(x_i, x)$ 避免进行变换情况下, 不同处有:

(1) 求函数 $Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j)$ 极大值点 α^* ;

(2) 计算最优权值: $w^* = \sum_{i=1}^{N_s} \alpha_i^* y_i \hat{y}_i$, 其中 \hat{y}_i 是隐层输出;

(3) 支持向量机的最优分界面为: $\sum_{i=1}^N \alpha_i^* y_i K(x_i, x) + b^* = 0$



- 支持向量机的特点

- 适合对小样本数据的学习，注重样本自身信息，而非产生样本的规律（概率及条件概率等）。
- 网络结构简单，只有一个隐层，隐层的节点数由所求得的支持向量个数自动决定。
- 可以根据核函数的选择自动计算重要的网络参数。
- 在解决模式分类问题方面，能提供较好的泛化性能。
- 有些参数（如控制对错分样本惩罚程度的C）不易确定。
- 虽然可以不需知道非线性映射 $\Phi(\bullet)$ 的具体形式，但非线性映射的核函数 $K(x_i, x)$ 不易得。判定一个给定的核函数是否满足Mercer定理条件是一件困难事。
- 在待分类的模式为线性不可分时，怎样控制支持向量的选择是一个困难的问题。



- 支持向量机类型的RBF网络和MLP网络

- RBF网络

- Gauss核函数 $K(X, X_i) = \exp(-\frac{1}{\delta^2} \|X - X_i\|)$;
- δ 对所有核相同，由设计者预先指定；
- 隐层节点的数量由支持向量的个数自动决定；
- 中心由支持向量的值自动决定。

- 单隐层MLP网络

- Sigmoid核函数 $K(X, X_i) = \tanh(k_1 X^T X + k_2)$ ，其中只有一些特定的 k_1, k_2 值满足Mercer定理；
- 隐层节点的个数由支持向量的个数自动决定；
- 隐节点的权值由支持向量的值自动决定。



• 支持向量机设计举例 — XOR问题

– 训练样本: $\{((-1, -1), -1), ((-1, +1), +1), ((+1, -1), +1), ((+1, +1), -1)\}$

– 方法一:

- 选择非线性映射函数: $\Phi(X) = [1, x_1^2, \sqrt{2}x_1x_2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2]^T$ 将二维输入样本映射到一个六维特征空间。

(1) 求极值

$$\begin{aligned} Q(\alpha) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \Phi^T(X_i) \Phi(X_j) \\ &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2} (9\alpha_1^2 - 2\alpha_1\alpha_2 - 2\alpha_1\alpha_3 + 2\alpha_1\alpha_4 + \\ &\quad 9\alpha_2^2 + 2\alpha_2\alpha_3 - 2\alpha_2\alpha_4 + 9\alpha_3^2 - 2\alpha_3\alpha_4 + 9\alpha_4^2) \end{aligned}$$

满足约束条件:

$$-\alpha_1 + \alpha_2 + \alpha_3 - \alpha_4 = 0 \quad (\text{即必要条件2: } \sum_{i=1}^4 \alpha_i y_i = 0)$$

$$\alpha_i \geq 0 \quad i = 1, 2, 3, 4$$



• 支持向量机设计举例 — XOR问题（续）

– 方法一（续）：

$Q(\alpha)$ 对 α_i 求导并令导数为零，得到下列联立方程组：

$$9\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 1$$

$$-\alpha_1 - 9\alpha_2 + \alpha_3 + \alpha_4 = 1$$

$$-\alpha_1 + \alpha_2 + 9\alpha_3 - \alpha_4 = 1$$

$$\alpha_1 - \alpha_2 - \alpha_3 + 9\alpha_4 = 1$$

解得极值点为 $\alpha_i = \frac{1}{8}$, $i = 1, 2, 3, 4$ 。

可见4个样本都是支持向量。

(2) 计算最优权值和偏置值：

$$\begin{aligned} W^* &= \sum_{i=1}^{N_s} \alpha_i^* y_i \Phi(X_i) = \frac{1}{8} (-\Phi(X_1) + \Phi(X_2) + \Phi(X_3) - \Phi(X_4)) \\ &= -(1, 1, \sqrt{2}, 1, -\sqrt{2}, -\sqrt{2}) + (1, 1, -\sqrt{2}, 1, -\sqrt{2}, \sqrt{2}) + \\ &\quad (1, 1, -\sqrt{2}, 1, \sqrt{2}, -\sqrt{2}) - (1, 1, \sqrt{2}, 1, \sqrt{2}, \sqrt{2}) \\ &= (0, 0, -1/\sqrt{2}, 0, 0, 0) \end{aligned}$$



