南昌大学
NANCHANG UNIVERSITY

统计机器学习

主讲人：彭振华

数学与计算机学院

2026年

目录

CONTENTS

支 持 向 量 机

- 支持向量机(support vector machines. SVM)

➢ 二类分类模型.它的基本模型是定义在特征空间上的间隔最大的线性分类器;

➢ 支持向量机还包括核技巧，这使它成为实质上的非线性分类器.

➢ 支持向量机的学习策略就是间隔最大化，可形式化为一个求解凸二次规划(convex quadratic programming)的问题，也等价于正则化的合页损失函数的最小化问题.支持向量机的学习算法是求解凸二次规划的最优化算法.

➢ **线性可分支持向量机(linear support vector machine in linearly separable case).——硬间隔**

➢ **线性支持向量机(linear support vector machine)——软间隔**

➢ **非线性支持向量机(non-linear support vector machine)——核技巧**

支 持 向 量 机

- 二分类问题：
- 输入空间：欧式空间或离散集合
- 特征空间：欧式空间或希尔伯特空间
- 线性可分支持向量机、线性支持向量机：假设这两个空间的元素一一对应，并将输入空间中的输入映射为特征空间中的特征向量；
- 非线性支持向量机：利用一个从输入空间到特征空间的非线性映射将输入映射为特征向量；
- 支持向量机的学习是在特征空间进行的.

支 持 向 量 机

- 假设特征空间上的训练数据集：

$$T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$$

$$x_i\in\mathcal{X}=\mathbf{R}^n,\quad y_i\in\mathcal{Y}=\{+1,-1\},\quad i=1,2,\cdots,N$$

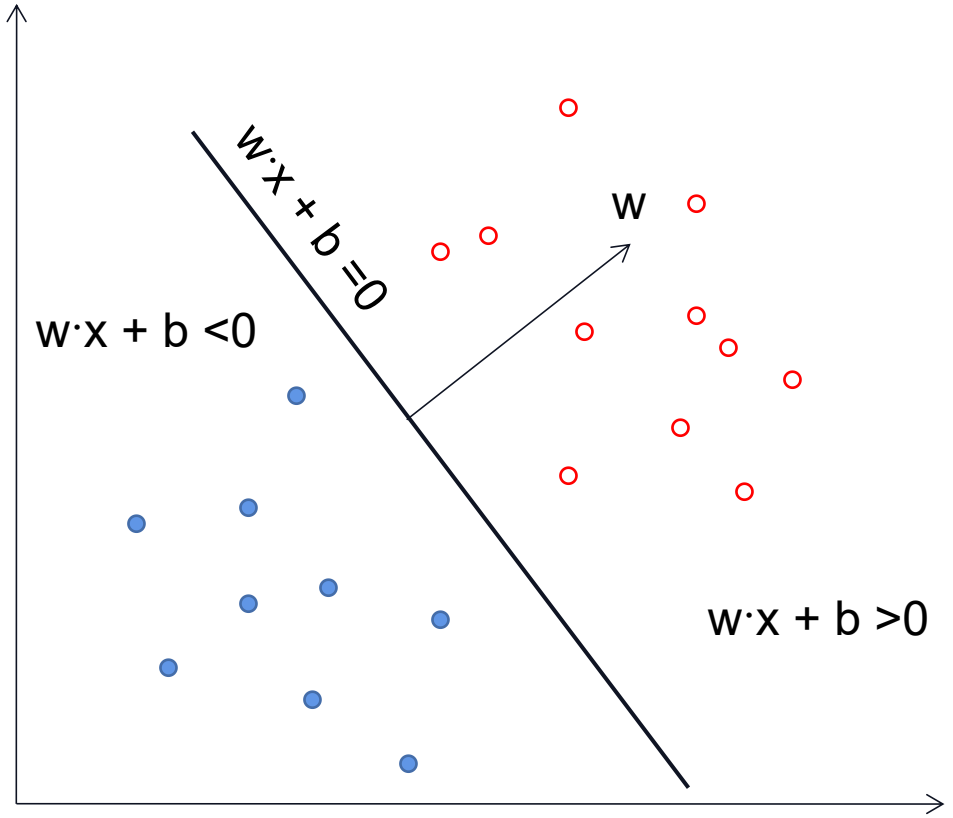
- 正例和负例
- 学习的目标：找到分类超平面，
- 线性可分支持向量机：给定线性可分训练数据集，通过间隔最大化或等价地求解相应的凸二次规划问题学习得到的分离超平面为

