

第 5 节 核函数和非线性

SVM分类问题



5.1 非线性分类问题分析

分类问题不一定是线性可分的。


要解决一个非线性分类问题，可以设法将其通过非线性变换，转化为高维空间中的线性问题，再在这个变换空间中求最优分类面。

那么，是否有必要显式地找到一个从低维向高维映射的函数呢？事实上，我们只需要在高维空间上定义变换后的内积运算，就可以对样本进行线性分类了。



在映射后得到的高维空间中，如果样本 $\Phi(x_i)$ 与样本 $\Phi(x_j)$ 的距离 $\|\Phi(x_i) - \Phi(x_j)\|$ 的值很小 (即样本点 $\Phi(x_i)$ 和 $\Phi(x_j)$ 很近)，那我们就可以把样本 x_i 和 x_j 分为同一类。

那么，在不知道映射函数 $\Phi(\cdot)$ 表达形式的条件下，能不能计算 $\|\Phi(x_i) - \Phi(x_j)\|$ 呢？核函数可以，它可以解决映射后高维空间样本距离 $\|\Phi(x_i) - \Phi(x_j)\|$ 的计算，同时又不显式地展示出映射函数 $\Phi(\cdot)$ 。



核函数通常可以表示为:

$$K(x_1, x_2) = \langle \Phi(x_1), \Phi(x_2) \rangle$$

从而有:

$$\|\Phi(x_1) - \Phi(x_2)\|^2$$

$$= \langle \Phi(x_1) - \Phi(x_2), \Phi(x_1) - \Phi(x_2) \rangle$$

$$= \langle \Phi(x_1), \Phi(x_1) \rangle - 2 \langle \Phi(x_1), \Phi(x_2) \rangle + \langle \Phi(x_2), \Phi(x_2) \rangle$$

$$= K(x_1, x_1) - 2 K(x_1, x_2) + K(x_2, x_2)$$



例5.1: XOR问题

x (二维) $\rightarrow \phi(x)$ (三维)

$$x = (x^1, x^2)$$

$$\phi(x) = (\phi_1(x), \phi_2(x), \phi_3(x)) = ((x^1)^2, (x^2)^2, \sqrt{2}x^1x^2)$$

