机器学习
Machine Learning

北京航空航天大学计算机学院
School of Computer Science and Engineering, Beihang University
刘庆杰　陈佳鑫

2025年春季学期
Spring 2025

5.1 支持向量机

- 支持向量机的发展简史
- 支持向量机的基本思想

支持向量机 (Support Vector Machine)

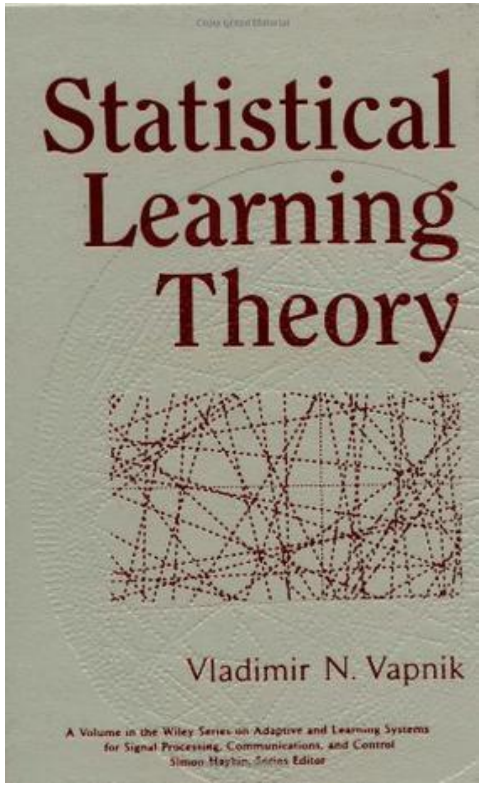
- **什么是支持向量机（SVM）**
  - 支持向量机是基于统计学习理论 (Statistical Learning Theory, SLT) 发展起来的一种机器学习的方法
  - 统计学习理论主要创立者是Vladimir N. Vapnik

1964年-1990年 莫斯科控制科学学院 曾担任计算机科学与研究系主任
1991年-2001年 美国AT&T贝尔实验室 发明支持向量机理论
2002年-2014年 NEC实验室 (美国) 从事机器学习研究
2014年-2016年 美国Facebook公司 从事人工智能研究
2016年 美国Vencore实验室 继续研究工作
2006年 入选美国国家工程院院士。


Vladimir N. Vapnik
（1936 - ）

支持向量机 (Support Vector Machine)

- V. Vapnik对于统计机器学习做出了巨大贡献
  - 1968年，Vapnik和Chervonenkis提出了VC熵和VC维的概念，这些是统计学习理论的核心概念。同时，他们发现了泛函空间的大数定理，得到了关于收敛速度的非渐进界的主要结论。
  - 1974年，Vapnik和Chervonenkis提出了结构风险最小化归纳原则。
  - 1989年，Vapnik和Chervonenkis发现了经验风险最小化归纳原则和最大似然方法一致性的充分必要条件，完成了对经验风险最小化归纳推理的分析。
  - 90年代中期，有限样本情况下的机器学习理论研究逐渐成熟起来，Vapnik对统计学习理论进行了较完善的总结，其中详细叙述了SVM理论。



支持向量机发展简史

● 支持向量机的发展简史

1963年，Vapnik首次提出了**支持向量方法**，实现对特征子集的划分等价于对整个数据集的划分。

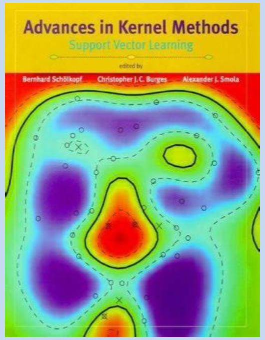
Grace Wahba

Bernhard Boser

Vapnik

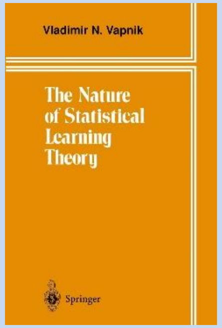
1990年，Grace、Boser和Vapnik等开始了SVM的研究

1999年，MIT出版了《Advances in Kernel Method》，SVM理论的研究与应用**推向了一个高潮**。



1971年，Kimeldorf提出**使用线性不等约束重新构造支持向量的核空间**

1995年，Vapnik的书《The Nature of Statistical Learning Theory》出版，**详细叙述了SVM理论**



之后，SVM的研究主要集中在对**SVM本身性质**的研究和完善，和加大SVM**应用的深度和广度**两方面。

Support-vector networks
C Cortes, V Vapnik
Machine learning 20, 273-297 63331 1995

The Nature of Statistical Learning Theory
V Vapnik
Data mining and knowledge discovery 104843 * 1995