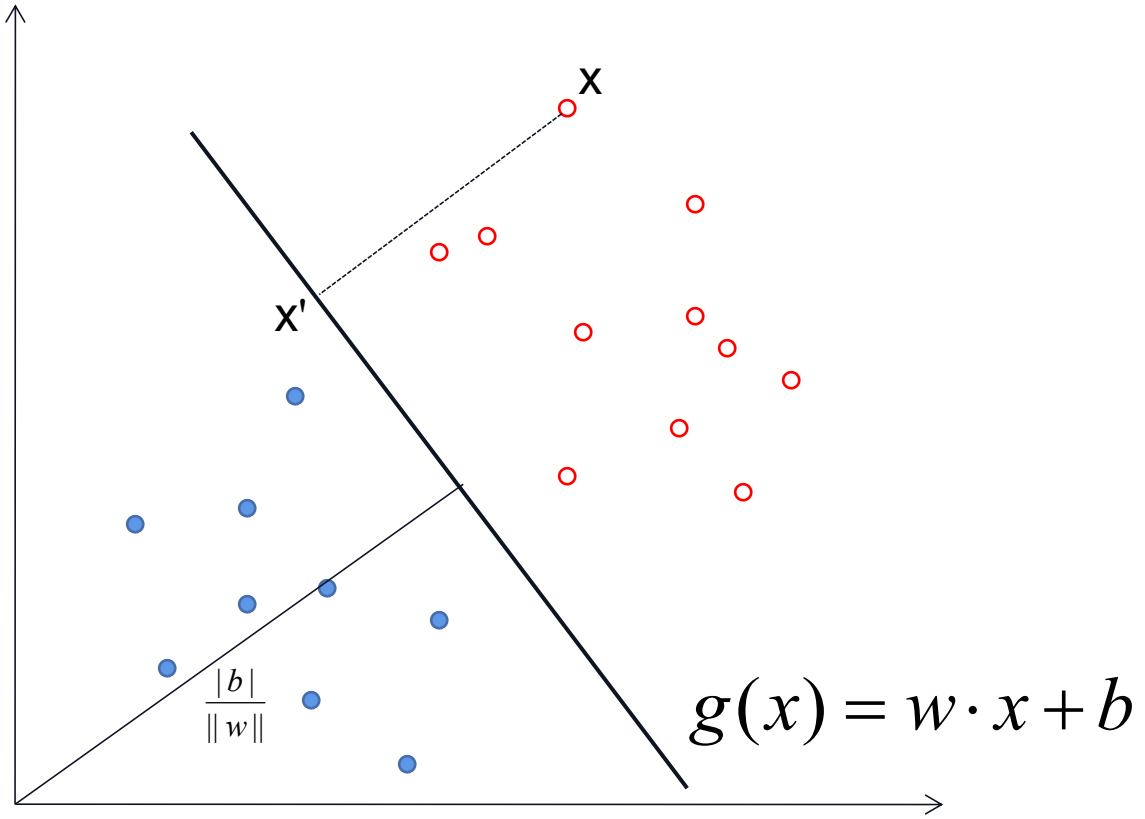
$$w^*\cdot x+b^*=0$$

- 决策函数：

$$f(x)=\text{sign}(w^*\cdot x+b^*)$$

点到超平面的距离：$\frac{|w\cdot x+b|}{\|w\|}$





支 持 向 量 机

■对于给定的训练数据集T和超平面，训练数据集的几何间隔

$$\gamma_i = y_i\left(\frac{w}{\|w\|}\cdot x_i + \frac{b}{\|w\|}\right)$$

■样本点到超平面的最小距离

$$\gamma = \min_{i=1,\cdots,N}\gamma_i$$

■最大间隔分类超平面

$$\max_{w,b} \quad \gamma$$

$$\text{s.t.} \quad y_i\left(\frac{w}{\|w\|}\cdot x_i + \frac{b}{\|w\|}\right) \geqslant \gamma, \quad i=1,2,\cdots,N$$

■根据几何间隔和函数间隔的关系

$$\max_{w,b} \quad \frac{\hat{\gamma}}{\|w\|}$$

$$\text{s.t.} \quad y_i(w\cdot x_i + b) \geqslant \hat{\gamma}, \quad i=1,2,\cdots,N$$

■考虑可以取 $\hat{\gamma}=1$

■最大化 $\frac{1}{\|w\|}$ 和最小化 $\frac{1}{2}\|w\|^2$ 等价

支 持 向 量 机

- 线性可分支持向量机学习的最优化问题

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$

$$\text{s.t.} \quad y_i(w \cdot x_i + b) - 1 \geqslant 0, \quad i = 1, 2, \cdots, N$$

- **凸二次规划(convex quadratic programming)**

支 持 向 量 机

- 输入：线性可分训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$

$$x_i \in \mathcal{X}=\mathbf{R}^n \quad y_i \in \mathcal{Y}=\{-1,+1\}, \quad i=1,2,\cdots,N$$

- 输出：最大间隔分离超平面和分类决策函数
- 1、构造并求解约束最优化问题

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$

$$\text{s.t.} \quad y_i(w\cdot x_i+b)-1\geqslant 0, \quad i=1,2,\cdots,N$$

- 求得w*和b*，得到分离超平面

$$w^*\cdot x+b^*=0$$

- 分类决策函数

$$f(x)=\text{sign}(w^*\cdot x+b^*)$$

支 持 向 量 机

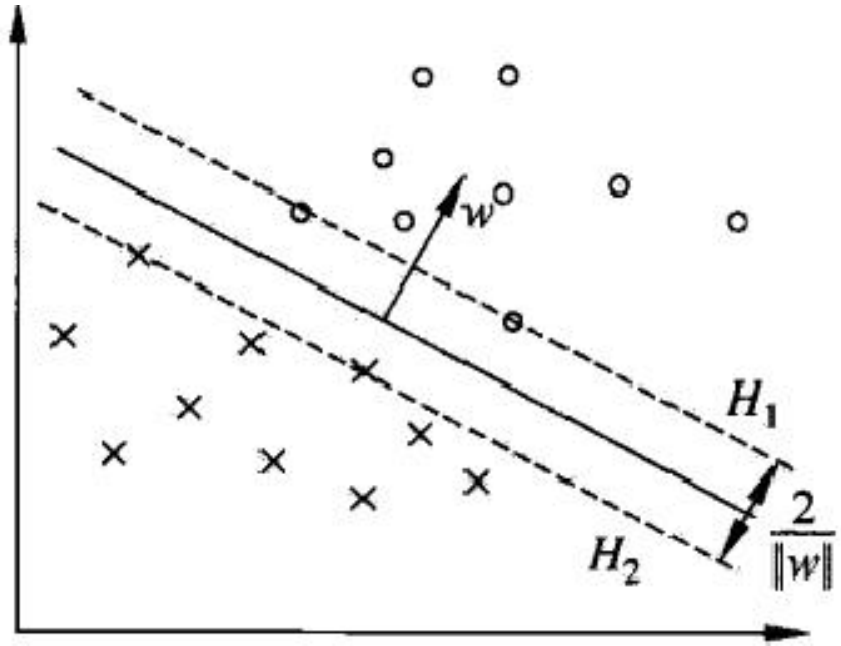
- 在线性可分情况下，训练数据集的样本点中与分离超平面距离最近的样本点的实例称为支持向量(support vector)；
- 支持向量是使约束条件式等号成立的点，即 $y_i(w \cdot x_i + b) - 1 = 0$

- 正例：

$$H_1 : w \cdot x + b = 1$$

- 负例：

$$H_2 : w \cdot x + b = -1$$



例题：正例：(3, 3)，(4, 3)，反例：(1, 1)

$$\min_{w,b} \quad \frac{1}{2}(w_1^2 + w_2^2)$$

$$\text{s.t.} \quad 3w_1 + 3w_2 + b \geqslant 1$$

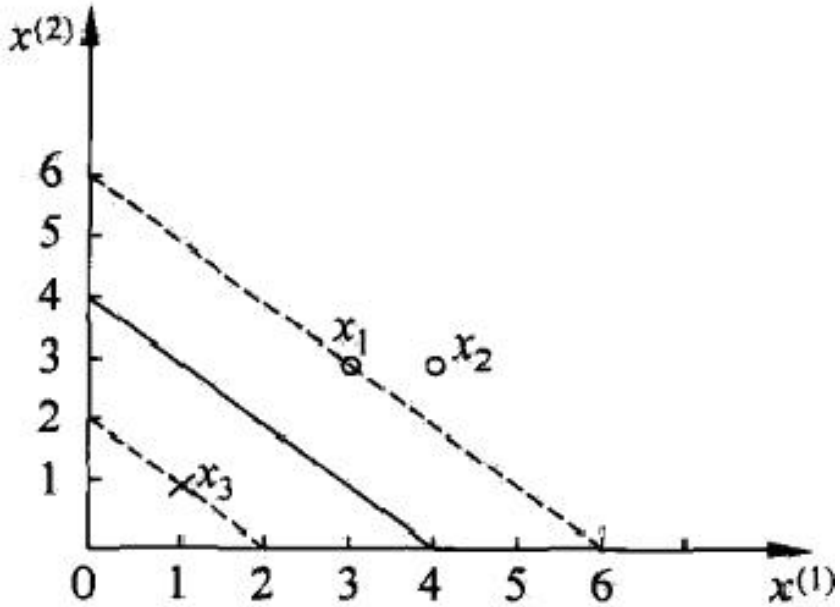
$$4w_1 + 3w_2 + b \geqslant 1$$

$$-w_1 - w_2 - b \geqslant 1$$

$$w_1 = w_2 = \frac{1}{2}, \quad b = -2$$

$$\frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)} - 2 = 0$$

$x_1 = (3,3)^T$ 与 $x_3 = (1,1)^T$ 为支持向量



支 持 向 量 机

- 对于线性可分支持向量机的优化问题，原始问题：

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$

$$\text{s.t.} \quad y_i(w \cdot x_i + b) - 1 \geqslant 0, \quad i = 1, 2, \cdots, N$$

