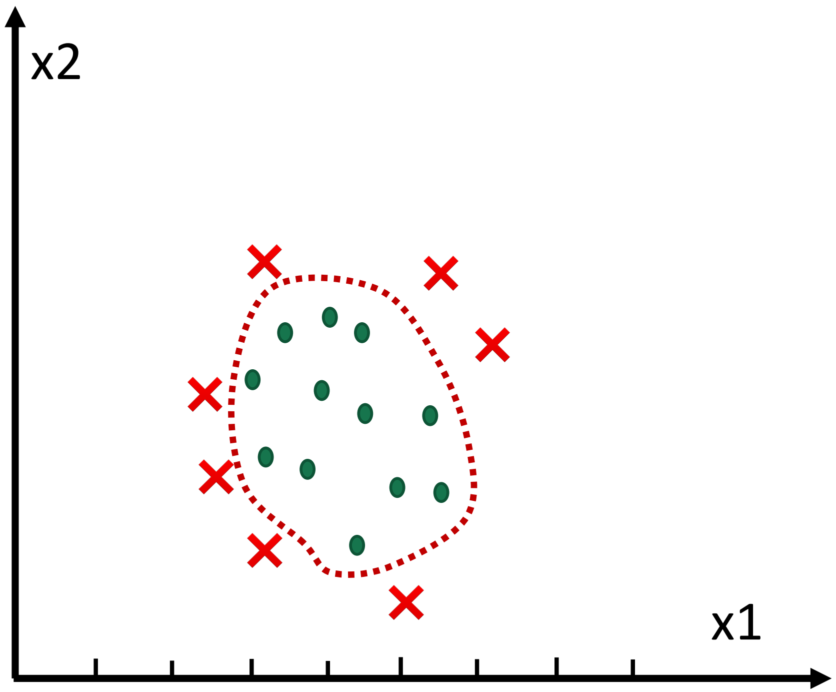
第5章 支持向量机*

1. 线性可分支持向量机与硬间隔最大化
2. 线性支持向量机与软间隔最大化（回顾与加强）
3. 非线性支持向量机与核函数
4. 序列最小最优化算法

*参阅《机器学习方法》第7章

线性不可分

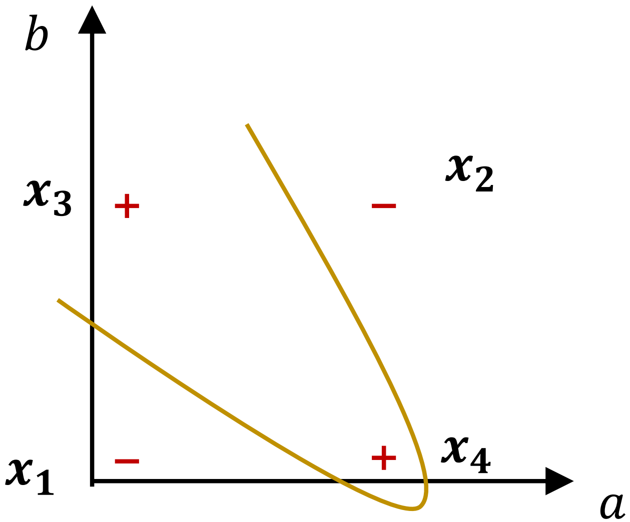
-Q:若不存在一个能正确划分两类样本的超平面，怎么办?



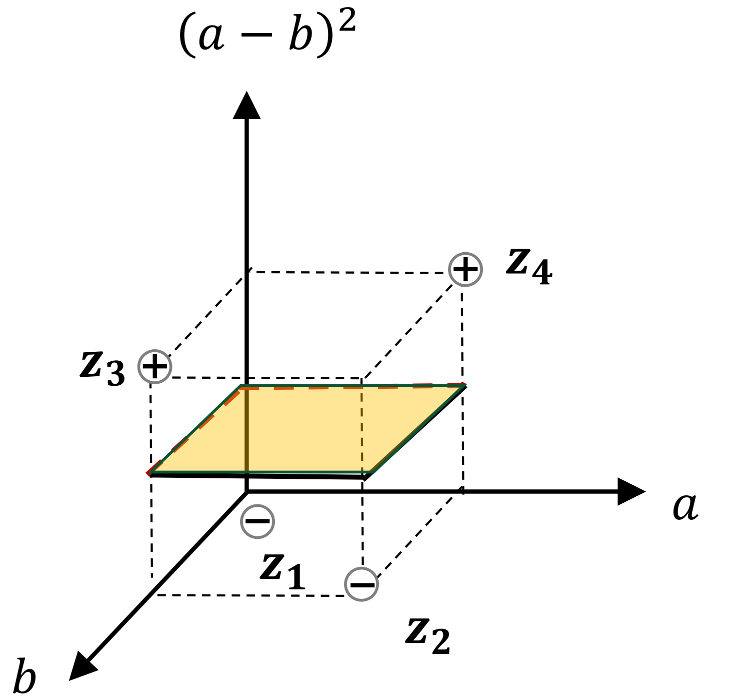
-A:将样本从原始空间映射到一个更高维的特征空间，使得样本在这个特征空间内线性可分.