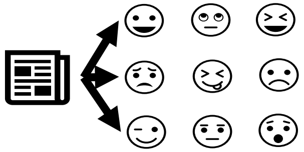
支持向量机的应用

● 支持向量机的应用

➢ 2000年~2010年，SVM被应用于诸多行业：

文本分类

垃圾邮件过滤



情感分类

图像分类

人脸识别

车牌识别

生物信息学

药物发现

蛋白质结构预测

金融预测

股市预测

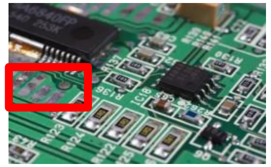
风险预测

医疗诊断

疾病诊断

用药筛选

异常检测



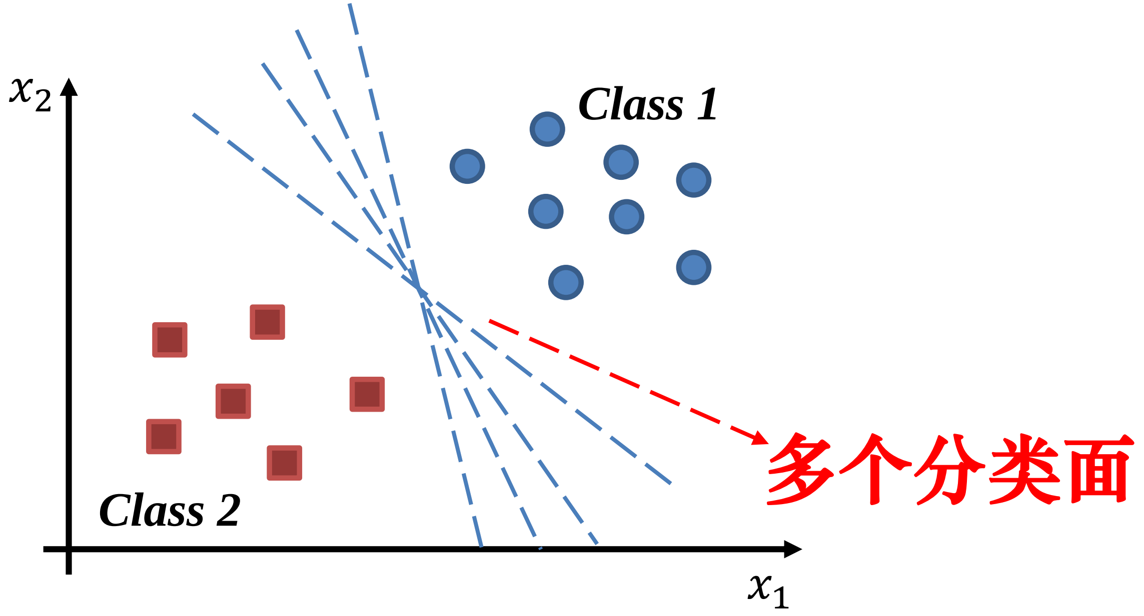
缺陷检测

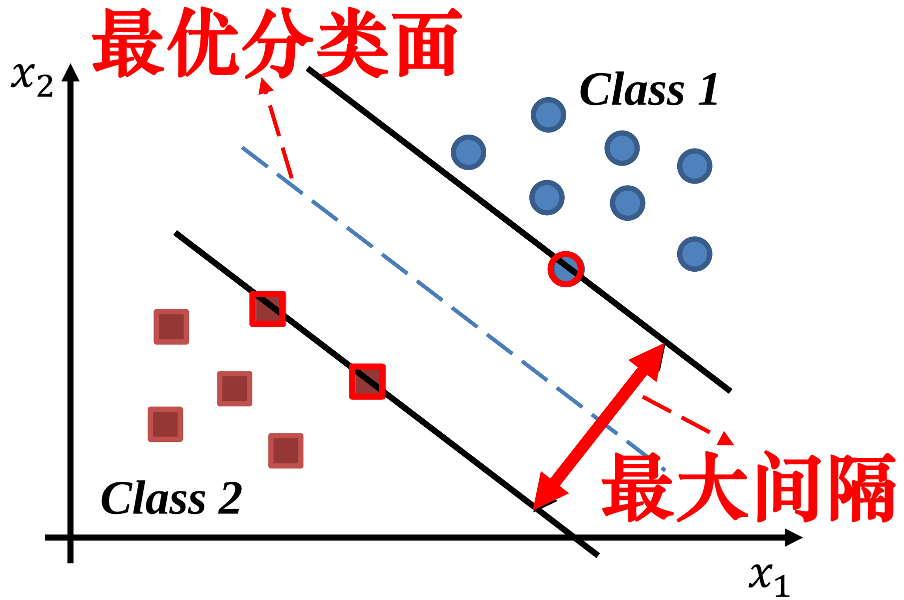
网络安全

➢ 当前，由于SVM对样本规模要求少，仍然被用于一些样本受限的研究领域，如疾病诊断、基因表达分析等。

支持向量机的基本思想

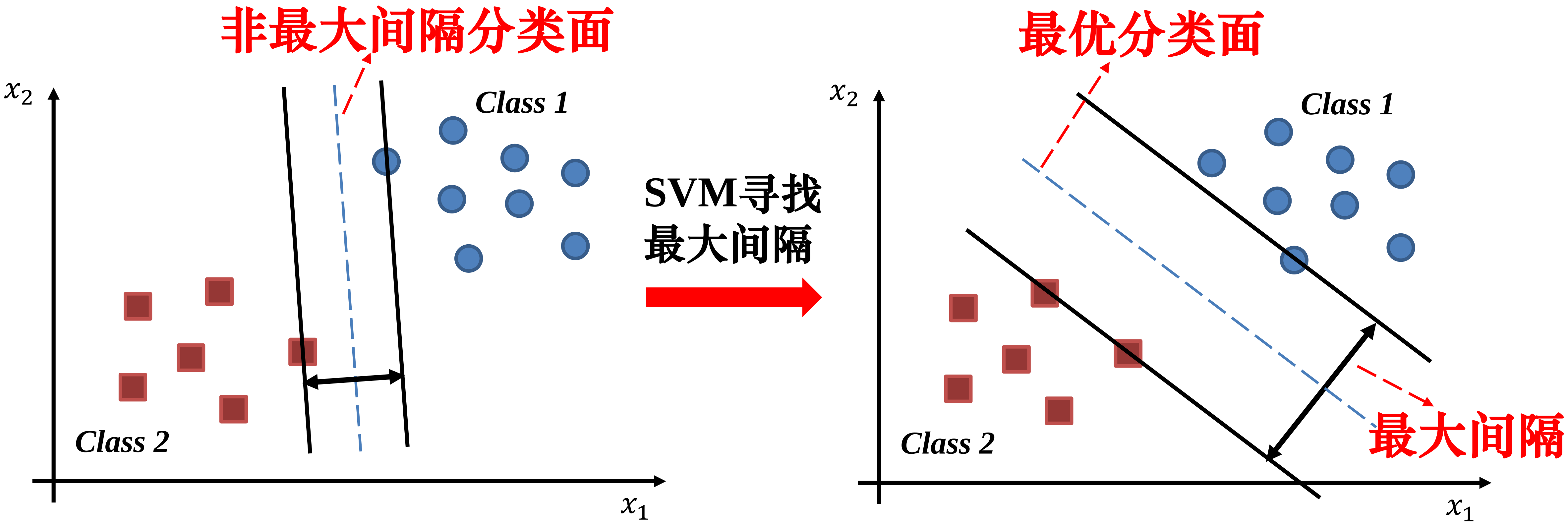
- SVM从线性可分情况下的**最优分类面**发展而来
  - 分类面可能存在多个，但是哪一个是最优的分类平面?
  - **最优分类面**：要求分类面**不但能将两类正确分开** (训练错误率为0)，且使分类**间隔**最大，这里的最优分类面是一个硬间隔分类面
  - **间隔 (Margin)**：距离分类面最近的两个异类样本到分类面的距离之和





支持向量机的基本思想

- **SVM如何求取最优分类面**
  - SVM考虑寻找一个满足分类要求的超平面，**使训练集中的点距离分类面尽可能的远**，即SVM是通过寻找最大间隔来求取最优分类面

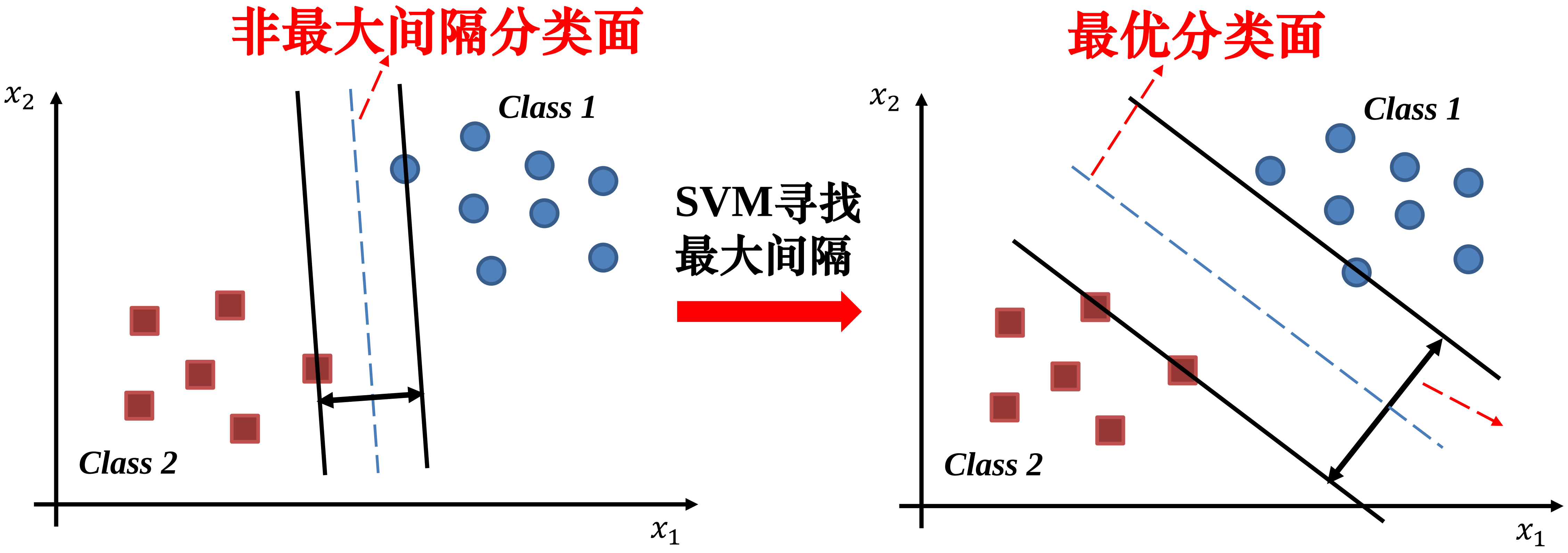


5.2 线性支持向量机

- 支持向量机的建模方法
- 支持向量机的模型推导

线性支持向量机

- 支持向量机如何求取最优分类面?
  - SVM通过寻找最大间隔来求取最优分类面



线性支持向量机建模方法

● 线性支持向量机建模

➢ 样本集： $\{x_n, t_n\}, n = 1, 2, \dots, N, x_n \in \mathbb{R}^d, t_n \in \{-1, 1\}$