线性不
可分

$$(0, 0) \rightarrow (0, 0, 0)$$

$$(1, 1) \rightarrow (1, 1, 2^{1/2})$$

$$(0, 1) \rightarrow (0, 1, 0)$$

$$(1, 0) \rightarrow (1, 0, 0)$$

线性可分



XOR数据点从二维向三维空间映射，如图5.1所示。

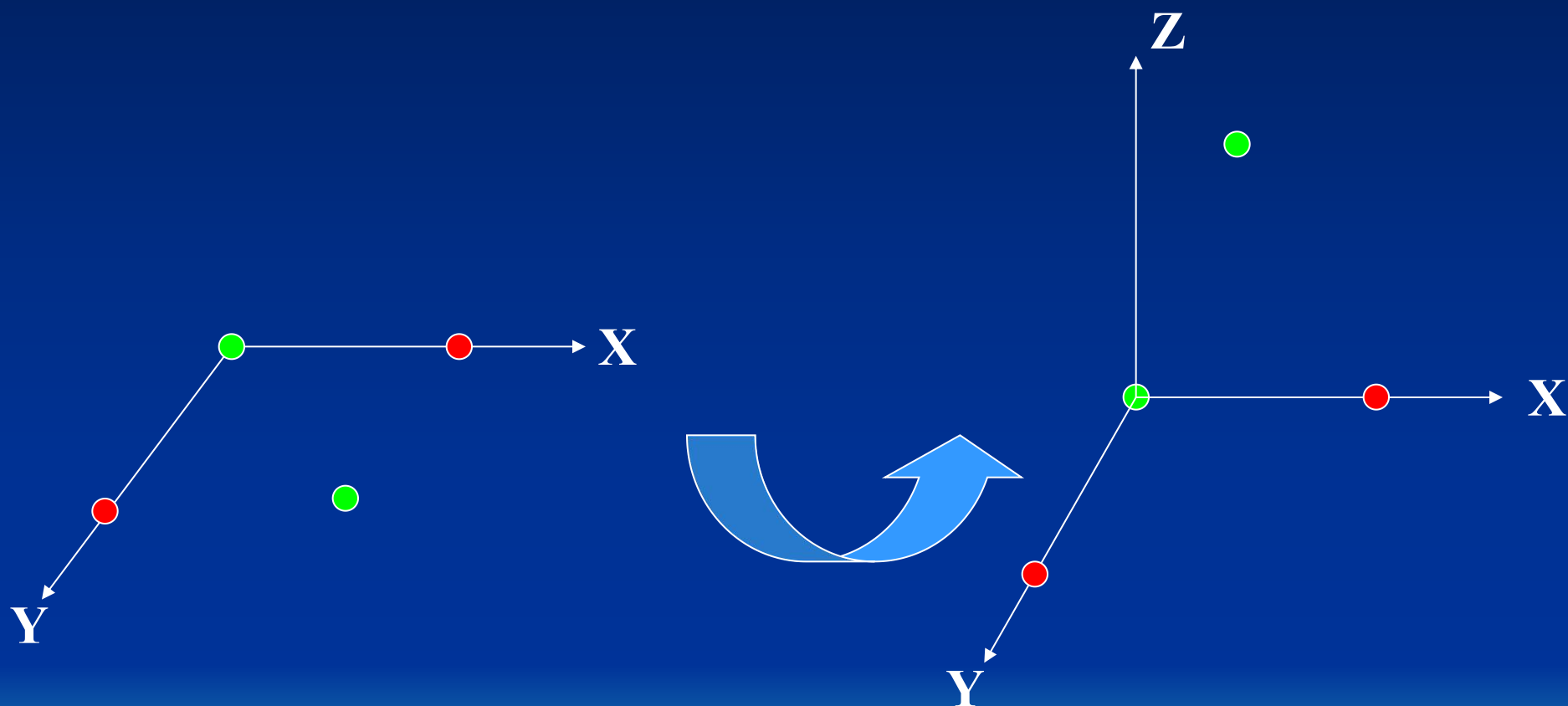


图5.1 XOR数据点从二维向三维映射示意图

5.1.1 核函数定义

设 X 是输入空间， H 是特征空间。如果存在一个从 X 到 H 的映射： $\varphi(x): X \rightarrow H$

使得对所有 $x, z \in X$ ，函数 $K(x, z)$ 满足条件：

$$K(x, z) = \varphi(x) \cdot \varphi(z)$$

则称 $K(x, z)$ 为核函数， $\varphi(x)$ 为映射函数，式中 $\varphi(x) \cdot \varphi(z)$ 为 $\varphi(x)$ 和 $\varphi(z)$ 的内积。



5.1.2 核变换和特征空间

设 X 是 n 维向量，令

$$\Phi(X) = [\varphi_1(X), \varphi_2(X), \dots, \varphi_M(X)]^T$$

表示从 n 维样本空间向 M 维特征空间的非线性变换。 M 维特征空间的维数较高，也称高维特征空间。

在 M 维特征空间中，如果特征数据可以线性分割，其分类面方程则可以表示为：

$$\sum_{j=1}^M W_j \cdot \varphi_j(X) + b = 0$$



其中, b 是分界面方程的偏置量(阈值)。

令 $\varphi_0(X) = 1, W_0 = b$, 则有:

$$\sum_{j=0}^M W_j \cdot \varphi(X) = 0$$

可写成: $W^T \cdot \Phi(X) = 0$

同理, 在高维 M 特征空间中, 权系数最优值为:

$$W = \sum_{i=1}^n \alpha_i \cdot y_i \cdot \Phi(X_i)$$



从而有变换特征空间的分类面为：

$$\sum_{i=1}^n \alpha_i \cdot y_i \cdot \Phi^T(X_i) \cdot \Phi(X) = 0$$

上式中， $\Phi^T(X_i) \cdot \Phi(X)$ 表示第 i 个样本向量 X_i 在特征变换空间中的像 $\Phi^T(X_i)$ 与输入向量 X 在特征变换空间中的像 $\Phi(X)$ 的内积。