$$\boldsymbol{x} \mapsto \phi(\boldsymbol{x})$$

例子1



异或问题



$$\boldsymbol{x} = [a, b]^T \xrightarrow{\boldsymbol{x} \mapsto \phi(\boldsymbol{x})} \boldsymbol{z} = \varphi(\boldsymbol{x})$$

$$\boldsymbol{z} = [a, b, (a-b)^2]^T$$

C1 C2

$\boldsymbol{x}_1 = [0,0]^T$ $\boldsymbol{x}_3 = [0,1]^T$

$\boldsymbol{x}_2 = [1,1]^T$ $\boldsymbol{x}_4 = [1,0]^T$

$\boldsymbol{z}_1 = [0,0,0]^T$ $\boldsymbol{z}_3 = [0,1,1]^T$

$\boldsymbol{z}_2 = [1,1,0]^T$ $\boldsymbol{z}_4 = [1,0,1]^T$

例子2

找不到一个超平面（二维：直线）将其分割开来

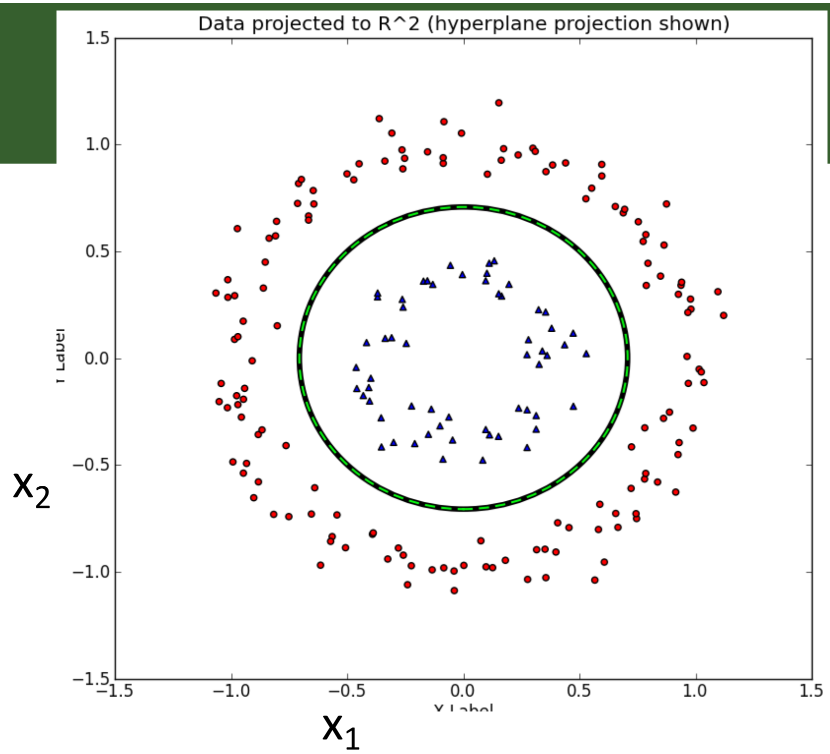
用圆或者椭圆将数据分类

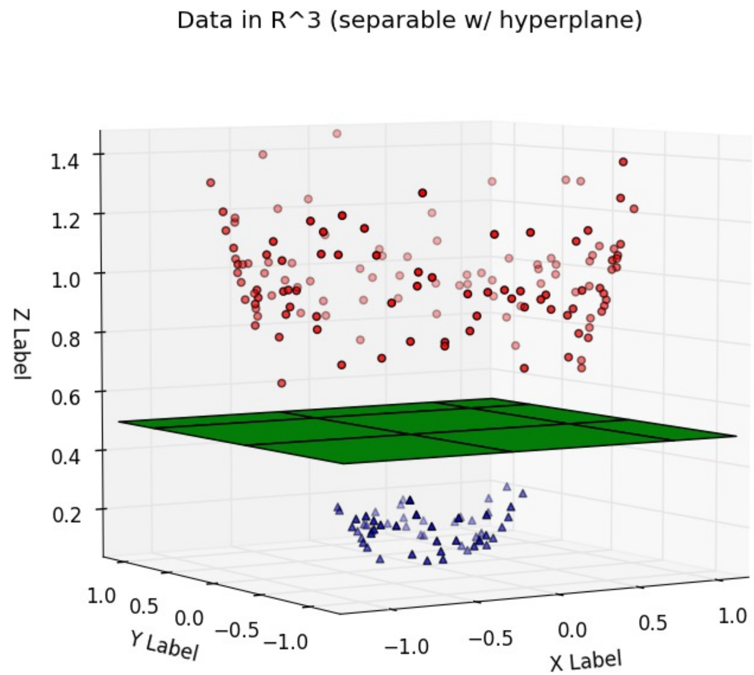
$$a_1x_1^2 + a_2x_1x_2 + a_3x_2^2 = 0$$

$$\boldsymbol{z} = \varphi(\boldsymbol{x})$$

$$z_1 = x_1 \quad z_2 = x_2 \quad z_3 = x_1^2 + x_2^2$$

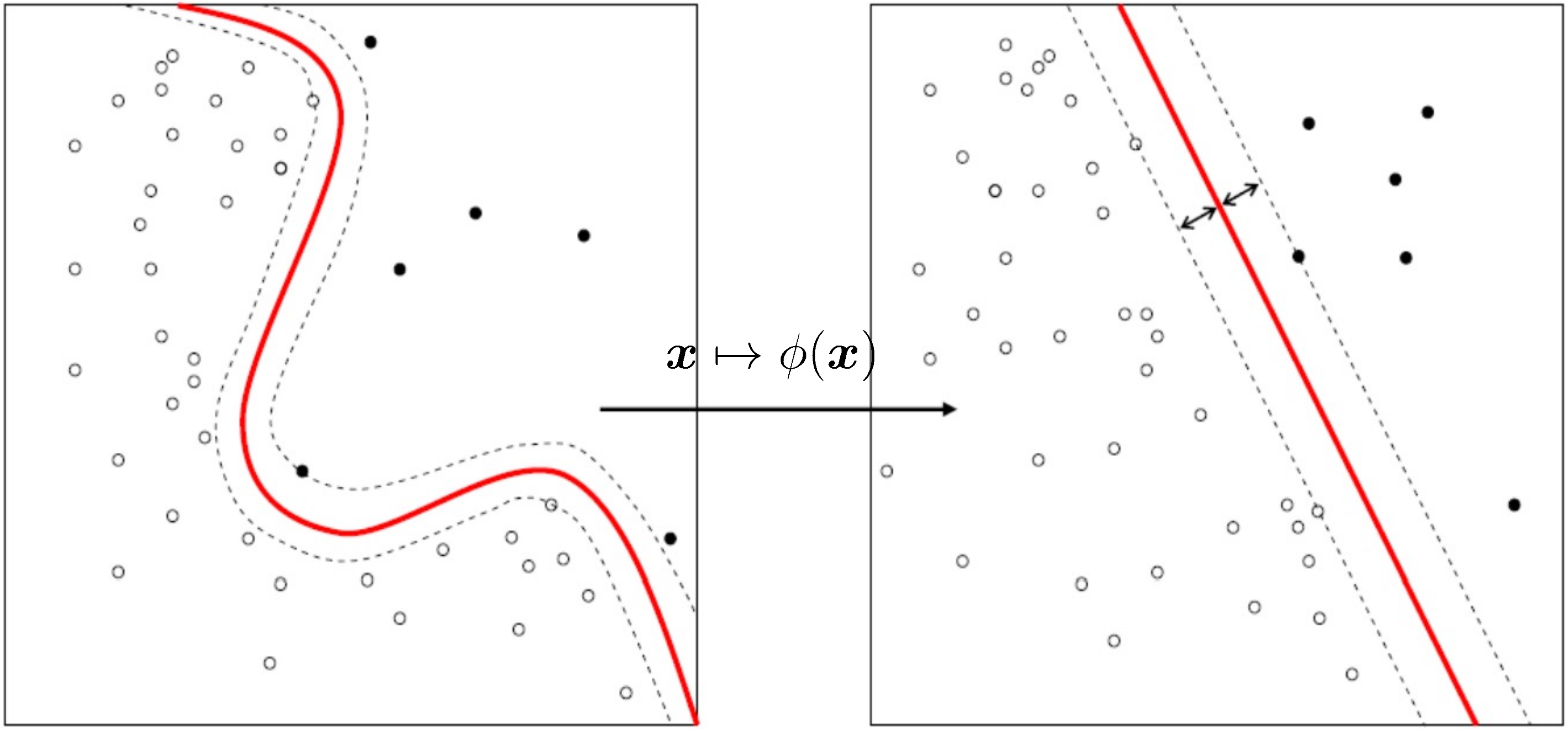
$$(x_1, x_2) \xrightarrow[\boldsymbol{z} = \varphi(\boldsymbol{x})]{} (z_1, z_2, z_3)$$