➢ 分类超平面：$y(x) = w^T x + b$

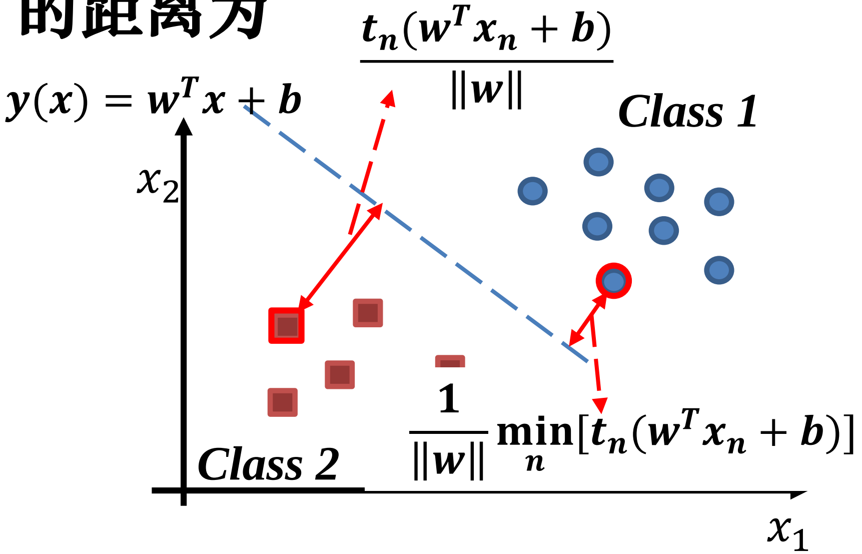
当分类超平面可以将训练样本正确分类时，有 $t_n = \begin{cases} +1, & w^T x_n + b > 0 \\ -1, & w^T x_n + b < 0 \end{cases}$，即 $t_n y(x_n) > 0$

样本集上任意一点 x_n 到分类超平面 (满足 $t_n y(x_n) > 0$) 的距离为

$$\frac{t_n y(x_n)}{\|w\|} = \frac{t_n(w^T x_n + b)}{\|w\|}$$

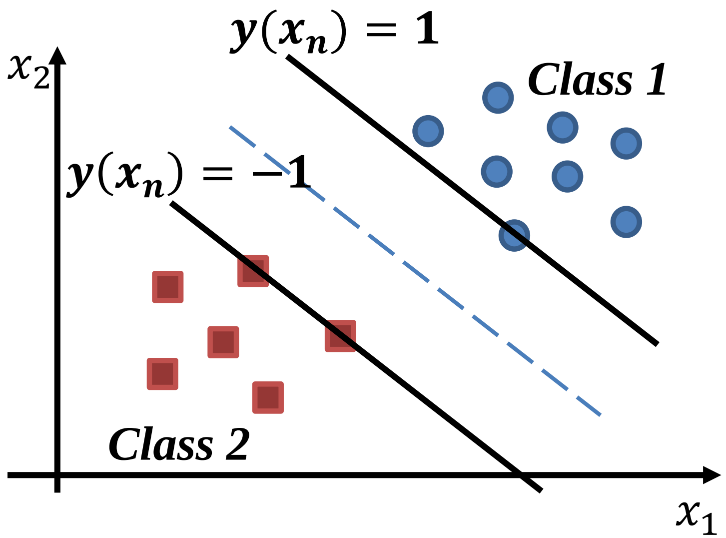
距离超平面最近的样本到分类超平面的距离为

$$\frac{1}{\|w\|} \min_n [t_n(w^T x_n + b)]$$



线性支持向量机建模方法

- 线性支持向量机建模
  - 求解有最大间隔的分类超平面的目标函数为



$$\max_{w,b}\{\frac{1}{\|w\|}\min_{n}[t_n(w^T x_n + b)]\} \Rightarrow \text{求解复杂}$$

由于$w \rightarrow kw, b \rightarrow kb$不影响求解目标，则令距离超平面最近的点为

$$t_n(w^T x_n + b) = 1$$

样本集中所有点满足 $t_n(w^T x_n + b) \geq 1$

则求解原问题可以转为求解一个更简单的问题：

$$\max_{w,b}\{\frac{1}{\|w\|}\min_{n}[t_n(w^T x_n + b)]\} \Rightarrow \max_{w,b} \frac{1}{\|w\|} \quad s.t. \quad t_n(w^T x_n + b) \geq 1$$

线性支持向量机模型推导

优化问题求解

- 原优化问题转化为二次规划问题

$$\max_{w,b} \frac{1}{\|w\|} \quad s.t. \quad t_n(w^T x_n + b) \geq 1 \quad \xrightarrow{\text{二次规划问题}} \quad \min_{w,b} \frac{1}{2}\|w\|^2 \quad s.t. \quad t_n(w^T x_n + b) \geq 1 \tag{5-2-1}$$

- 使用拉格朗日乘子法将上述二次规划问题转化为等价的对偶问题来简化求解
- 该问题的拉格朗日函数可以表示为

$$L(w,b,a) = \frac{1}{2}\|w\|^2 - \sum_{n=1}^{N} a_n \{t_n(w^T x_n + b) - 1\}, \ \boxed{a_n} \geq 0 \tag{5-2-2}$$

拉格朗日乘子

线性支持向量机模型推导

● 优化问题求解

➢ 分别对变量求导

$$\frac{\partial L(w,b,a)}{\partial w}=w-\sum_{n=1}^{N}a_n t_n x_n=0,\quad \frac{\partial L(w,b,a)}{\partial b}=\sum_{n=1}^{N}a_n t_n=0 \tag{5-2-3}$$

➢ 带入拉格朗日函数，得到 (5-2-1) 的对偶形式

$$\begin{aligned}&\max_{a}\sum_{n=1}^{N}a_n-\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}a_n a_m t_n t_m x_n^T x_m\\ &s.t.\quad a_n\geq 0,\ n=1,\dots,N,\ \sum_{n=1}^{N}a_n t_n=0\end{aligned} \tag{5-2-4}$$

➢由于 (5-2-4) 存在不等式约束，需要验证最优解是否满足KKT (Karush-Kuhn-Tucker) 条件，才可以确定对偶问题是否等价于原问题

线性支持向量机模型推导

拉格朗日对偶性

- 不失一般性，设存在不等式约束的优化问题

优化目标：

$$\min_{w} f(w)$$

$$s.t.\quad g_i(w) \le 0, i = 1, \dots, k.$$

$$h_i(w) = 0, i = 1, \dots, l.$$

拉格朗日函数：

$$L(w,\alpha,\beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

- 定义 $\theta_P(w) = \max_{\alpha,\beta\,;\,\alpha_i \ge 0} L(w,\alpha,\beta)$，只有当满足约束条件，才有最大值

$$\theta_P(w) = \begin{cases} f(w), & \text{若 } w \text{ 满足约束条件} \\ +\infty, & \text{否则} \end{cases}$$

线性支持向量机模型推导

- **拉格朗日对偶性**
  - 原问题 $\min_{w} f(w)$ 转化为 p^*

$$p^* = \min_{w} \theta_P(w) = \min_{w} \max_{\alpha,\beta\,;\,\alpha_i \geq 0} L(w,\alpha,\beta)$$

  - 直接求解较困难，进而转向求解其对偶问题 d^*

$$\theta_D(\alpha,\beta) = \min_{w} L(w,\alpha,\beta)$$

$$d^* = \max_{\alpha,\beta\,;\,\alpha_i \geq 0} \theta_D(\alpha,\beta) = \max_{\alpha,\beta\,;\,\alpha_i \geq 0} \min_{w} L(w,\alpha,\beta)$$

线性支持向量机模型推导

● 拉格朗日对偶性

➢ p^*和 d^* 存在如下关系

$$d^* = \max \theta_D(\alpha,\beta) = \theta_D(\alpha^*,\beta^*) = \min L(w,\alpha^*,\beta^*) \le L(w^*,\alpha^*,\beta^*)$$