- 应用拉格朗日对偶性，通过求解对偶问题，得到原始问题的解。
- 优点：
  - 对偶问题往往容易解
  - 引入核函数，推广到非线性分类问题

支 持 向 量 机

- 定义拉格朗日函数

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

- 原问题：极小极大，对偶问题：极大极小

$$\min_x \theta_p(x)=\min_x \max_{\alpha,\beta:\alpha_i\geq 0} L(x,\alpha,\beta)$$

$$\max_{\alpha} \min_{w,b} L(w,b,\alpha)$$

支 持 向 量 机

- 先求L(w,b,α)对w，b的极小，再求对α的极大
- 1、求 $\min_{w,b} L(w,b,\alpha)$，对w，b分别求偏导并令等于0
- 由

$$\nabla_w L(w,b,\alpha) = w - \sum_{i=1}^{N} \alpha_i y_i x_i = 0 \quad \Rightarrow \quad w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\nabla_b L(w,b,\alpha) = \sum_{i=1}^{N} \alpha_i y_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{N} \alpha_i y_i = 0$$

- 得:

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

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i \quad \Leftrightarrow \quad \min_{w,b} L(w,b,\alpha)$$

支 持 向 量 机

- 求 $\min\limits_{w,b} L(w,b,\alpha)$ 对α的极大，即是对偶问题：

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

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

$$\alpha_i \geqslant 0, \quad i=1,2,\cdots,N$$

$$\Rightarrow \quad \min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j)-\sum_{i=1}^{N}\alpha_i$$

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

$$\alpha_i \geqslant 0, \quad i=1,2,\cdots,N$$

- 定理：设 $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_l^*)^T$ 是对偶最优问题的解，则存在下标j，使得$\alpha_j^* > 0$，并可按下式求得原始问题的解。

$$w^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i \quad (7.25)$$

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j) \quad (7.26)$$

证明：由

$$\nabla_w L(w^*, b^*, \alpha^*) = w^* - \sum_{i=1}^{N} \alpha_i^* y_i x_i = 0$$

$$\nabla_b L(w^*, b^*, \alpha^*) = -\sum_{i=1}^{N} \alpha_i^* y_i = 0$$

$$\alpha_i^* (y_i (w^* \cdot x_i + b^*) - 1) = 0, \quad i = 1, 2, \cdots, N$$

$$y_i (w^* \cdot x_i + b^*) - 1 \geqslant 0, \quad i = 1, 2, \cdots, N$$

$$\alpha_i^* \geqslant 0, \quad i = 1, 2, \cdots, N$$

得：

$$w^* = \sum_{i} \alpha_i^* y_i x_i$$

证明：由 $w^* = \sum_i \alpha_i^* y_i x_i$，其中至少有一个 $\alpha_j^* > 0$

反证法：

假设：$\alpha^* = 0$，可知w*=0，

但这不是原始优化问题的解，产生矛盾

对此：j 有 $y_j(w^* \cdot x_j + b^*) - 1 = 0$ (7.28)

将式 (7.25) 代入式 (7.28) 并注意到 $y_j^2 = 1$，

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j)$$

- 由此定理可知，分离超平面可以写成：$\sum_{i=1}^{N} \alpha_i^* y_i (x \cdot x_i) + b^* = 0$
- 分类决策函数可以写成：

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i^* y_i (x \cdot x_i) + b^* \right)$$

支 持 向 量 机

- 输入：线性可分训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$

$$x_i \in \mathcal{X}=\mathbf{R}^n \qquad y_i \in \mathcal{Y}=\{-1,+1\}\ , \quad i=1,2,\cdots,N$$

- 输出：最大间隔分离超平面和分类决策函数
- 1、构造并求解约束最优化问题

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

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

$$\alpha_i \geqslant 0\ , \quad i=1,2,\cdots,N$$

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

支 持 向 量 机

- 2、计算

$$w^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$$

- 并选择α*的一个正分量 $\alpha_j^* > 0$，计算

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j)$$

- 3、求得分离超平面

$$w^* \cdot x + b^* = 0$$

- 分类决策函数

$$f(x) = \text{sign}(w^* \cdot x + b^*)$$

支 持 向 量 机

- 考虑原始优化问题和对偶优化问题，
- 将数据集中对应于 $\alpha_j^* > 0$ 的样本 (x_i, y_i) 的实例 x_i 称为支持向量
- 支持向量一定在分割边界上，由KKT互补条件：

$$\alpha_i^*(y_i(w^* \cdot x_i + b^*) - 1) = 0, \quad i = 1, 2, \cdots, N$$

- 对应于 $\alpha_j^* > 0$ 的样本 x_i

$$y_i(w^* \cdot x_i + b^*) - 1 = 0$$

- 或

$$w^* \cdot x_i + b^* = \pm 1$$

- 正例点 $x_1=(3,3)^{\mathrm{T}}$, $x_2=(4,3)^{\mathrm{T}}$　负例点 $x_3=(1,1)^{\mathrm{T}}$
- 解：对偶形式

$$\begin{aligned}\min_{\alpha}\quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)-\sum_{i=1}^{N}\alpha_i\\ &=\frac{1}{2}(18\alpha_1^2+25\alpha_2^2+2\alpha_3^2+42\alpha_1\alpha_2-12\alpha_1\alpha_3-14\alpha_2\alpha_3)-\alpha_1-\alpha_2-\alpha_3\\ \text{s.t.}\quad & \alpha_1+\alpha_2-\alpha_3=0\\ & \alpha_i\geqslant 0,\quad i=1,2,3\end{aligned}$$

- 将 $\alpha_3=\alpha_1+\alpha_2$ 带入目标函数并记为

$$s(\alpha_1,\alpha_2)=4\alpha_1^2+\frac{13}{2}\alpha_2^2+10\alpha_1\alpha_2-2\alpha_1-2\alpha_2$$