https://www.eric-kim.net/eric-kim-net/posts/1/kernel_trick.html

Kernel SVM

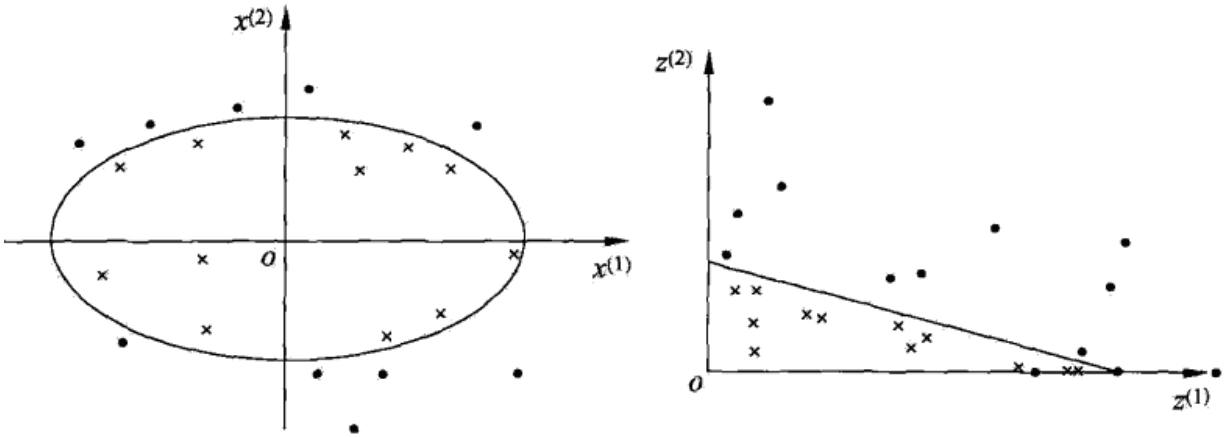


核技巧

- 非线性分类问题

用线性分类方法求解非线性分类问题分为两步：

- 首先使用一个变换将原空间的数据映射到新空间；
- 然后在新空间里用线性分类学习方法从训练数据中学习分类模型。



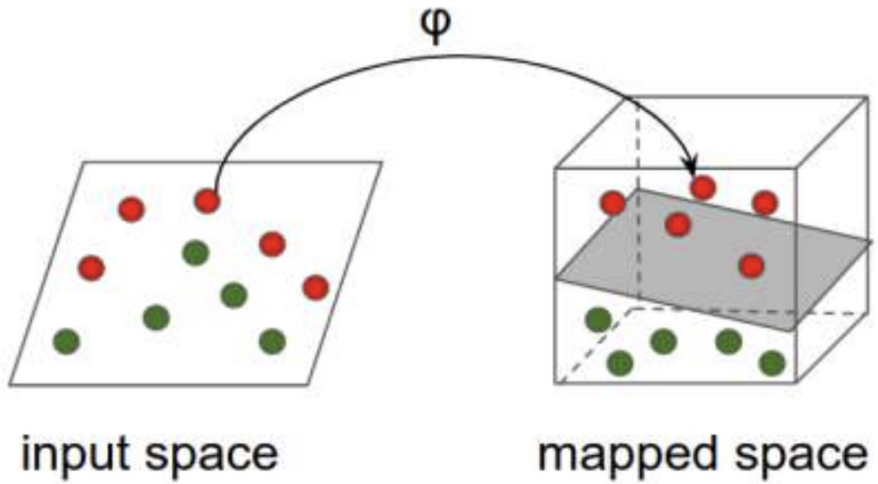
核技巧就属于这样的方法

核技巧

- **非线性分类问题**

核技巧应用到支持向量机，其基本想法：

- 通过一个非线性变换将输入空间（欧氏空间$\mathbb{R}^d$或离散集合）对应于一个特征空间（希尔伯特空间$\mathcal{H}$），使得在输入空间$\mathbb{R}^d$中的超曲面模型对应于特征空间$\mathcal{H}$中的超平面模型（支持向量机）；
- 分类问题的学习任务通过在特征空间中求解线性支持向量机就可以完成。



核技巧——回顾

◻ 线性支持向量机学习的对偶算法

第三步：回代，得到对偶问题

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boxed{\boldsymbol{x}_i^T\boldsymbol{x}_j} - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \sum_{i=1}^{m}\alpha_i y_i = 0,$$