取最大值时的解　　取最小值时的解

$$= f(w^*) + \sum_{i=1}^{k} \alpha_i^* g_i(w^*) + \sum_{i=1}^{l} \beta_i^* h_i(w^*) \le f(w^*) = p^*$$

$g_i(w) \le 0$　　$h_i(w) = 0$　　$p^* = \min_{w} \theta_P(w) = f(w^*)$

➢在无其他条件约束下，$d^* \le p^*$，两个对偶问题是不一定等价的，那满足什么条件可以使原问题与对偶问题等价，即 $d^* = p^*$呢？

线性支持向量机模型推导

拉格朗日对偶性

➢ 若使 $d^* = p^*$，则应将下式中的不等号取等号：

$$d^* = \max \theta_D(\alpha,\beta) = \theta_D(\alpha^*,\beta^*) = \min L(w,\alpha^*,\beta^*) \boxed{\le} L(w^*,\alpha^*,\beta^*)$$

$$= f(w^*) + \boxed{\sum_{i=1}^{k} \alpha_i^* g_i(w^*)} + \boxed{\sum_{i=1}^{l} \beta_i^* h_i(w^*)} \boxed{\le} f(w^*) = p^*$$

当 $w = w^*$ 时，此时取等号，即满足 $\frac{\partial}{\partial w_i} L(w^*,\alpha^*,\beta^*) = 0$ 时，等号成立

若 $\alpha_i^* g_i(w^*) = 0$ 时，此部分整体为0

$h_i(w) = 0$，此部分已知为0

当 $\alpha_i^* g_i(w^*) = 0$ 时，此时取等号

➢ 所以，当同时满足 $\frac{\partial}{\partial w_i} L(w^*,\alpha^*,\beta^*) = 0$ 和 $\alpha_i^* g_i(w^*) = 0$ 时，原问题与对偶问题等价。这两个条件和之前已知的约束条件共同组成了KKT条件

（详细推导过程见附录）

线性支持向量机模型推导

● KKT (Karush-Kuhn-Tucker) 条件

$$\frac{\partial}{\partial w_i}L(w^*,\alpha^*,\beta^*)=0, i=1,\dots,n$$

$$\boxed{\alpha_i^* g_i(w^*)=0, i=1,\dots,k}$$

KKT互补松弛条件

$$g_i(w^*)\le 0, i=1,\dots,k$$

$$\alpha^*\ge 0, i=1,\dots,k$$

如果w^*, α^*, β^*满足上述的KKT条件，那么它们就是原问题和对偶问题的解。

➢ KKT互补松弛条件分析：

补充条件隐含如果 $\alpha^*>0$，那么 $g_i(w^*)=0$，即w处于可行域的边界上，是起作用的(Active)约束，而位于可行域内部的点都是不起作用的约束，其 $\alpha^*=0$。

线性支持向量机模型推导

- 求解分类超平面

➢ 回归到SVM求解问题，已知对偶问题：

$$\max_{a} \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n\, a_m t_n t_m x_n^T x_m \quad s.t. \quad a_n \geq 0,\ n = 1, \dots, N, \ \sum_{n=1}^{N} a_n\, t_n = 0 \qquad (5\text{-}2\text{-}4)$$

如果原问题和对偶问题等价，必须满足KKT条件：

$$a_n \geq 0, \quad t_n y(x_n) - 1 \geq 0, \quad a_n\{t_n y(x_n) - 1\} = 0 \text{（互补松弛条件）} \qquad (5\text{-}2\text{-}5)$$

由 $\frac{\partial L(w,b,a)}{\partial w} = w - \sum_{n=1}^{N} a_n\, t_n x_n = 0$ 和 $\frac{\partial L(w,b,a)}{\partial b} = \sum_{n=1}^{N} a_n\, t_n = 0$，可以得到超平面为：

$$y(x) = \sum_{n=1}^{N} a_n\, t_n x_n^T x + b, b = \frac{1}{N_S}\sum_{n\in S}(t_n - \sum_{m\in S} a_m\, t_m x_n^T x_m) \qquad (5\text{-}2\text{-}6)$$

线性支持向量机模型推导

● KKT条件分析

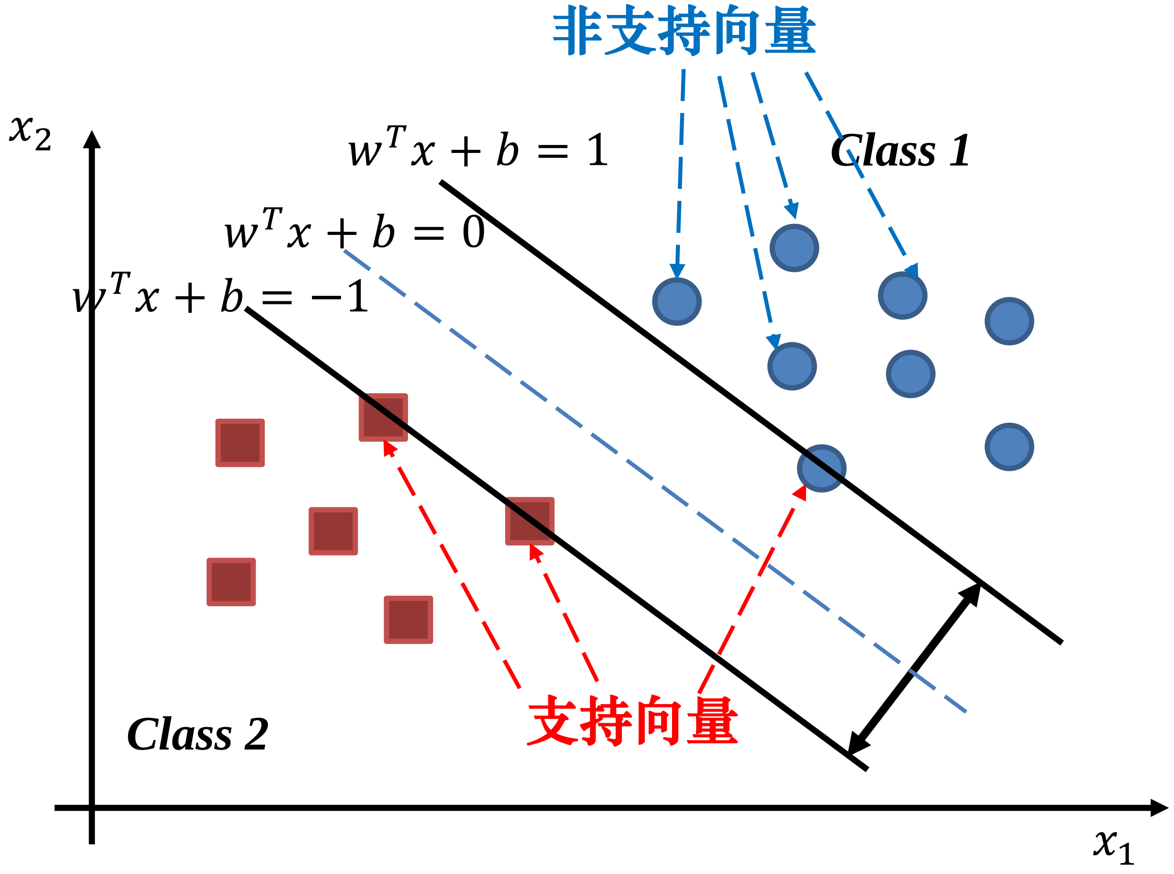
分类超平面 $y(x) = \sum_{n=1}^{N} a_n t_n x_n^T x + b$