- 对 α_1,α_2 求偏导数，并令其为0，易知 $s(\alpha_1,\alpha_2)$ 在$\left(\frac{3}{2},-1\right)^T$ 取极值，但该点不满足约束条件 $\alpha_2 \geqslant 0$，所以最小值应在边界上达到
- 当 $\alpha_1=0$时，最小值$s\left(0,\frac{2}{13}\right)=-\frac{2}{13}$
- 当$\alpha_2=0$ 时，最小值 $s\left(\frac{1}{4},0\right)=-\frac{1}{4}$
- 于是 $s(\alpha_1,\alpha_2)$ 在$\alpha_1=\frac{1}{4},\alpha_2=0$ 获得极小，$\alpha_3=\alpha_1+\alpha_2=\frac{1}{4}$
- 这样 $\alpha_1^*=\alpha_3^*=\frac{1}{4}$ 对应的实例向量为支持向量

$$w_1^*=w_2^*=\frac{1}{2}$$

$$b^*=-2$$

$$\frac{1}{2}x^{(1)}+\frac{1}{2}x^{(2)}-2=0$$

$$f(x)=\text{sign}\left(\frac{1}{2}x^{(1)}+\frac{1}{2}x^{(2)}-2\right)$$

支 持 向 量 机

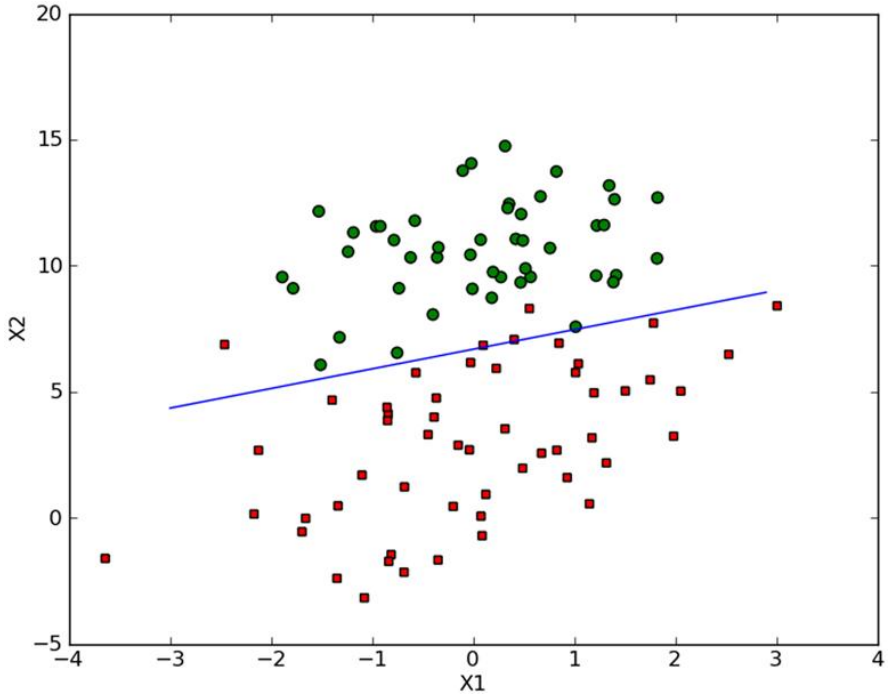
- 训练数据中有一些特异点（outlier），不能满足函数间隔大于等于1的约束条件。
- 解决方法：对每个样本点 (x_i, y_i) 引进一个松弛变量 $\xi_i \geqslant 0$
- 使得函数间隔加上松弛变量

大于等于1，约束条件变为：

$$y_i(w \cdot x_i + b) \geqslant 1 - \xi_i$$

目标函数变为：$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$

其中，C > 0 为惩罚参数



支 持 向 量 机

- 线性不可分的线性支持向量机的学习问题：

$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$\text{s.t.} \quad y_i(w \cdot x_i + b) \geqslant 1 - \xi_i, \quad i = 1,2,\cdots,N$$

$$\xi_i \geqslant 0, \quad i = 1,2,\cdots,N$$

- 可证明w的解是唯一的，b不是，
- 设该问题的解是w*,b*,可得到分离超平面和决策函数

$$w^* \cdot x + b^* = 0$$

$$f(x) = \text{sign}(w^* \cdot x + b^*)$$

支 持 向 量 机

- 原始问题的拉格朗日函数：

$$L(w,b,\xi,\alpha,\mu)\equiv\frac{1}{2}\|w\|^2+C\sum_{i=1}^{N}\xi_i-\sum_{i=1}^{N}\alpha_i(y_i(w\cdot x_i+b)-1+\xi_i)-\sum_{i=1}^{N}\mu_i\xi_i$$

- 其中：$\alpha_i\geqslant 0,\mu_i\geqslant 0$
- 对偶问题是拉格朗日函数的极大极小问题
- 首先求 $L(w,b,\xi,\alpha,\mu)$ 对w,b,ξ的极小，由

$$\nabla_w L(w,b,\xi,\alpha,\mu)=w-\sum_{i=1}^{N}\alpha_i y_i x_i=0$$

$$\nabla_b L(w,b,\xi,\alpha,\mu)=-\sum_{i=1}^{N}\alpha_i y_i=0$$

$$\nabla_{\xi_i} L(w,b,\xi,\alpha,\mu)=C-\alpha_i-\mu_i=0$$

得：

$$w=\sum_{i=1}^{N}\alpha_i y_i x_i$$

$$\sum_{i=1}^{N}\alpha_i y_i=0$$

$$C-\alpha_i-\mu_i=0$$

代入

- 得：

$$\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i$$

- 再对 $\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu)$ 求α的极大，得到对偶问题：

$$\max_{\alpha} \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i$$

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

$$C - \alpha_i - \mu_i = 0 \quad \Rightarrow \quad 0 \leqslant \alpha_i \leqslant C$$

$$\alpha_i \geqslant 0$$

$$\mu_i \geqslant 0, \quad i = 1,2,\cdots,N$$

- 原始问题的对偶问题：

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

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

$$0 \leqslant \alpha_i \leqslant C, \quad i = 1,2,\cdots,N$$

- 定理：设 $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_N^*)^T$ 是对偶问题的一个解，若存在α*的一个分量 α_j^*，$0 < \alpha_j^* < C$，则原始问题的解w*,b*

$$w^* = \sum_{i=1}^{N}\alpha_i^* y_i x_i \qquad b^* = y_j - \sum_{i=1}^{N} y_i \alpha_i^* (x_i \cdot x_j)$$

支 持 向 量 机

- 输入：线性不可分训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$

$$x_i \in \mathcal{X}=\mathbf{R}^n \qquad y_i \in \mathcal{Y}=\{-1,+1\}\ , \quad i=1,2,\cdots,N$$