$$0 \leq \alpha_i \leq C, i = 1, \cdots, m$$

第四步：解出$\boldsymbol{\alpha}^*$，根据KKT条件求出$(\boldsymbol{\omega}^*, b^*)$

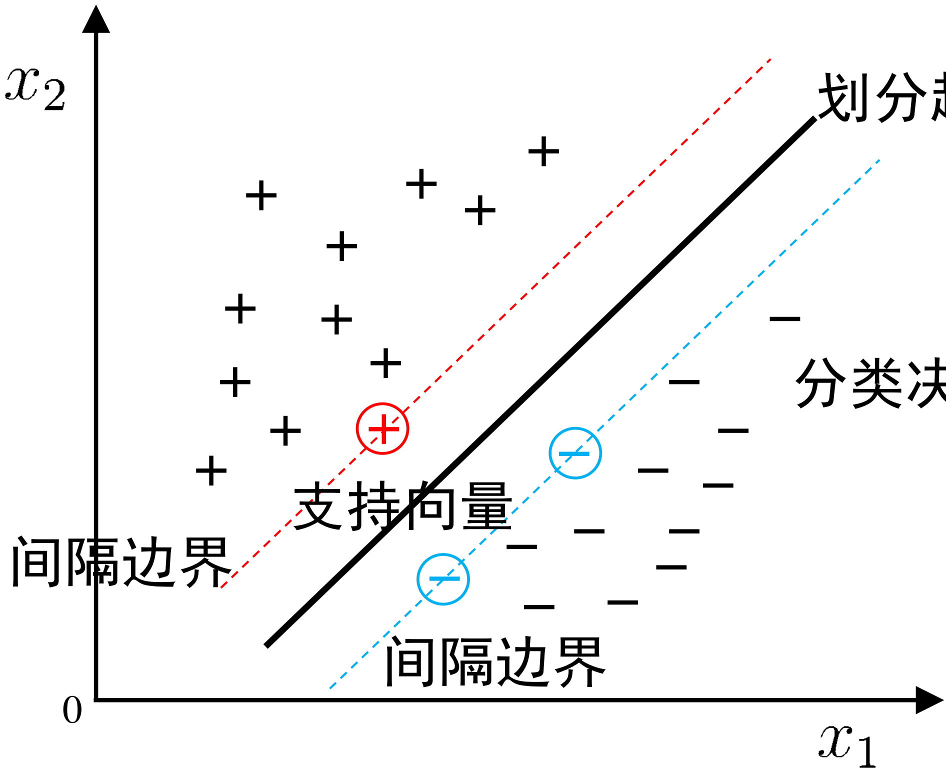
$$\boldsymbol{\omega}^* = \sum_{i=1}^{m}\alpha_i^* y_i \boldsymbol{x}_i, b^* = y_j - \sum_{i=1}^{m}\alpha_i^* y_i \boxed{\boldsymbol{x}_i^T\boldsymbol{x}_j}$$

j为$\boldsymbol{\alpha}^*$中的一个正分量 $0 < \alpha_j^* < C$所对应的下标

核技巧——回顾

□ 线性支持向量机学习的对偶算法

第五步：最终模型$f(\boldsymbol{x}) = \mathrm{sign}\left(\boldsymbol{\omega}^{*\mathrm{T}}\boldsymbol{x} + b^*\right)$



划分超平面：$\sum_{i=1}^{m} \alpha_i^* y_i \boxed{\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}} + b^* = 0$

分类决策函数：$f(\boldsymbol{x}) = \mathrm{sign}\left(\sum_{i=1}^{m} \alpha_i^* y_i \boxed{\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}} + b^*\right)$

现象：分类决策函数只依赖于输入$\boldsymbol{x}$和训练样本输入的内积。

核技巧

☐ 核函数

设$\mathcal{X}$是输入空间（欧式空间$\mathbb{R}^d$的子集或离散集合），又设$\mathcal{H}$为特征空间（希尔伯特空间），如果存在一个从$\mathcal{X}$到$\mathcal{H}$的映射

$$\phi(\boldsymbol{x}): \mathcal{X} \rightarrow \mathcal{H}$$

使得对所有$\boldsymbol{x}_i, \boldsymbol{x}_j \subset \mathcal{X}$，函数$K(\boldsymbol{x}_i, \boldsymbol{x}_j)$满足条件

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i) \cdot \phi(\boldsymbol{x}_j) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle = \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)$$

则称$K(\boldsymbol{x}_i, \boldsymbol{x}_j)$为核函数，$\phi(\boldsymbol{x})$为映射函数，$\langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle$为$\phi(\boldsymbol{x}_i)$和$\phi(\boldsymbol{x}_j)$的内积。

“映射”后高维空间的内积，可通过“映射”前原始空间的核函数来进行计算，并不需要“真正”地映射到高维空间。

核技巧

- 核技巧的基本想法

在学习与预测中只定义核函数$K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)$，而不显式地定义映射函数$\phi(\boldsymbol{x})$。

- ϕ是输入空间$\mathbb{R}^d$到特征空间$\mathcal{H}$的映射，特征空间$\mathcal{H}$一般是高维，甚至是无穷维的；
- 直接计算$K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)$比较容易，而通过$\phi(\boldsymbol{x}_i)$和$\phi\left(\boldsymbol{x}_j\right)$计算$K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)$并不容易。