互补松弛条件 $a_n\{t_n y(x_n) - 1\} = 0$

➢ 支持向量： $t_n(w^T x + b) = 1, a_n > 0$

非支持向量： $t_n(w^T x + b) > 1, a_n = 0$

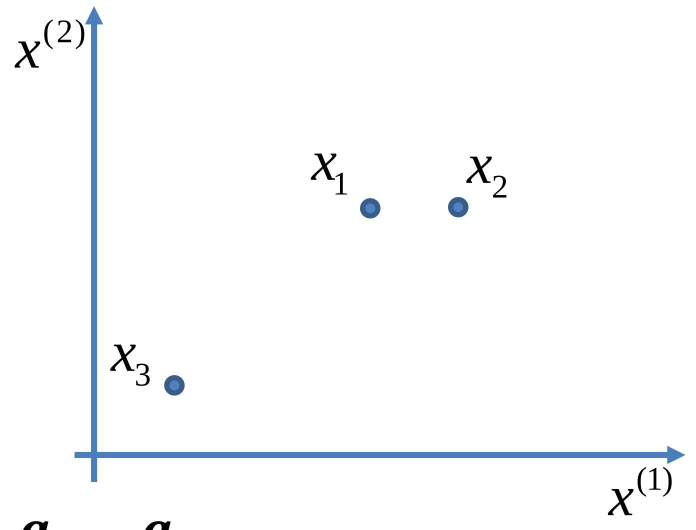
只有支持向量决定分类超平面，非支持向量与其无关



线性支持向量机的求解示例

示例

- 已知正样本点为 $x_1=(3,3)^T$，$x_2=(4,3)^T$，负样本点为 $x_3=(1,1)^T$，求线性可分支持向量机。



- 解：由公式 (5-2-4)，得到对偶问题为：

$$\min_{a}\frac{1}{2}(18a_1^2+25a_2^2+2a_3^2+42a_1a_2-12a_1a_3-14a_2a_3)-a_1-a_2-a_3$$

$$s.t.\quad a_1+a_2-a_3=0,\ a_i\geq 0, i=1,2,3$$

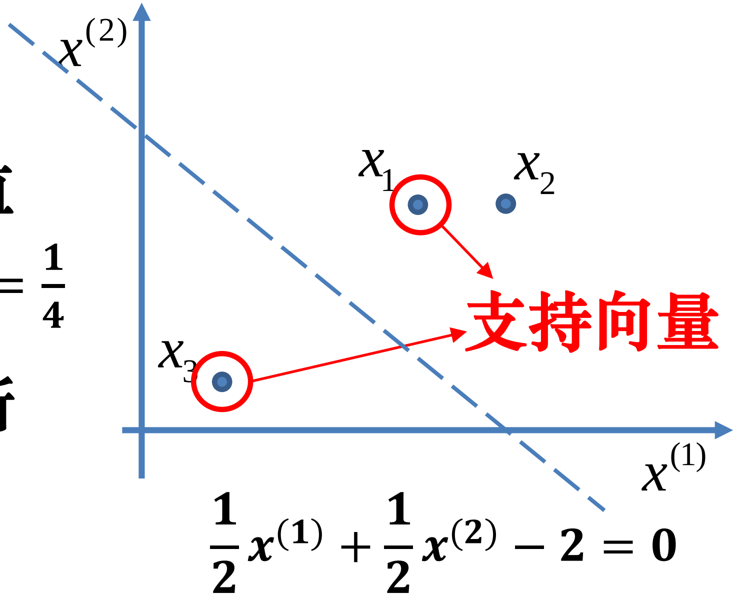
$a_3=a_1+a_2$带入对偶问题优化目标函数，记为$s(a_1,a_2)=4a_1^2+\frac{13}{2}a_2^2+10a_1a_2-2a_1-2a_2$

可知$s(a_1,a_2)$在$(\frac{3}{2},-1)^T$取极值，但此时$a_2=-1$不满足$a_2\geq 0$，则最小值只能在约束边界上取得。

——本例参考自 李航《统计学习方法》7.2节

线性支持向量机的求解举例

● 举例



则有，当$a_1=0$时，最小值 $s(0,\frac{2}{13})=-\frac{2}{13}$，当$a_2=0$时，最小值 $s\left(\frac{1}{4},0\right)=-\frac{1}{4}$，于是$s(a_1,a_2)$在$a_1=\frac{1}{4},a_2=0$时取最小值，则$a_3=\frac{1}{4}$

- 基于**KKT条件**，可知在 $a_n>0$时为支持向量，所以a_1和 a_3，所对应的实例点x_1和x_3是支持向量。

- 由公式 (5-2-3) 可求

$$w=\sum_{n=1}^{N}a_nt_nx_n=\frac{1}{4}\times 1\times(3,3)^T+0+\frac{1}{4}\times(-1)\times(1,1)^T=(\frac{1}{2},\frac{1}{2})$$

$$b=\frac{1}{N_S}\sum_{n\in S}(t_n-\sum_{m\in S}a_m\,t_mx_n^Tx_m)$$ ，则有 $b=-2$

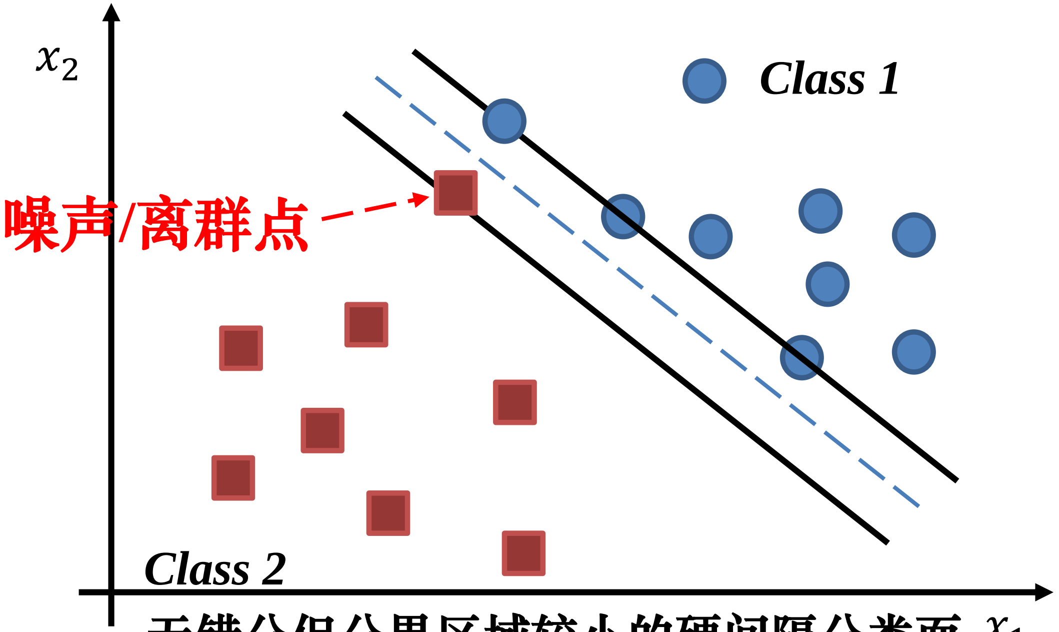
- 则分类超平面为： $\frac{1}{2}x^{(1)}+\frac{1}{2}x^{(2)}-2=0$