- 输出：分离超平面和分类决策函数
- 1、构造并求解约束最优化问题

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

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

$$0\leqslant \alpha_i \leqslant C\ , \quad i=1,2,\cdots,N$$

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

- 2、计算

$$w^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$$

- 并选择α*，适合条件 $0 < \alpha_j^* < C$，计算

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j)$$

- 3、求得分离超平面

$$w^* \cdot x + b^* = 0$$

- 分类决策函数

$$f(x) = \text{sign}(w^* \cdot x + b^*)$$

- 线性支持向量机学习还有另外一种解释，就是最小化以下目标函数：

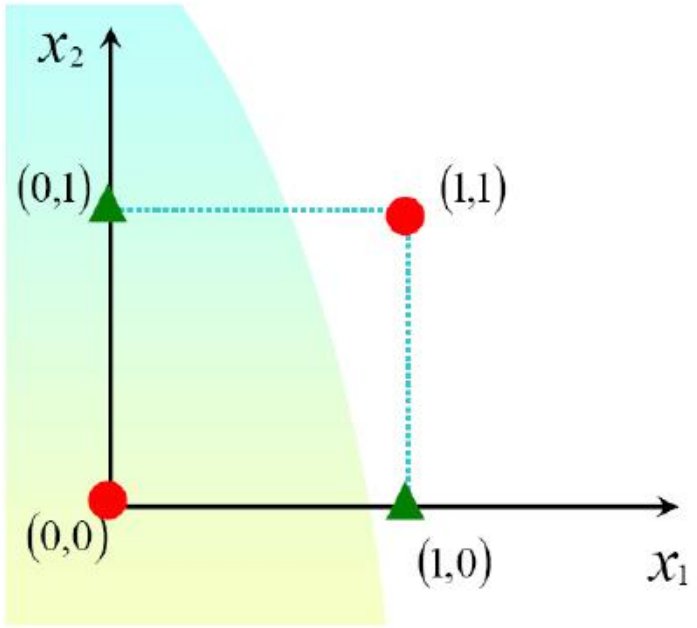
$$\sum_{i=1}^{N}[1-y_i(w\cdot x_i+b)]_+ +\lambda\|w\|^2$$

- 第一项： $L(y(w\cdot x+b))=[1-y(w\cdot x+b)]_+$ 称为合页损失函数

$$[z]_+=\begin{cases}z, & z>0\\ 0, & z\leqslant 0\end{cases}$$

- 线性支持向量机等价于

$$\min_{w,b} \sum_{i=1}^{N}[1-y_i(w\cdot x_i+b)]_+ +\lambda\|w\|^2$$



$\omega_1 : \{(0,0),(1,1)\}$

$\omega_2 : \{(1,0),(0,1)\}$

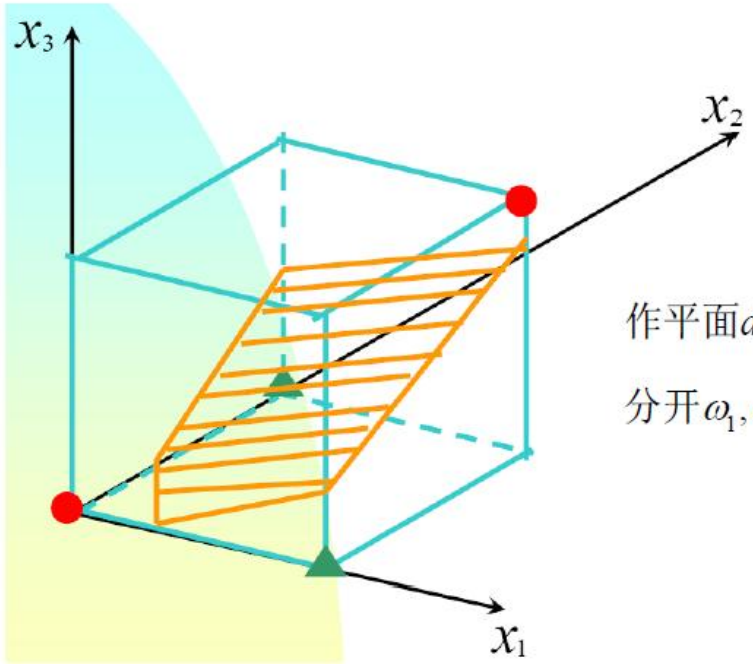
线性识别函数不存在

把2维空间(x_1, x_2)

变为3维空间(x_1, x_2, x_3)

$$x_3 = x_1 x_2$$

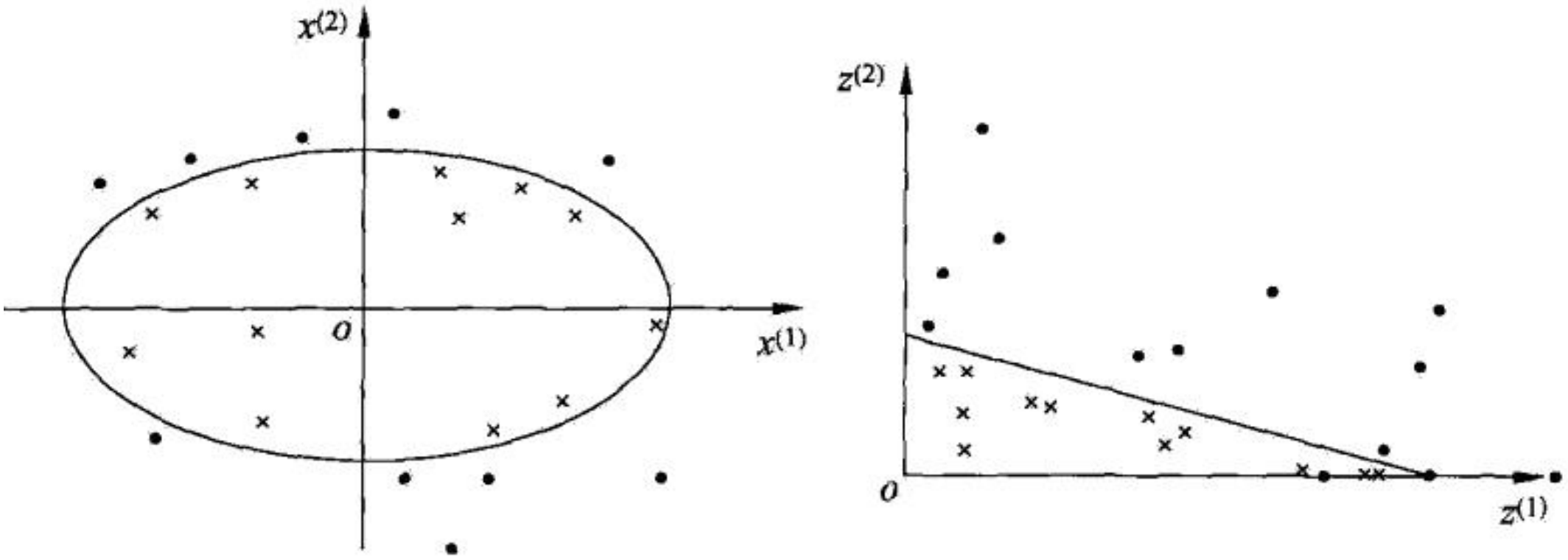
(x_1, x_2)	(x_1, x_2, x_3)
$(0,0)$	$(0,0,0)$
$(1,1)$	$(1,1,1)$
$(1,0)$	$(1,0,0)$
$(0,1)$	$(0,1,0)$



作平面$d(X) = x_1 + x_2 - 2x_3 - \frac{1}{3} = 0$

分开ω_1, ω_2

支 持 向 量 机



如果能用R^n中的一个超曲面将正负例正确分开，则称这个问题为非线性可分问题.