核技巧

核技巧在支持向量机中的应用

对偶问题：

$$\min_{\boldsymbol{\alpha}} \ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \sum_{i=1}^{m}\alpha_i y_i = 0,$$

$$0 \le \alpha_i \le C, i = 1, \cdots, m$$

$\Rightarrow$

$$\min_{\boldsymbol{\alpha}} \ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \sum_{i=1}^{m}\alpha_i y_i = 0,$$

$$0 \le \alpha_i \le C, i = 1, \cdots, m$$

决策函数：

$$f(\boldsymbol{x}) = \text{sign}\left(\sum_{i=1}^{m}\alpha_i^* y_i \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x} + b^*\right)$$

$\Rightarrow$

$$f(\boldsymbol{x}) = \text{sign}\left(\sum_{i=1}^{m}\alpha_i^* y_i K\left(\boldsymbol{x}_i, \boldsymbol{x}\right) + b^*\right)$$

核技巧

核技巧在支持向量机中的应用

对偶问题：

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)-\sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \sum_{i=1}^{m}\alpha_i y_i = 0,$$

$$0 \le \alpha_i \le C, i = 1,\cdots,m$$

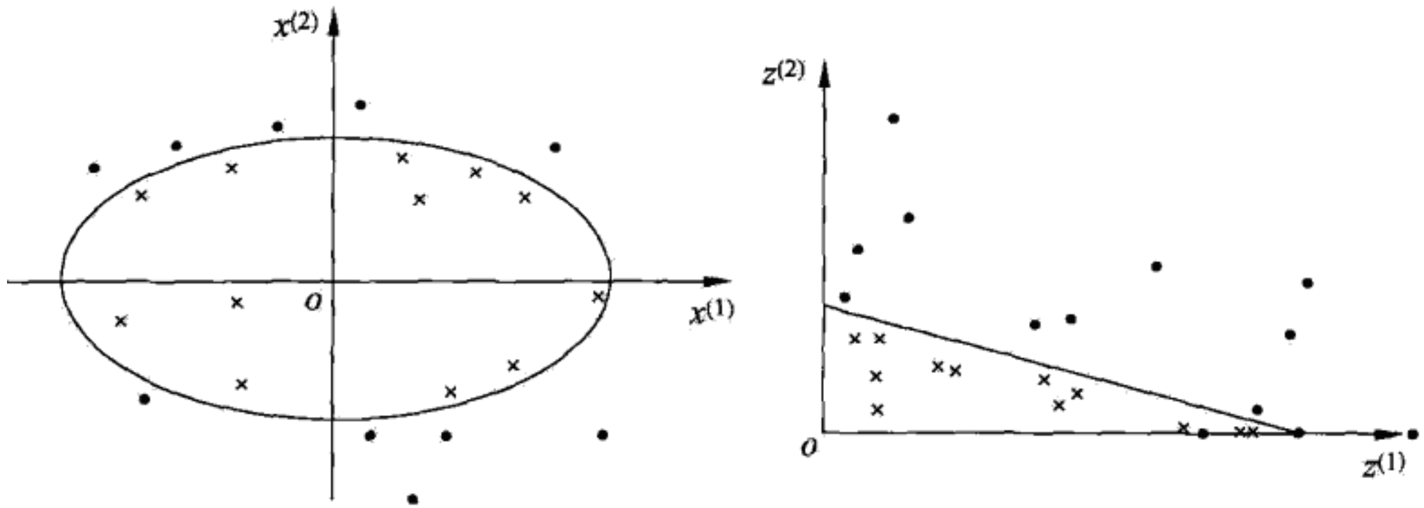
决策函数：

$$f(\boldsymbol{x}) = \operatorname{sign}\left(\sum_{i=1}^{m}\alpha_i^* y_i K(\boldsymbol{x}_i,\boldsymbol{x}) + b^*\right)$$

这等价于经过映射函数ϕ将原来的输入空间变换到一个新的特征空间，将输入空间中的内积$\boldsymbol{x}_i^T\boldsymbol{x}_j$变换为特征空间中的内积$\phi(\boldsymbol{x}_i)\cdot\phi\left(\boldsymbol{x}_j\right)$，在新的特征空间里从训练样本中学习线性支持向量机。当映射函数是非线性函数时，学习得到的含有核函数的支持向量机是非线性分类模型。