5.3 软间隔

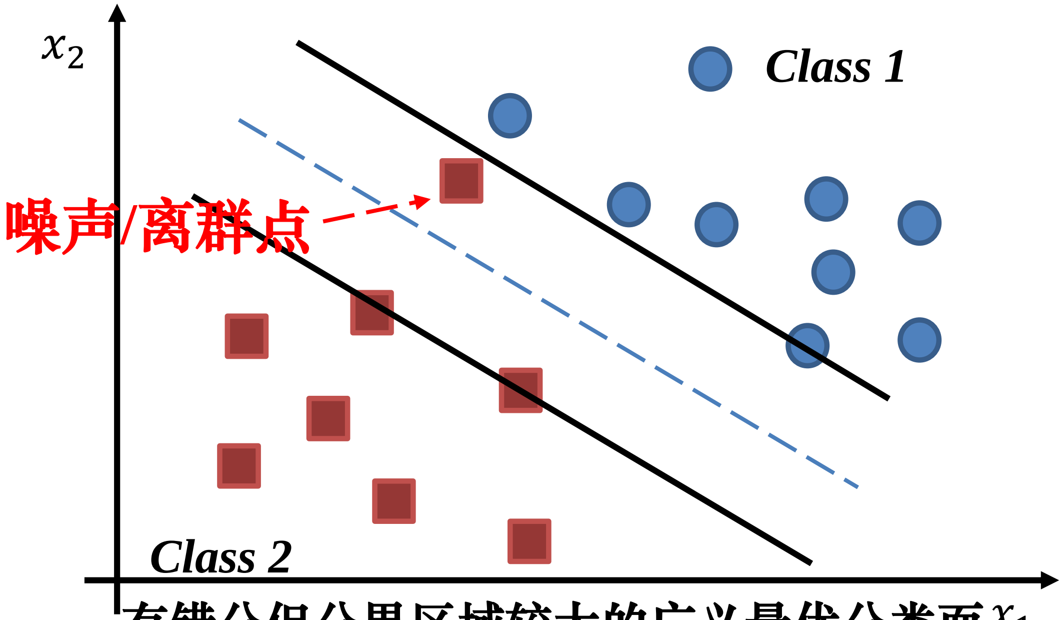
- 什么是软间隔
- 软间隔的求解

硬间隔的局限

- **硬间隔难以处理噪声和离群点**
  - 由于样本中不可避免地存在噪声和离群点，可能导致硬间隔求解的最优分类面**泛化性能差**。
  - 对于分布有交集的数据，需要有一定范围内的“错分”，又有较大分界区域的**广义最优分类面。**