支 持 向 量 机

- 非线性问题往往不好求解，所以希望能用解线性分类间题的方法解决这个问题。
- 采取的方法是进行一个非线性变换，将非线性问题变换为线性问题，通过解变换后的线性问题的方法求解原来的非线性问题。
- 原空间：

$$\mathcal{X} \subset \mathbf{R}^2,\ x = (x^{(1)}, x^{(2)})^{\mathrm{T}} \in \mathcal{X}$$

- 新空间：$\mathcal{Z} \subset \mathbf{R}^2,\ z = (z^{(1)}, z^{(2)})^{\mathrm{T}} \in \mathcal{Z}$　　$z = \phi(x) = ((x^{(1)})^2, (x^{(2)})^2)^{\mathrm{T}}$

$$w_1(x^{(1)})^2 + w_2(x^{(2)})^2 + b = 0 \quad \Rightarrow \quad w_1 z^{(1)} + w_2 z^{(2)} + b = 0$$

支 持 向 量 机

- 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R”或离散集合)对应于一个特征空间(希尔伯特空间)，使得在输入空间中的超曲面模型对应于特征空间中的超平面模型(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.
- **核函数**定义:
- 设X是输入空间(欧氏空间R^n的子集或离散集合)，又设H为特征空间(希尔伯特空间)，如果存在一个从X到H的映射 $\phi(x):\mathcal{X}\to\mathcal{H}$
- 使得对所有 $x,z\in\mathcal{X}$
- 函数K(x,z)满足条件 $K(x,z)=\phi(x)\cdot\phi(z)$，则称 $K(x,z)$ 为核函数，$\phi(x)$ 为映射函数，
- 式中 $\phi(x)\cdot\phi(z)$ 为 $\phi(x)$ 和 $\phi(z)$ 的内积

- 例：假设输入空间是R^2，核函数是 $K(x,z)=(x\cdot z)^2$,试找出其相关的特征空间H和映射$\phi(x):\mathbf{R}^2\to\mathcal{H}$

- 解：取特征空间$\mathcal{H}=\mathbf{R}^3$，记$x=(x^{(1)},x^{(2)})^{\mathrm{T}}$，$z=(z^{(1)},z^{(2)})^{\mathrm{T}}$

$$(x\cdot z)^2=(x^{(1)}z^{(1)}+x^{(2)}z^{(2)})^2=(x^{(1)}z^{(1)})^2+2x^{(1)}z^{(1)}x^{(2)}z^{(2)}+(x^{(2)}z^{(2)})^2$$

- 可以取：$\phi(x)=((x^{(1)})^2,\sqrt{2}x^{(1)}x^{(2)},(x^{(2)})^2)^{\mathrm{T}}$

- 容易验证：$\phi(x)\cdot\phi(z)=(x\cdot z)^2=K(x,z)$

- 同样：

$$\phi(x)=\frac{1}{\sqrt{2}}((x^{(1)})^2-(x^{(2)})^2,2x^{(1)}x^{(2)},(x^{(1)})^2+(x^{(2)})^2)^{\mathrm{T}}$$

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

- 都满足条件。

- 注意到：
- 线性支持向量机对偶问题中，无论是目标函数还是决策函数都只涉及输入实例和实例之间的内积。
- 目标函数中的内积 $x_i \cdot x_j$ 用核函数 $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ 代替，目标函数：

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

- 决策函数：

$$f(x) = \text{sign}\left(\sum_{i=1}^{N_s} a_i^* y_i \phi(x_i) \cdot \phi(x) + b^*\right) = \text{sign}\left(\sum_{i=1}^{N_s} a_i^* y_i K(x_i, x) + b^*\right)$$

- 正定核的等价定义
- 设$\mathcal{X} \subset \mathbf{R}^n$，是K(x,z)是定义在$\mathcal{X} \times \mathcal{X}$ 对称函数，如果对任意的$x_i \in \mathcal{X}$，$i=1,2,\cdots,m$，K(x,z)对应的Gram矩阵

$$K=\left[K(x_i,x_j)\right]_{m\times m}$$

- 半正定的，则称K(x,z)为正定核。
- 这一定义在构造核函数时很有用。但对于一个具体函数K(x,z) 来说，检验它是否为正定核函数并不容易，因为要求对任意有限输入集$\{x_1,x_2,\cdots,x_m\}$验证K对应的Gram矩阵是否为半正定的。
- 在实际问题中往往应用已有的核函数。

支 持 向 量 机

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	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\delta^2}\right)$	$\delta > 0$为高斯核的带宽(width)
拉普拉斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|}{\delta}\right)$	$\delta > 0$
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, $\beta > 0, \theta < 0$

- 输入：线性不可分训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$

$$x_i \in \mathcal{X}=\mathbf{R}^n \qquad y_i \in \mathcal{Y}=\{-1,+1\}\ ,\quad i=1,2,\cdots,N$$

- 输出：分类决策函数
- 1、选取适当的核函数和参数C，构造最优化问题：

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

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

$$0\leqslant\alpha_i\leqslant C\ ,\quad i=1,2,\cdots,N$$

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

支 持 向 量 机

- 2、并选择α*，适合条件 $0<\alpha_j^*<C$，计算

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i K(x_i \cdot x_j)$$

- 3、构造决策函数

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i^* y_i K(x \cdot x_i) + b^* \right)$$

- 当K(x,z)是正定核函数时，非线性支持向量机问题是凸二次规划问题，解是存在的。

支 持 向 量 机

- 序列最小最优化(sequential minimal optimization SMO)算法：1998年由Platt提出。

John C. Platt, "Using Analytic QP and Sparseness to Speed Training of Support Vector Machines" in *Advances in Neural Information Processing Systems* 11, M. S. Kearns, S. A. Solla, D. A. Cohn, eds (MIT Press, 1999), 557–63.