从上面可知：在特征变换空间中构造线性分界面时，仅用特征变换空间的内积即可。



很明显，如果能找到一个函数 k ，使得：

$$K(X, X_i) = \Phi^T(X) \cdot \Phi(X_i) = \sum_{j=1}^M \varphi_j(X) \cdot \varphi_j(X_i)$$

则在变换特征空间中建立分类面时，无需考虑变换 φ 的形式。 $K(X, X_i)$ 称为内积核函数，可以用它来代替内积运算 $\Phi^T(X_i) \cdot \Phi(X)$ ，此时对偶问题的目标函数可以改写为：

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



5.1.3 核函数的判定:

函数 $K(X, X_i)$ 满足什么条件才能成为核函数?

Mercer定理:

如果函数 $K(X, X_i)$ 是实数域上的一个有效核函数(也称为Mercer核函数), 那么当且仅当对于训练样例 $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$, 其相应的内积矩阵是对称半正定的。

Mercer定理指出如何确定一个候选核是否是有效的核函数, 但没有指出如何构造映射函数 $\varphi_i(X)$ 。

5.1.4 常用的核函数:

对于核函数 $K(X, X^T)$, 只要求它满足Mercer定理即可。因此, 核函数可以有多种。常用的核函数有4种:

(1) 多项式核函数

$$K(X, X_i) = [(X \cdot X_i) + 1]^q$$

(2) Gauss核函数

$$K(X, X_i) = \exp\left(-\frac{|X - X_i|^2}{\sigma^2}\right)$$

(3) Sigmoid核函数

$$K(X, X_i) = \tanh[v(X \cdot X_i) + c]$$

(4) 线性核函数

$$K(X, X_i) = X \cdot X_i$$


高斯核为什么可以映射到无限维：

高斯核函数：

$$K(X, X_i) = \exp\left(-\frac{|X - X_i|^2}{\sigma^2}\right)$$

指数函数的泰勒展开级数：

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

则可得到：



$$\begin{aligned} K(X, X_i) &= \exp \left[-\frac{1}{\sigma^2} \langle X - X_i, X - X_i \rangle \right] \\ &= \exp \left[-\frac{1}{\sigma^2} (\|X\|^2 + \|X_i\|^2 - 2X^T X_i) \right] \\ &= \exp \left(-\frac{\|X\|^2}{\sigma^2} \right) * \exp \left(-\frac{\|X_i\|^2}{\sigma^2} \right) * \exp \left(\frac{2X^T X_i}{\sigma^2} \right) \end{aligned}$$



其中:

$$\begin{aligned} \exp\left(-\frac{\|X\|^2}{\sigma^2}\right) &= \sum_{n=0}^{\infty} \frac{\left(-\frac{\|X\|^2}{\sigma^2}\right)^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{-\frac{1}{\sigma^2} (\|X_1\|^2 + \|X_2\|^2 + \dots + \|X_k\|^2)^n}{n!} \end{aligned}$$

$$X = (X_1, X_2, \dots, X_k)^T$$

也就是说, $\exp\left(-\frac{\|X\|^2}{\sigma^2}\right)$ 含有了无穷多项的多项式, 对应的映射将 k 维空间映射成了无限维空间。

5.1.5 如何选择正确的核函数：

核函数的选择包括两部分工作：

- 1) 核函数类型的选择；
- 2) 确定核函数类型后相关参数的选择。

目前，没有具体的理论或方法来指导核函数的选取，但在在选取核函数解决实际问题时，有些通常采用的方法。



- a) 利用专家的先验知识预先选定核函数;
- b) 分别试用不同的核函数, 归纳误差最小的核函数就是最好的核函数。

例如, 对傅立叶核、RBF核, 结合信号处理问题中的函数回归问题, 通过仿真实验, 对比分析了在相同数据条件下, 采用傅立叶核的SVM要比采用RBF核的SVM误差小很多。