核技巧

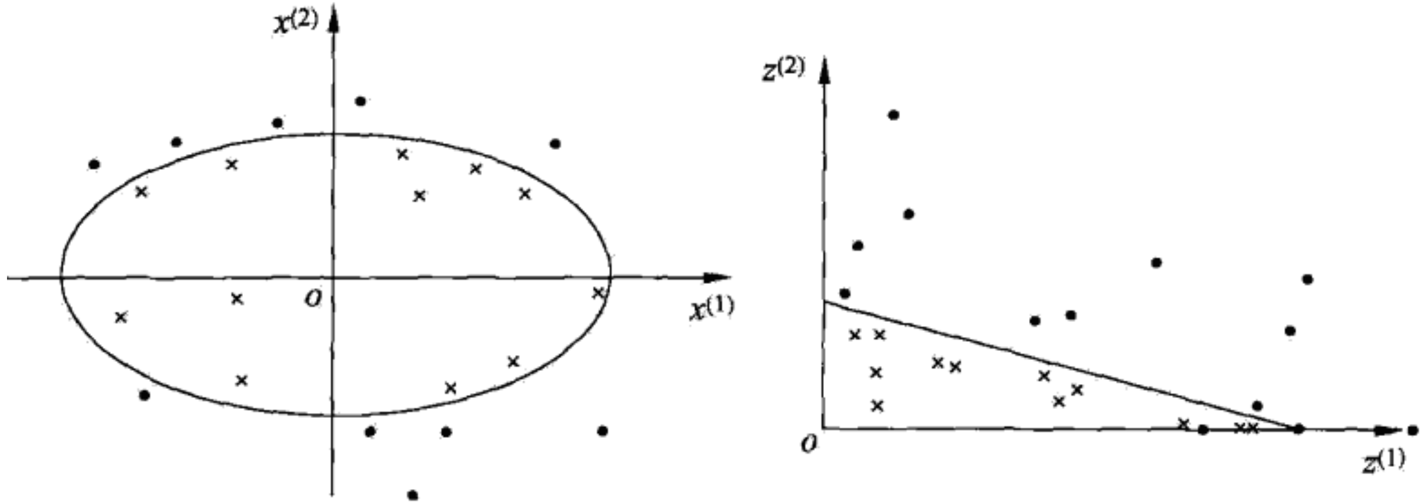
- 核技巧在支持向量机中的应用



这等价于经过映射函数ϕ将原来的输入空间变换到一个新的特征空间，将输入空间中的内积$\boldsymbol{x}_i^T\boldsymbol{x}_j$变换为特征空间中的内积$\phi(\boldsymbol{x}_i)\cdot\phi(\boldsymbol{x}_j)$，在新的特征空间里从训练样本中学习线性支持向量机。当映射函数是非线性函数时，学习得到的含有核函数的支持向量机是非线性分类模型。

核技巧

□ 核技巧在支持向量机中的应用



$$K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\phi(\boldsymbol{x}_i)\cdot\phi(\boldsymbol{x}_j)=\langle\phi(\boldsymbol{x}_i),\phi(\boldsymbol{x}_j)\rangle=\phi(\boldsymbol{x}_i)^T\phi(\boldsymbol{x}_j)$$

在核函数$K(\boldsymbol{x}_i,\boldsymbol{x}_j)$给定的条件下，可以利用解线性问题的方法求解非线性分类问题的支持向量机。学习是隐式地在特征空间进行的，不需要显示地定义特征空间和映射函数。这样的技巧称为核技巧。

核技巧举例1

- 基本想法：不显式地设计核映射，而是设计核函数.

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^\top \phi(\boldsymbol{x}_j)$$

- Kernel Trick

$$\boldsymbol{x} = (a, b)^T \Rightarrow \varphi(\boldsymbol{x}) = \left(a^2, b^2, \sqrt{2ab}\right)^T$$

$$\varphi(\boldsymbol{x}_1)^T \varphi(\boldsymbol{x}_2) = \left(a_1^2, b_1^2, \sqrt{2a_1 b_1}\right) \begin{pmatrix} a_2^2 \\ b_2^2 \\ \sqrt{2a_2 b_2} \end{pmatrix}$$

$$= a_1^2 a_2^2 + b_1^2 b_2^2 + 2a_1 b_1 a_2 b_2$$

$$\kappa(x_1, x_2) = (x_1^T x_2)^2$$

$$= (a_1 a_2 + b_1 b_2)^2 = (x_1^T x_2)^2$$

核技巧举例1：2次多项式kernel

- d-Feature

$$\boldsymbol{x} = (x_1, x_2, \dots, x_d)^T$$

$$\varphi_2(\boldsymbol{x}) = (x_1x_1, x_1x_2, \dots, x_1x_d, x_2x_1, x_2x_2, \dots, x_2x_d, \dots\dots x_dx_d)^T$$

$$\varphi_2(\boldsymbol{x^1})^{\mathrm{T}}\varphi_2(\boldsymbol{x^2}) = \sum_{i=1}^{d}\sum_{j=1}^{d} x_i^1\, x_j^1\, x_i^2\, x_j^2$$

$$= \sum_{i=1}^{d} x_i^1\, x_i^2 \sum_{j=1}^{d} x_j^1\, x_j^2$$

$$\kappa(\boldsymbol{x^1}, \boldsymbol{x^2}) = (\ \boldsymbol{x^1}^{\mathrm{T}}\boldsymbol{x^2})^2$$

核函数

- 基本想法：不显式地设计核映射，而是设计核函数.