无错分但分界区域较小的硬间隔分类面



有错分但分界区域较大的广义最优分类面

什么是软间隔

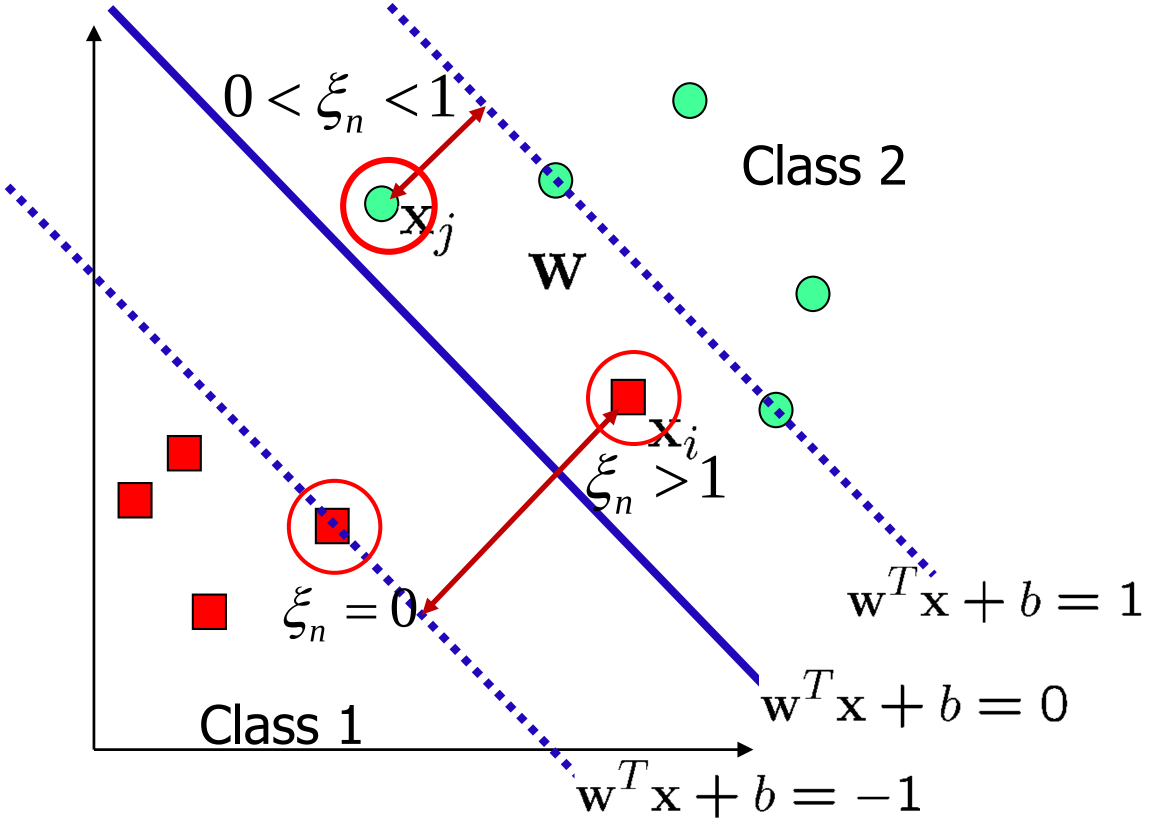
- 引入松弛变量处理噪声和离群点

➢ 引入松弛变量 $\xi_n \geq 0$

$\xi_n = 0$ 最大间隔边界

$\boxed{\xi_n = |t_n - y(x_n)|}$ ➡ $0 < \xi_n \leq 1$ 最大间隔内部且未被错分

$\xi_n > 1$ 被错分



原有不等式约束修改为

$t_n y(x_n) \geq 1$ ➡ $t_n y(x_n) \geq 1 - \xi_n$

➢ 这种处理方式也被视为是从硬间隔 (Hard Margin) 向**软间隔** (Soft Margin) 的转变。

什么是软间隔

- 基于软间隔的优化问题

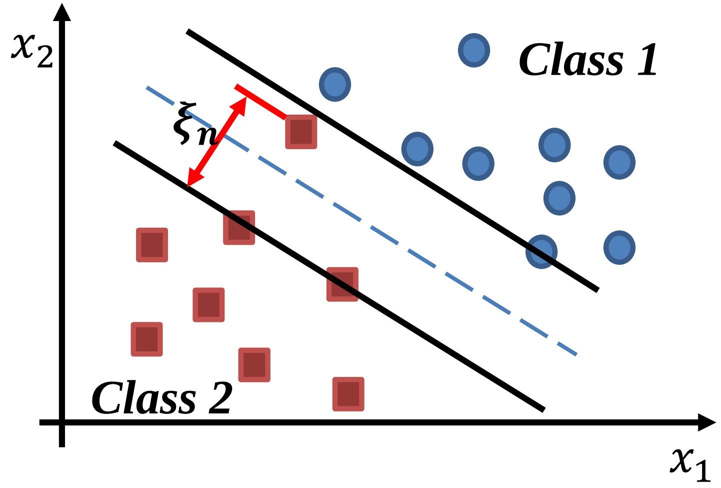
惩罚系数：反映对离群点容忍度　松弛变量

软间隔

$$\min_{w,b} \frac{1}{2}\|w\|^2 + \boxed{C}\boxed{\sum_{n=1}^{N} \xi_n}$$

$$s.t.\ t_n(w^T x_n + b) \geq 1 - \xi_n, n = 1, \ldots, N$$

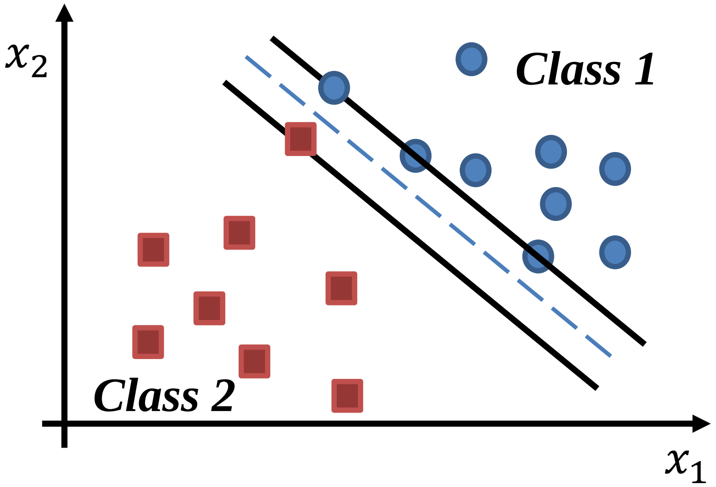
$$\xi_n \geq 0$$



硬间隔

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

$$s.t.\ t_n(w^T x_n + b) \geq 1, n = 1, \ldots, N$$



软间隔虽然存在少量错分，但带来更大的间隔，具有更好的泛化性能

软间隔的求解

- 利用拉格朗日乘子法进行软间隔求解
  - 优化目标： $$\min_{w,b} \frac{1}{2}\|w\|^2 + \boxed{C\sum_{n=1}^{N}\xi_n}$$
    $$s.t.\ t_n\left(w^T x_n + b\right) \geq 1 - \boxed{\xi_n}\ \xi_n \geq 0, n = 1, \ldots, N$$
  - 拉格朗日函数：
    $$L(w,b,\xi,a,\mu) = \frac{1}{2}\|w\|^2 + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N} a_n\{t_n y(x_n) - 1 + \xi_n\} - \boxed{\sum_{n=1}^{N}\mu_n\xi_n}$$
    $$a_n \geq 0\,; \boxed{\mu_n \geq 0}$$
  - 优化$w, b, \{\xi_n\}$：
    $$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{n=1}^{N} a_n t_n x_n$$
    $$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{n=1}^{N} a_n t_n = 0$$
    $$\frac{\partial L}{\partial \xi_n} = 0 \Rightarrow \boxed{a_n = C - \mu_n}$$

相比硬间隔添加了松弛变量和对应的惩罚系数

松弛变量引入约束条件

由于$\xi_n \geq 0$，松弛变量添加为拉格朗日函数的一项，$\mu_n \geq 0$ 是对应的拉格朗日乘子

松弛变量作为一个变量求对应偏导，偏导为0，此时有 $a_n + \mu_n = C$

软间隔的求解

- 利用拉格朗日乘子法进行软间隔求解

  - 带入拉格朗日函数，得到对偶问题：

$$\max_{a} \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m x_n^T x_m$$

$$s.t.\ \boxed{0 \le a_n \le C}, n = 1, 2, \ldots, N$$

$$\sum_{n=1}^{N} a_n t_n = 0$$

与硬间隔对偶问题相比，只有约束条件不同，硬间隔中要求$a_n \ge 0$

  - KKT条件：

$$a_n \ge 0, \mu_n \ge 0, \xi_n \ge 0$$

$$t_n y(x_n) - 1 + \xi_n \ge 0$$

互补松弛条件：

$$\boxed{a_n\{t_n y(x_n) - 1 + \xi_n\} = 0}$$

$$\boxed{\mu_n \xi_n = 0}$$

约束 $t_n(w^T x_n + b) \ge 1 - \xi_n$的互补松弛条件

约束 $\xi_n \ge 0$ 的互补松弛条件

软间隔的求解

● KKT互补松弛条件分析

$a_n = 0$ 样本x_n不受约束，与分类超平面无关，是非支持向量

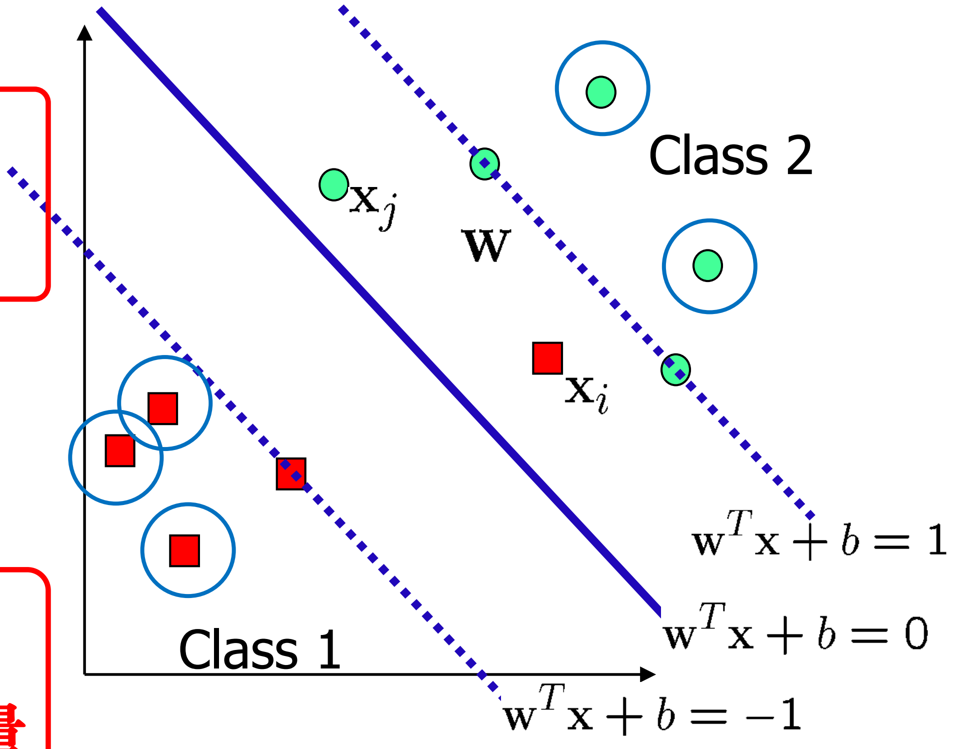
$$a_n\{t_n y(x_n) - 1 + \xi_n\} = 0$$
$$\mu_n \xi_n = 0$$