- c) 采用由Smits等人提出的混合核函数方法, 将不同的核函数结合起来后会有更好的特性, 这是混合核函数方法的基本思想。



5.2 非线性问题的支持向量机

支持向量机对非线性数据样本，采用核方法，即核函数映射等一系列非线性数据处理。

在本质上讲，核方法实现了样本空间、特征空间和类别空间之间的非线性变换。

设 X 、 X_i 为样本空间的样本，考虑对样本空间的非线性分类。样本空间到特征空间的映射函数为 Φ ，则核方法的基础就是实现向量的内积变换，如下式所示：

$$(X, X_i) \rightarrow K(X, X_i) = \Phi(X) \cdot \Phi(X_i)$$



非线性变换 $\Phi(\cdot)$ 相当复杂，但核函数 $K(\cdot, \cdot)$ 则简单得多，在实际计算时，只用 $K(\cdot, \cdot)$ ，而不需要用 $\Phi(\cdot)$ 。

在线性分类中，有几个式子十分重要，分别是：

(1) Lagrange乘子寻优函数：

$$A(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j X_i X_j$$

(2) 分类面函数：

$$d(X) = W^* \cdot X + b^* = \sum_{i=1}^n \alpha_i^* y_i X_i \cdot X + b^*$$


如果用核函数 $K(\cdot, \cdot)$ 取代上面两个公式中的 X 内积，很明显就相当于进行了从样本空间到特征空间的非线性变换，因为这种取代实现了核函数分解成内积的变换。