- 支持向量机设计举例 — XOR问题（续）

- 方法一（续）：

$$b^* = y_2 - w^{*T} \Phi(X_2)$$

$$= 1 - (0, 0, -1/\sqrt{2}, 0, 0, 0) \Phi(-1, 1)$$

$$= 1 - (0, 0, -1/\sqrt{2}, 0, 0, 0) (1, 1, -\sqrt{2}, 1, -\sqrt{2}, \sqrt{2})^T$$

$$= 1 - 1 = 0$$

(3) 支持向量机的最优分界面为：

$$(W^*)^T \Phi(X) + b^* = (0, 0, -1/\sqrt{2}, 0, 0, 0) (1, x_1^2, \sqrt{2}x_1x_2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2)^T$$

$$= -x_1x_2 = 0$$



• 支持向量机设计举例 — XOR问题（续）

– 方法二:

- 选择核函数为: $K(X, X_i) = (1 + X^T X_i)^2$

将 $X = (x_1, x_2)^T$ 和 $X_i = (x_{i1}, x_{i2})^T$ 代入上式, 核函数可表为:

$$K(X_i, X) = 1 + x_1^2 x_{i1}^2 + 2x_1 x_2 x_{i1} x_{i2} + x_2^2 x_{i2}^2 + 2x_1 x_{i1} + 2x_2 x_{i2}$$

将各训练样本代入, 可计算出4*4对称矩阵为:

$$\mathbf{K} = \begin{pmatrix} K(X_1, X_1) & K(X_1, X_2) & K(X_1, X_3) & K(X_1, X_4) \\ K(X_2, X_1) & K(X_2, X_2) & K(X_2, X_3) & K(X_2, X_4) \\ K(X_3, X_1) & K(X_3, X_2) & K(X_3, X_3) & K(X_3, X_4) \\ K(X_4, X_1) & K(X_4, X_2) & K(X_4, X_3) & K(X_4, X_4) \end{pmatrix} = \begin{pmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{pmatrix}$$

(1) 用拉格朗日乘子法求极大值点:

$$\begin{aligned} Q(\alpha) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(X_i, X_j) \\ &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2} (9\alpha_1^2 - 2\alpha_1\alpha_2 - 2\alpha_1\alpha_3 + 2\alpha_1\alpha_4 + \\ &\quad 9\alpha_2^2 + 2\alpha_2\alpha_3 - 2\alpha_2\alpha_4 + 9\alpha_3^2 - 2\alpha_3\alpha_4 + 9\alpha_4^2) \end{aligned}$$



• 支持向量机设计举例 — XOR问题（续）

– 方法二（续）：

$Q(\alpha)$ 对 α_i 求导并令导数为零，得到下列联立方程组：

$$9\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 1$$

$$-\alpha_1 - 9\alpha_2 + \alpha_3 + \alpha_4 = 1$$

$$-\alpha_1 + \alpha_2 + 9\alpha_3 - \alpha_4 = 1$$

$$\alpha_1 - \alpha_2 - \alpha_3 + 9\alpha_4 = 1$$

解得极值点： $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \frac{1}{8}$ 。由于4个样本都是支持向量，隐层应有4个节点，各节点的输出为：

$$\hat{y}_i = K(X_i, X) = (1 + X_i^T X)^2, \quad i = 1, 2, 3, 4$$

$$(2) \text{ 计算最优权值和偏置值: } w^* = \sum_{i=1}^4 \alpha_i^* y_i \hat{y}_i = (-1, 1, 1, -1)$$

$$b^* = 0$$

$$\text{其中 } w^* = \frac{1}{8} (-[K(X_1, X_j)]_{4 \times 1} + [K(X_2, X_j)]_{4 \times 1} +$$

$$[K(X_3, X_j)]_{4 \times 1} - [K(X_4, X_j)]_{4 \times 1})$$



- 支持向量机设计举例 — XOR问题（续）

- 方法二（续）：

(3) 支持向量机的最优分界面为：
$$\sum_{i=1}^4 \alpha_i^* y_i K(x_i, x) + b^* = 0$$

即：

$$\frac{1}{8}(-K(X_1, X) + K(X_2, X) + K(X_3, X) - K(X_4, X)) = 0$$

将 $b^* = 0$ 和下式($i = 1, 2, 3, 4$)代入上式

$$K(X_i, X) = 1 + x_1^2 x_{i1}^2 + 2x_1 x_2 x_{i1} x_{i2} + x_2^2 x_{i2}^2 + 2x_1 x_{i1} + 2x_2 x_{i2}$$

整理后有：

$$\sum_{i=1}^4 \alpha_i^* y_i K(x_i, x) + b^* = -x_1 x_2 = 0$$



从结构、学习、功能和性能等方面，对含单隐层MLP网络、RBF网络和SVM网络进行详细比较，并结合自己的研究领域讨论它们的应用前景。