- 动机：

- 支持向量机的学习问题可以形式化为求解凸二次规划问题.这样的凸二次规划问题具有全局最优解，并且有许多最优化算法可以用于这一问题的求解；

- 但是当训练样本容量很大时，这些算法往往变得非常低效，以致无法使用. 所以如何高效地实现支持向量机学习就成为一个重要的问题。

支 持 向 量 机

- SMO（Sequential minimal optimization ）
- 解如下凸二次规划的对偶问题

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

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

$$0 \leqslant \alpha_i \leqslant C, \quad i=1,2,\cdots,N$$

- 注意：变量是拉格朗日乘子α_i，一个对应一个样本

- 启发式算法，基本思路：
- 如果所有变量的解都满足此最优化问题的KKT条件，那么得到解；
- 否则，选择两个变量，固定其它变量，针对这两个变量构建一个二次规划问题，称为子问题，可通过解析方法求解，提高了计算速度。
- 子问题的两个变量：一个是违反KKT条件最严重的那个，另一个由约束条件自动确定。

$$\alpha_1 = -y_1 \sum_{i=2}^{N} \alpha_i y_i$$

- SMO算法包括两个部分：
  - 求解两个变量二次规划的解析方法
  - 选择变量的启发式方法

支 持 向 量 机

- 选择两个变量，其它固定，SMO的子问题：

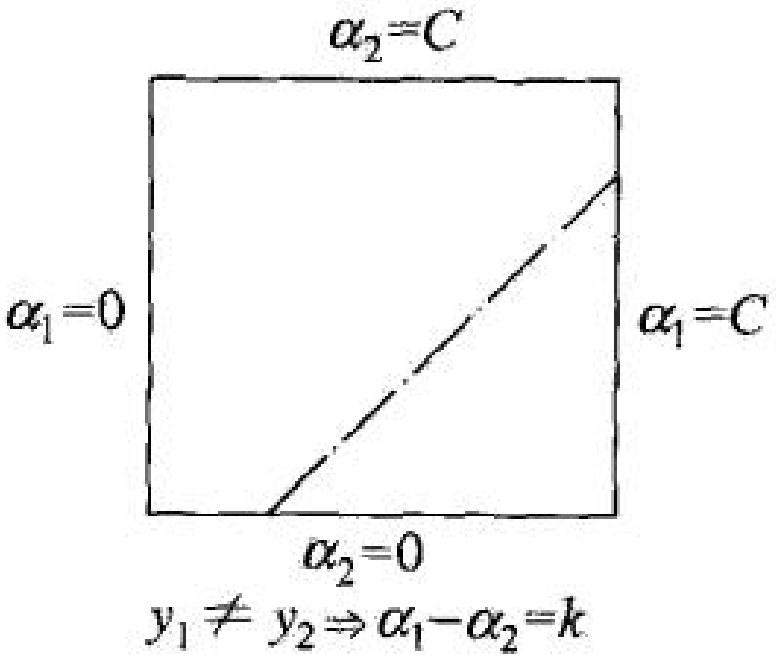
$$\min_{\alpha_1,\alpha_2} \quad W(\alpha_1,\alpha_2)=\frac{1}{2}K_{11}\alpha_1^2+\frac{1}{2}K_{22}\alpha_2^2+y_1y_2K_{12}\alpha_1\alpha_2$$
$$-(\alpha_1+\alpha_2)+y_1\alpha_1\sum_{i=3}^{N}y_i\alpha_iK_{i1}+y_2\alpha_2\sum_{i=3}^{N}y_i\alpha_iK_{i2}$$

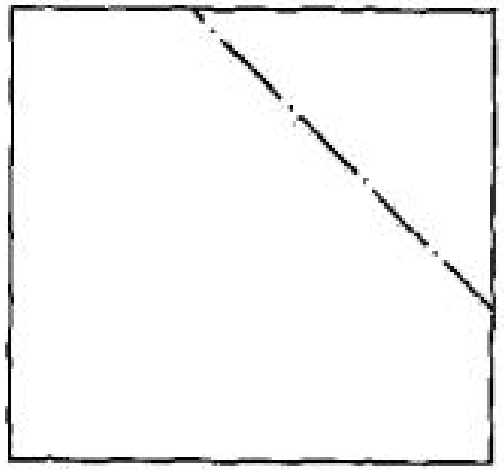
$$\text{s.t.} \quad \alpha_1y_1+\alpha_2y_2=-\sum_{i=3}^{N}y_i\alpha_i=\varsigma$$

$$0\leqslant\alpha_i\leqslant C, \quad i=1,2$$

支 持 向 量 机

- 两个变量，约束条件用二维空间中的图形表示





- 假设子问题的初始可行解为 $\alpha_1^{old}, \alpha_2^{old}$， 最优解 $\alpha_1^{new}, \alpha_2^{new}$
- 设α_2未经剪辑时的最优解为 $\alpha_2^{new,unc}$

- 根据不等式条件 α_2^{new} 的取值范围：

$$L \leqslant \alpha_2^{new} \leqslant H$$

- 左图： $L = \max(0, \alpha_2^{old} - \alpha_1^{old})$ $H = \min(C, C + \alpha_2^{old} - \alpha_1^{old})$
- 右图： $L = \max(0, \alpha_2^{old} + \alpha_1^{old} - C)$ $H = \min(C, \alpha_2^{old} + \alpha_1^{old})$
- 求解过程：
- 先求沿着约束方向未经剪辑时的 $\alpha_2^{new,unc}$
- 再求剪辑后的 α_2^{new}
- 记： $g(x) = \sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b$ 令： $E_i = g(x_i) - y_i = \left(\sum_{j=1}^{N} \alpha_j y_j K(x_j, x_i) + b \right) - y_i$， $i = 1, 2$
- E 为输入x的预测值和真实输出y的差，i=1，2

- **定理：**

- 最优化子问题沿约束方向未经剪辑的解：

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$

$$\eta = K_{11} + K_{22} - 2K_{12} = \left\| \Phi(x_1) - \Phi(x_2) \right\|^2$$

- 剪辑后的解

$$\alpha_2^{new} = \begin{cases} H, & \alpha_2^{new,unc} > H \\ \alpha_2^{new,unc}, & L \leqslant \alpha_2^{new,unc} \leqslant H \\ L, & \alpha_2^{new,unc} < L \end{cases}$$

- 得到α_1的解 $\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_2^{new})$

- 证明： 引进记号

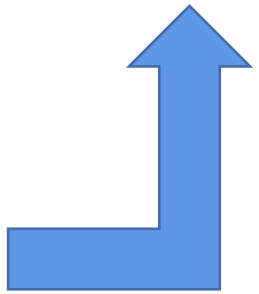
$$v_i=\sum_{j=3}^{N}\alpha_j y_j K(x_i,x_j)=g(x_i)-\sum_{j=1}^{2}\alpha_j y_j K(x_i,x_j)-b,\quad i=1,2$$

- 目标函数写成：

$$W(\alpha_1,\alpha_2)=\frac{1}{2}K_{11}\alpha_1^2+\frac{1}{2}K_{22}\alpha_2^2+y_1y_2K_{12}\alpha_1\alpha_2$$
$$-(\alpha_1+\alpha_2)+y_1v_1\alpha_1+y_2v_2\alpha_2$$