这时有非线性问题寻优和分类函数：

$$A(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \cdot K(X_i, X_j)$$

$$d(X) = \sum_{i=1}^n \alpha_i^* y_i \cdot K(X_i, X) + b^*$$



很明显，这也是一种线性分类，它就是非线性支持向量机，可以用结构图表示，如图5.2所示。它由一个三层的网络组成：输入层、中间层和输出层。

输入层是输入样本向量，中间层是基于 S 个支持向量 $X_i, i = 1, 2, \dots, S$ 的非线性变换，即核函数 $K(X, X_i)$ 。

输出层是决策规则 $d(X)$ 。



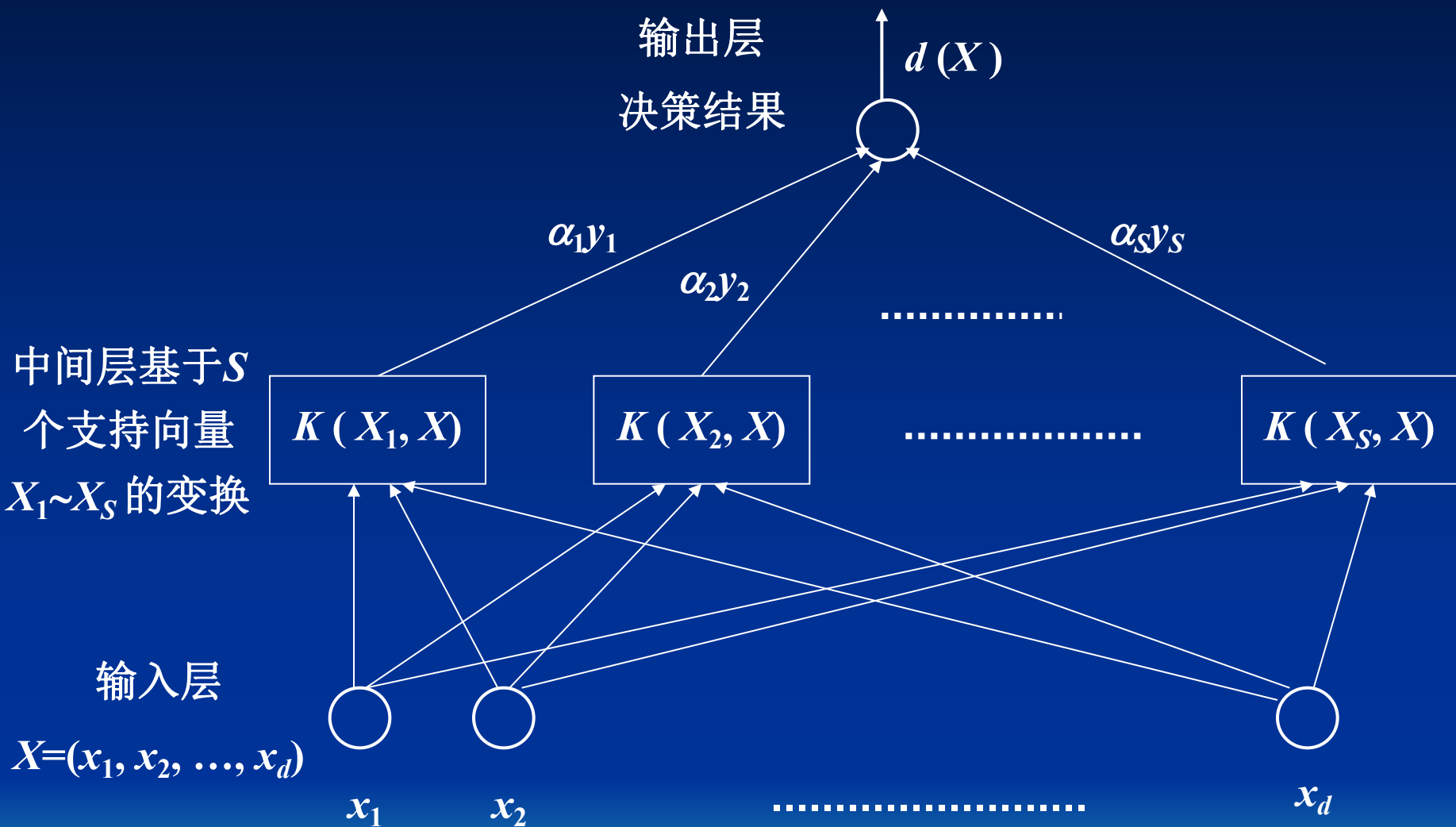


图5.2 非线性支持向量机网络结构示图

5.3 非线性支持向量机实例

在实际应用中，用支持向量机求分类问题步骤如下：

1. 选择适当核函数 $K(\cdot, \cdot)$ ；
2. 求解寻优方程 $A(\alpha)$ ，得出Lagrange乘子 α_i ，
 $i=1, 2, \dots$ 和 $\alpha_i \neq 0$ 对应的支持向量；
3. 给出最优分类面方程：

$$d(X) = \sum_{i=1}^n \alpha_i^* y_i \cdot K(X_i, X) + b^*$$

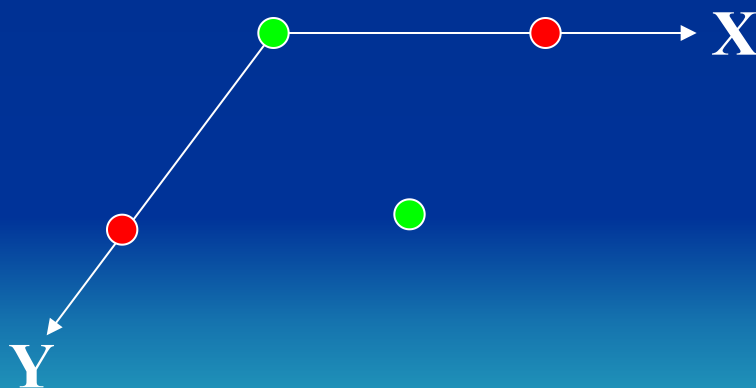
例5.2

XOR (异或) 分类问题:

$$X_1 = \{ (1, 0), (0, 1) \}$$

$$X_{-1} = \{ (1, 1), (0, 0) \}$$

求它们的最优分类面。



解：

1. 选择核函数：选择以下的非线性映射
将二维数据映射到三维特征空间：

$$\varphi(x_i) = (z_{i1}, z_{i2}, z_{i3}) = (x_{i1}, x_{i2}, x_{i1} \cdot x_{i2})$$

$x_{i1}(z_{i1})$	$x_{i2}(z_{i2})$	$x_{i1} \cdot x_{i2}(z_{i3})$	Label
1	0	0	1
0	1	0	1
1	1	1	-1
0	0	0	-1

2. 解寻优方程

首先求出三维空间中的核函数:

$$K(z_i, z) = \langle z_i, z \rangle$$

$$= \begin{vmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{vmatrix}$$

接着，对Lagrange寻优函数求导以获得最优值 α^* ：

$$A(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \cdot K(z_i, z_j)$$

得到相应的优化方程：

$$\begin{cases} \alpha_1 - \alpha_3 = 1 \\ \alpha_2 - \alpha_3 = 1 \\ \alpha_1 + \alpha_2 - 3\alpha_3 = -1 \end{cases}$$