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^\top \phi(\boldsymbol{x}_j)$$

- 常用核函数：

名称	表达式	参数
线性核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^\top \boldsymbol{x}_j$	
多项式核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^\top \boldsymbol{x}_j)^d$	$d \geq 1$为多项式的次数
高斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\lVert \boldsymbol{x}_i - \boldsymbol{x}_j \rVert^2}{2\delta^2}\right)$	$\delta > 0$为高斯核的带宽(width)
拉普拉斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\lVert \boldsymbol{x}_i - \boldsymbol{x}_j \rVert}{\delta}\right)$	$\delta > 0$
Sigmoid核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(\beta \boldsymbol{x}_i^\top \boldsymbol{x}_j + \theta)$	tanh为双曲正切函数，$\beta > 0$, $\theta < 0$

- Mercer定理(充分非必要)：只要一个对称函数所对应的核矩阵半正定，则它就能作为核函数来使用.

$$K_{ij} = K(x^{(i)}, x^{(j)}) = \phi(x^{(i)})^\top \phi(x^{(j)}) = \phi(x^{(j)})^\top \phi(x^{(i)}) = K(x^{(j)}, x^{(i)}) = K_{ji}$$

非线性支持向量机学习算法

◻ 非线性支持向量机

输入：训练数据集$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_m,y_m)\}$，其中$\boldsymbol{x}_i\in\mathbb{R}^d, y_i\in\{-1,1\}$，$\quad i=1,\cdots,m$

输出： 划分超平面和分类决策函数

（1） 选择适当的核函数$K(\boldsymbol{x}_i,\boldsymbol{x}_j)$和惩罚参数$C>0$，构造并求解约束最优化问题

$$\min_{\boldsymbol{\alpha}}\ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j K(\boldsymbol{x}_i,\boldsymbol{x}_j)-\sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.}\sum_{i=1}^{m}\alpha_i y_i=0,$$

$$0\le\alpha_i\le C, i=1,\cdots,m$$

求得最优解$\boldsymbol{\alpha}^*=(\alpha_1^*,\alpha_2^*,\cdots,\alpha_d^*)^{\mathrm{T}}$。

非线性支持向量机学习算法

非线性支持向量机

输入：训练数据集$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_m,y_m)\}$，其中$\boldsymbol{x}_i\in\mathbb{R}^d,y_i\in\{-1,1\}$， $i=1,\cdots,m$

输出： 划分超平面和分类决策函数

（2）选择$\boldsymbol{\alpha}^*$中的一个正分量$0<\alpha_j^*<C$，计算

$$b^*=y_j-\sum_{i=1}^{m}\alpha_i^*y_iK(\boldsymbol{x}_i,\boldsymbol{x}_j)$$

非线性支持向量机学习算法

非线性支持向量机

输入：训练数据集$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_m,y_m)\}$，其中 $\boldsymbol{x}_i\in\mathbb{R}^d, y_i\in\{-1,1\}$， $i=1,\cdots,m$

输出： 划分超平面和分类决策函数

（3） 求得划分超平面

$$\boldsymbol{\omega}^{*\mathrm{T}}\boldsymbol{x}+b^*=0 \text{即} \sum_{i=1}^{m}\alpha_i^* y_i K(\boldsymbol{x}_i,\boldsymbol{x})+b^*=0$$

和分类决策函数

$$f(\boldsymbol{x})=\operatorname{sign}\left(\sum_{i=1}^{m}\alpha_i^* y_i K(\boldsymbol{x}_i,\boldsymbol{x})+b^*\right)$$

非线性支持向量机

□ 非线性支持向量机

数据：$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_m,y_m)\}$，$\boldsymbol{x}_i\in\mathbb{R}^d,y_i\in\{-1,1\}$

模型：$f(\boldsymbol{x}_i)=\operatorname{sign}\left(\boldsymbol{\omega}^{\mathrm{T}}\phi(\boldsymbol{x}_i)+b\right)$，使得$f(\boldsymbol{x}_i)\simeq y_i$，其中$\boldsymbol{\omega}\in\mathbb{R}^d$

策略：经过映射函数ϕ将原来的输入空间变换到一个新的特征空间，将输入空间中的内积$\boldsymbol{x}_i^T\boldsymbol{x}_j$变换为特征空间中的内积$\phi(\boldsymbol{x}_i)\cdot\phi(\boldsymbol{x}_j)$，在新的特征空间里从训练样本中学习线性支持向量机。

算法：对偶算法

学得模型：分离超平面$\sum_{i=1}^m\alpha_i^*y_iK(\boldsymbol{x}_i,\boldsymbol{x})+b^*=0$和分类决策函数为$f(\boldsymbol{x})=\operatorname{sign}(\sum_{i=1}^m\alpha_i^*y_iK(\boldsymbol{x}_i,\boldsymbol{x})+b^*)$，即非线性支持向量机。

谢谢！