- 由 $\alpha_1y_1=\varsigma-\alpha_2y_2$ 及 $y_i^2=1$

$$\alpha_1=(\varsigma-y_2\alpha_2)y_1$$



- 得到只是α2 的函数的目标函数

$$W(\alpha_2)=\frac{1}{2}K_{11}(\varsigma-\alpha_2 y_2)^2+\frac{1}{2}K_{22}\alpha_2^2+y_2K_{12}(\varsigma-\alpha_2 y_2)\alpha_2$$
$$-(\varsigma-\alpha_2 y_2)y_1-\alpha_2+v_1(\varsigma-\alpha_2 y_2)+y_2v_2\alpha_2$$

- 对α2求导

$$\frac{\partial W}{\partial \alpha_2}=K_{11}\alpha_2+K_{22}\alpha_2-2K_{12}\alpha_2-K_{11}\varsigma y_2+K_{12}\varsigma y_2+y_1y_2-1-v_1y_2+y_2v_2$$

- 令其为0：

$$(K_{11}+K_{22}-2K_{12})\alpha_2=y_2(y_2-y_1+\varsigma K_{11}-\varsigma K_{12}+v_1-v_2)$$
$$=y_2\left[y_2-y_1+\varsigma K_{11}-\varsigma K_{12}+\left(g(x_1)-\sum_{j=1}^{2}y_j\alpha_jK_{1j}-b\right)-\left(g(x_2)-\sum_{j=1}^{2}y_j\alpha_jK_{2j}-b\right)\right]$$

- 将 $\varsigma = \alpha_1^{old} y_1 + \alpha_2^{old} y_2$ 代入：

$$\begin{aligned}(K_{11}+K_{22}-2K_{12})\alpha_2^{new,unc} &= y_2((K_{11}+K_{22}-2K_{12})\alpha_2^{old} y_2 + y_2 - y_1 + g(x_1) - g(x_2)) \\ &= (K_{11}+K_{22}-2K_{12})\alpha_2^{old} + y_2(E_1 - E_2)\end{aligned}$$

- 将 $\eta = K_{11} + K_{22} - 2K_{12}$ 代入：

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$

- 得到定理：

- 最优化子问题沿约束方向未经剪辑的解：

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$

$$\eta = K_{11} + K_{22} - 2K_{12} = \left\| \Phi(x_1) - \Phi(x_2) \right\|^2$$

- 剪辑后的解

$$\alpha_2^{new} = \begin{cases} H, & \alpha_2^{new,unc} > H \\ \alpha_2^{new,unc}, & L \leqslant \alpha_2^{new,unc} \leqslant H \\ L, & \alpha_2^{new,unc} < L \end{cases}$$

- 得到α_1的解 $\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_2^{new})$

支 持 向 量 机

- SMO算法在每个子问题中选择两个变量优化，其中至少一个变量是违反KKT条件的
- 1、第一个变量的选择：外循环
- 违反KKT最严重的样本点，
- 检验样本点是否满足KKT条件：

先检查 →

$$\alpha_i = 0 \Leftrightarrow y_i g(x_i) \geqslant 1$$

$$0 < \alpha_i < C \Leftrightarrow y_i g(x_i) = 1$$

$$\alpha_i = C \Leftrightarrow y_i g(x_i) \leqslant 1$$

$$g(x_i) = \sum_{j=1}^{N} \alpha_j y_j K(x_i, x_j) + b$$

- 2、第二个变量的检查：内循环，
  - 选择的标准是希望能使目标函数有足够大的变化即对应 $|E_1 - E_2|$ 最大，即E1，E2 的符号相反，差异最大
  - 如果内循环通过上述方法找到的点不能使目标函数有足够的下降则：遍历间隔边界上的样本点，测试目标函数下降
  - 如果下降不大，则遍历所有样本点。如果依然下降不大，则丢弃外循环点，重新选择

$$\sum_{i=1}^{N}\alpha_i y_i K_{i1} + b = y_1 \qquad b_1^{\text{new}} = y_1 - \sum_{i=3}^{N}\alpha_i y_i K_{i1} - \alpha_1^{\text{new}} y_1 K_{11} - \alpha_2^{\text{new}} y_2 K_{21}$$

$$E_i = g(x_i) - y_i = \left(\sum_{j=1}^{N}\alpha_j y_j K(x_j, x_i) + b\right) - y_i, \quad i = 1,2$$

$$E_1 = \sum_{i=3}^{N}\alpha_i y_i K_{i1} + \alpha_1^{\text{old}} y_1 K_{11} + \alpha_2^{\text{old}} y_2 K_{21} + b^{\text{old}} - y_1$$

支 持 向 量 机

$$y_1 - \sum_{i=3}^{N} \alpha_i y_i K_{i1} = -E_1 + \alpha_1^{old} y_1 K_{11} + \alpha_2^{old} y_2 K_{21} + b^{old}$$

$$b_1^{new} = -E_1 - y_1 K_{11}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{21}(\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

$$0 < \alpha_2^{new} < C$$

$$b_2^{new} = -E_2 - y_1 K_{12}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{22}(\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

$$E_i^{new} = \sum_{S} y_j \alpha_j K(x_i, x_j) + b^{new} - y_i$$

S 是所有支持向量 x_j 的集合

支 持 向 量 机

- 输入：训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$
- $x_i\in\mathcal{X}=\mathbf{R}^n$ $y_i\in\mathcal{Y}=\{-1,+1\}$，$i=1,2,\cdots,N$，精度ε
- 输出：近似解 α

（1）取初值 $\alpha^{(0)}=0$，令 $k=0$

（2）选取优化变量 $\alpha_1^{(k)},\alpha_2^{(k)}$，解析求解两个变量的最优化问题

求得最优解 $\alpha_1^{(k+1)},\alpha_2^{(k+1)}$，更新 α 为 $\alpha^{(k+1)}$；

（3）若在精度 ε 范围内满足停机条件

$$\sum_{i=1}^{N}\alpha_i y_i=0$$

$$0\leqslant\alpha_i\leqslant C,\quad i=1,2,\cdots,N$$

则转 (4)；否则令 $k=k+1$，转 (2)；

（4）取 $\hat{\alpha}=\alpha^{(k+1)}$

$$y_i\cdot g(x_i)=\begin{cases}\geqslant 1, & \{x_i\mid\alpha_i=0\}\\ =1, & \{x_i\mid 0<\alpha_i<C\}\\ \leqslant 1, & \{x_i\mid\alpha_i=C\}\end{cases}$$

$$g(x_i)=\sum_{j=1}^{N}\alpha_j y_j K(x_j,x_i)+b$$

感 谢 观 看

统 计 机 器 学 习

主讲人：彭振华

数学与计算机学院

2026年