再加上约束： $\sum_{i=1}^4 \alpha_i y_i = 0$ 即： $\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 = 0$

联立上述四个方程组成方程组，求解得到拉格朗日系数 α 的最优值为：

$$\begin{cases} \alpha_1 = 4 \\ \alpha_2 = 4 \\ \alpha_3 = 3 \\ \alpha_4 = 5 \end{cases}$$

说明此例中的四个样本都是支持向量。



3. 给出最优分类面方程:

① 根据最优权向量 W^* 的计算公式:


$$W^* = \sum_{i=1}^n \alpha_i y_i \varphi(z_i)$$

计算得出最优的权值向量 $W^* = [1 \ 1 \ -3]^T$

② 根据最优阈值 b^* 的计算公式:

$$b^* = -\frac{1}{2} W^* \cdot (z_i + z_j)$$

其中, z_i 和 z_j 是来自两个不同类的支持向量。计算得出最优阈值为:

$$b^* = -\frac{1}{2}$$


最优分界面方程为：

$$d(z) = W^* \cdot z + b^* = [1 \quad 1 \quad -3] \cdot \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} - \frac{1}{2}$$

$$= z_1 + z_2 - 3z_3 - \frac{1}{2}$$

映射回二维空间，可得：

$$x_1 + x_2 - 3x_1 \cdot x_2 - \frac{1}{2} = 0$$

如图5.3所示。

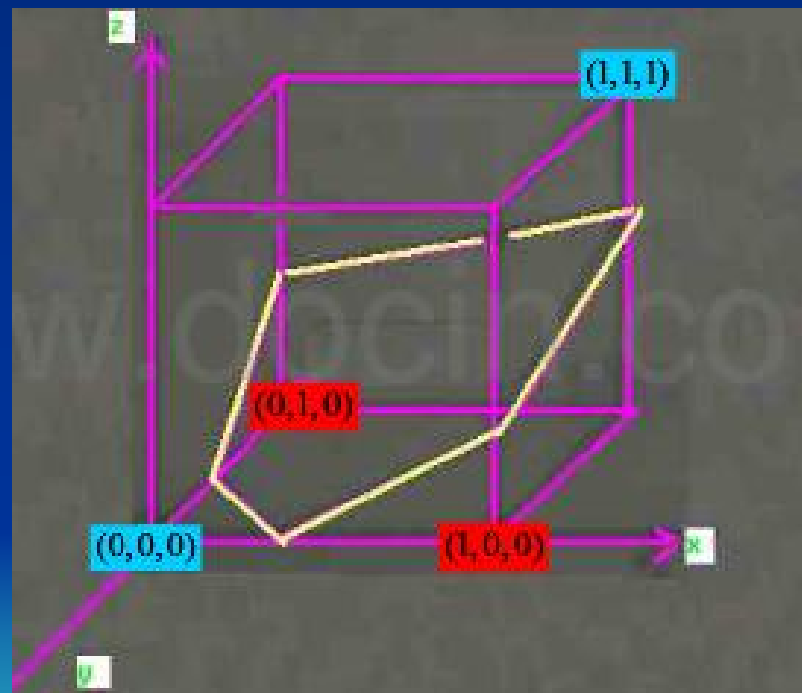


图5.3 XOR问题SVM解答示意图

例5.3 XOR问题的另一个核函数

输入向量 x	期望的类别 y
$(+1, -1)$	$+1$
$(-1, +1)$	$+1$
$(-1, -1)$	-1
$(+1, +1)$	-1

1、核函数取多项式核函数

$$K(x_i, x_j) = (1 + x_i^T x_j)^2$$

$$x_i = [x_{i1}, x_{i2}]^T$$

$$K(x_i, x_j) = 1 + x_{i1}^2 x_{j1}^2 + 2x_{i1} x_{i2} x_{j1} x_{j2} + x_{i2}^2 x_{j2}^2 + 2x_{i1} x_{j1} + 2x_{i2} x_{j2}$$



多项式核的基函数:

多项式核函数展开后, 可以得到基函数, 也就是构成核函数的内积基。它把二维输入向量映射到高维特征空间。

$$K(\mathbf{x}_i, \mathbf{x}_j) = 1 + \mathbf{x}_{i1}^2 \mathbf{x}_{j1}^2 + 2\mathbf{x}_{i1} \mathbf{x}_{i2} \mathbf{x}_{j1} \mathbf{x}_{j2} + \mathbf{x}_{i2}^2 \mathbf{x}_{j2}^2 + 2\mathbf{x}_{i1} \mathbf{x}_{j1} + 2\mathbf{x}_{i2} \mathbf{x}_{j2}$$
$$= \begin{pmatrix} 1 & \mathbf{x}_{i1}^2 & \sqrt{2}\mathbf{x}_{i1}\mathbf{x}_{i2} & \mathbf{x}_{i2}^2 & \sqrt{2}\mathbf{x}_{i1} & \sqrt{2}\mathbf{x}_{i2} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \mathbf{x}_{j1}^2 \\ \sqrt{2}\mathbf{x}_{j1}\mathbf{x}_{j2} \\ \mathbf{x}_{j2}^2 \\ \sqrt{2}\mathbf{x}_{j1} \\ \sqrt{2}\mathbf{x}_{j2} \end{pmatrix} = \boldsymbol{\varphi}(\mathbf{x}_i) \cdot \boldsymbol{\varphi}(\mathbf{x}_j)^T$$

所以，多项式核函数的内积基为：

$$\varphi(x_i) = [1, x_{i1}^2, \sqrt{2}x_{i1}x_{i2}, x_{i2}^2, \sqrt{2}x_{i1}, \sqrt{2}x_{i2}]^T$$

根据此特征基，每个二维空间的样本都可以映射成为六维空间中的一个向量。例如， $x_1 = (+1, -1)$ 这个向量可以映射为：

$$\varphi(x_1) = [1, 1, -\sqrt{2}, 1, \sqrt{2}, -\sqrt{2}]^T$$

$\varphi(x_2)$, $\varphi(x_3)$, $\varphi(x_4)$ 的计算类似可以完成。



2、求解寻优方程 $A(\alpha)$

首先计算高维 (本例是六维) 空间中的核函数

$K(x_i, x_j)$ ，得到：

$$K = \begin{vmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{vmatrix}$$

接着，对Lagrange寻优函数求导以获得最优值 α^* ：

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

得到相应的优化方程：

$$\begin{cases} 9\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 = 1 \\ \alpha_1 + 9\alpha_2 - \alpha_3 - \alpha_4 = 1 \\ \alpha_1 + \alpha_2 - 9\alpha_3 - \alpha_4 = -1 \\ \alpha_1 + \alpha_2 - \alpha_3 - 9\alpha_4 = -1 \end{cases}$$

再加上约束： $\sum_{i=1}^4 \alpha_i y_i = 0$ 即： $\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 = 0$

联立上述五个方程组成方程组，求解得到拉格朗日系数 α 的最优值为：

$$\begin{cases} \alpha_1 = \frac{1}{8} \\ \alpha_2 = \frac{1}{8} \\ \alpha_3 = \frac{1}{8} \\ \alpha_4 = \frac{1}{8} \end{cases}$$

说明此例中的四个样本都是支持向量。



3. 给出最优分类面方程:

① 根据最优权向量 W^* 的计算公式:

$$W^* = \sum_{i=1}^n \alpha_i y_i \varphi(x_i)$$

计算得出最优的权值向量: $W^* = [0 \ 0 \ -1/\sqrt{2} \ 0 \ 0 \ 0]^T$

② 根据最优阈值 b^* 的计算公式:

$$b^* = -\frac{1}{2} W^* \cdot [\varphi(x_i) + \varphi(x_j)]$$

其中, $\varphi(x_i)$ 和 $\varphi(x_j)$ 是高维空间中来自两个不同类的支持向量。计算得出最优阈值 $b^* = 0$

最优分界面方程为:

$$d[\varphi(x)] = W^* \cdot \varphi(x) + b^* = \begin{bmatrix} 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \varphi(x_1) \\ \varphi(x_2) \\ \varphi(x_3) \\ \varphi(x_4) \\ \varphi(x_5) \\ \varphi(x_6) \end{bmatrix} + 0$$
$$= -\frac{1}{\sqrt{2}} \varphi(x_3)$$

映射回二维空间, 可得二维分界面为:

$$-x_1 \cdot x_2 = 0$$


例5.4 存在两类样本集:

$$SX_A = \{ (1, 1), (1, 4), (4, 1), (4, 4) \}$$

$$SX_B = \{ (2, 2), (2, 3), (3, 2), (3, 3) \}$$

求它们的最优分类面。



解：1. 选择核函数

选二阶多项式核函数：

$$K(X, X_i) = [(X \cdot X_i) + 1]^2$$

2. 解寻优方程

$$A(\alpha) = \sum_{i=1}^8 \alpha_i - \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \alpha_i \alpha_j y_i y_j K(X_i \cdot X_j)$$

从而求出最优的 α 值以及对应的支持向量 X_i^* 及分界面的最优权重 W^* 、阈值 b^* 。

3. 给出最优分类面方程:

$$d(X) = \sum_{i=1}^8 \alpha_i^* y_i [(X_i \cdot X) + 1]^2 + b^*$$

计算得出分类面方程如下, 结果如图5.4所示。

$$x_1^2 + x_2^2 - 5x_1 - 5x_2 + 10 = 0$$

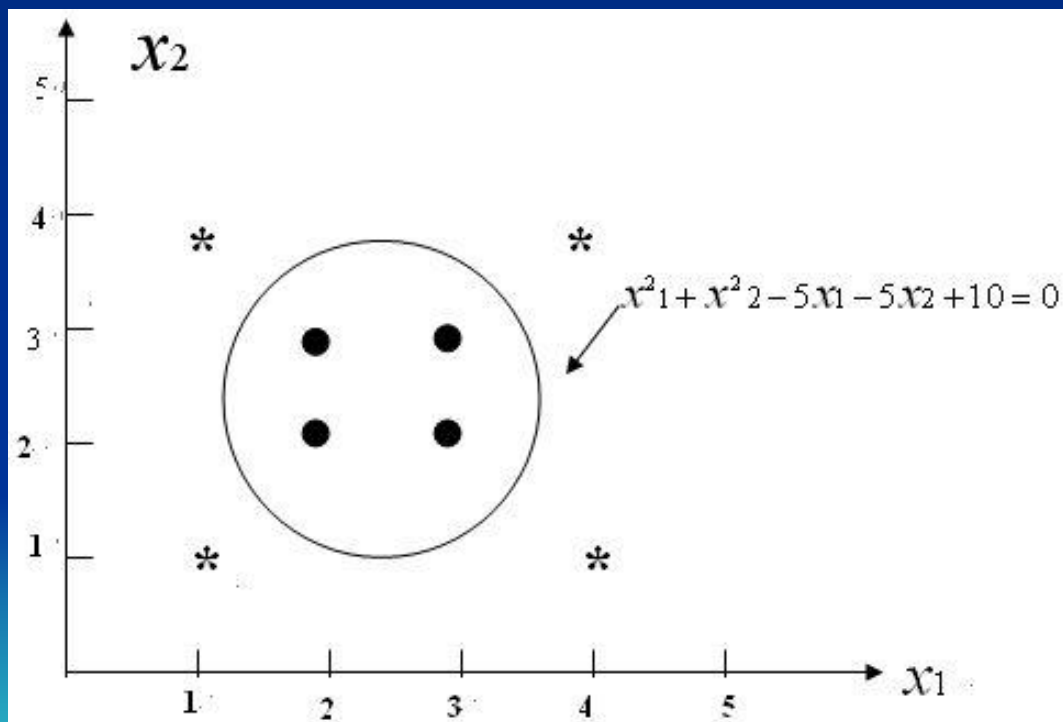


图5.4 例5.4解图

习题：

- 试采用核映射SVM方法对“癌症患者生存期”数据集进行分类。给出具体代码，以及训练和测试的准确率结果。

[UCI数据集\haberman.names](#)

[UCI数据集\haberman.data](#)