$a_n \geq 0$

$a_n > 0$

$t_n y(x_n) - 1 + \xi_n = 0$ ➡ 样本x_n是支持向量

$\mu_n \xi_n = 0$



软间隔的求解

● **KKT互补松弛条件分析**

$a_n > 0$

$t_n y(x_n) - 1 + \xi_n = 0$

$\mu_n \xi_n = 0$

$0 < a_n \le C,$

$a_n + \mu_n = C$

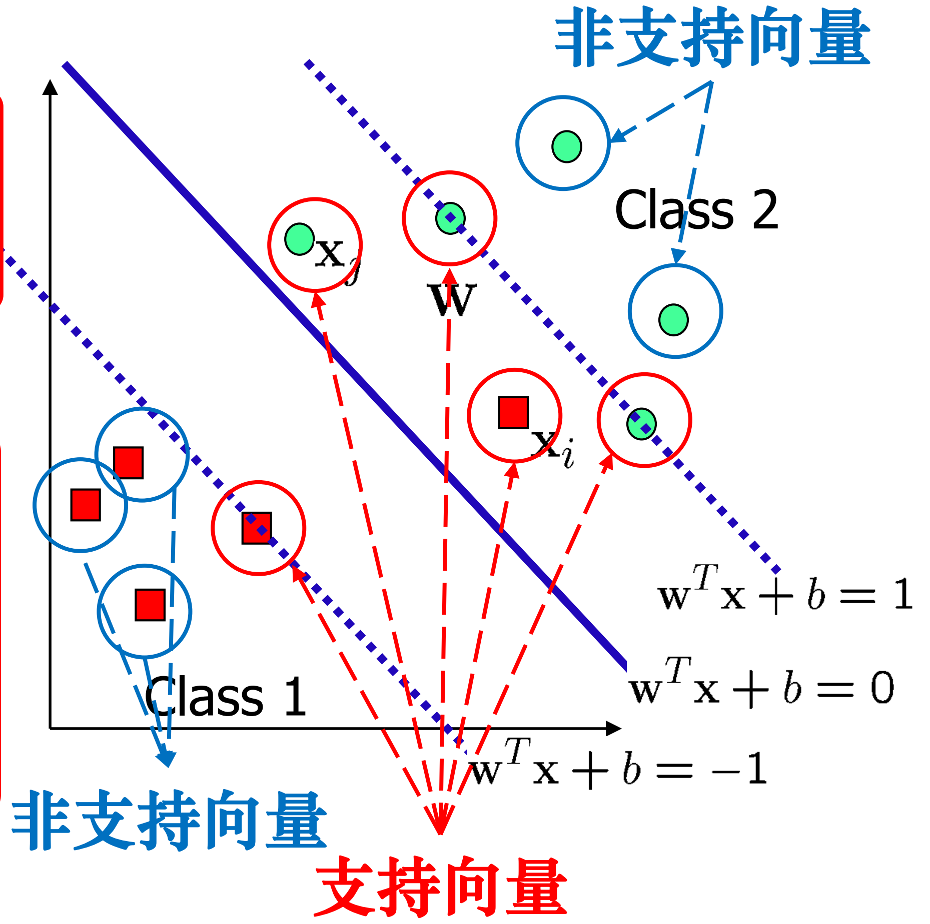
$a_n < C, \mu_n > 0$

$\xi_n = 0$ 样本x_n在最大分隔边界上

$a_n = C, \mu_n = 0, \mu_n$ 不再约束 ξ_n

$\xi_n \le 1$ 样本x_n在最大分隔内且被未错分

$\xi_n > 1$ 样本x_n被错分

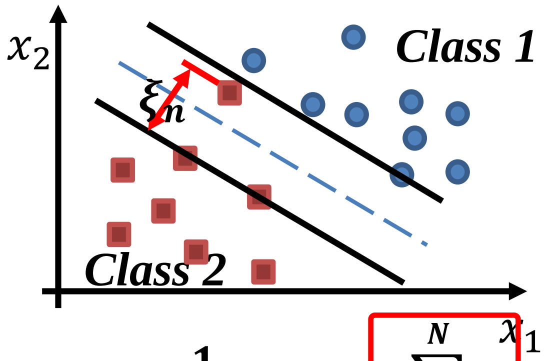


软间隔支持向量机的最终超平面也仅与支持向量有关

总结

● 总结

软间隔



允许对噪声/离群点有一定错分，最大间隔更大，**泛化性更好**

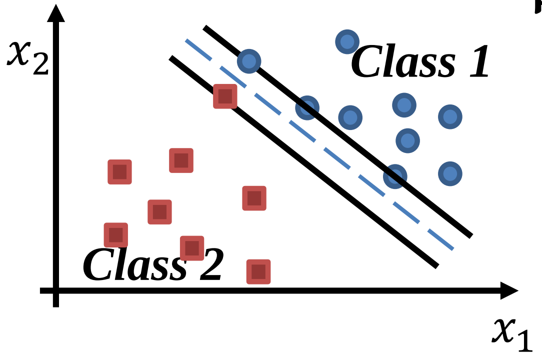
$$\arg\min_{w,b} \frac{1}{2}\|w\|^2 + \boxed{C\sum_{n=1}^{N}\xi_n}$$

$$\text{s.t. } t_n\left(w^T x_n + b\right) \geq 1-\xi_n,$$
$$n = 1,\dots,N, \xi_n \geq 0$$

求解时有超参惩罚系数，求解比硬间隔复杂

软间隔是权衡泛化性和准确性的综合选择

硬间隔



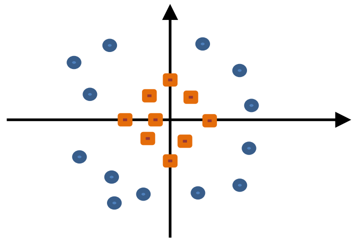
不允许错分，**准确性高**，最大间隔小，**泛化性差**

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

$$\text{s.t. } t_n\left(w^T x_n + b\right) \geq 1,$$
$$n = 1,\dots,N$$

求解过程较简单

➢思考新问题：介绍了使用硬间隔SVM和软间隔SVM求解线性分类问题，**SVM是否可以解决非线性问题?**



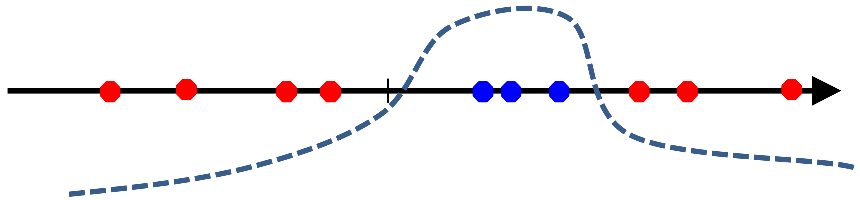
5.4 非线性支持向量机

- 非线性支持向量机原理
- 核函数原理

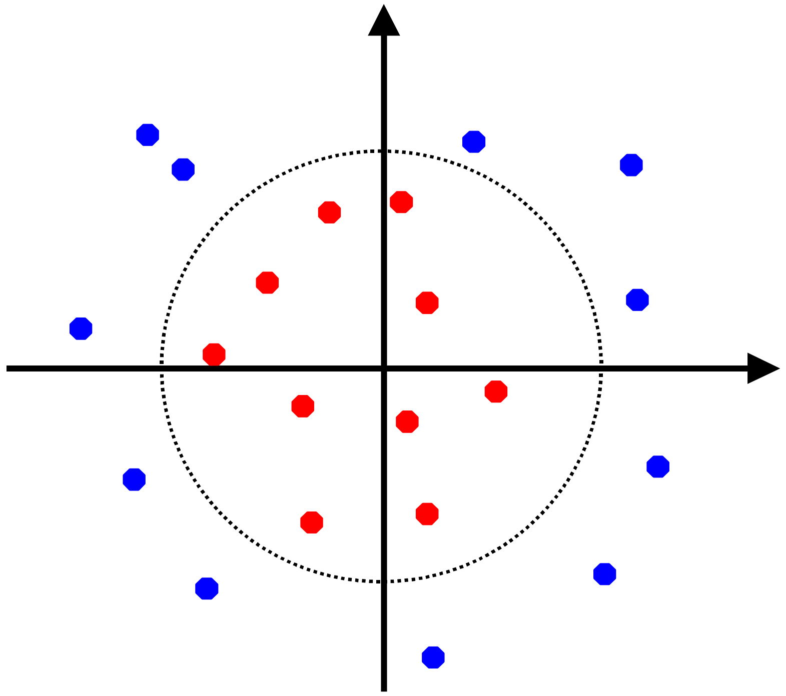
非线性支持向量机原理

- 支持向量机如何解决非线性问题

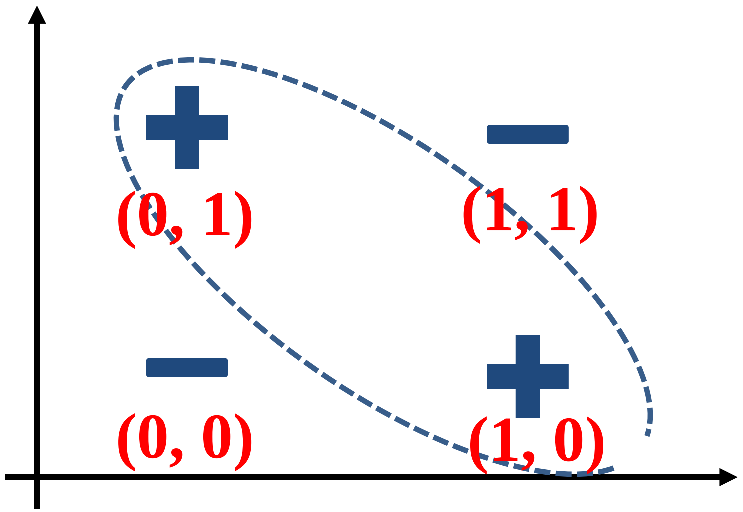
(1) 一维线性不可分问题



(2) 二维线性不可分问题



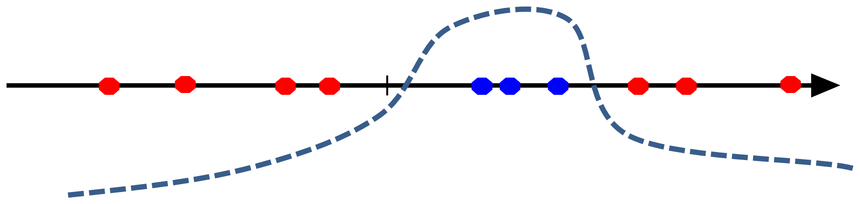
(3) 异或问题



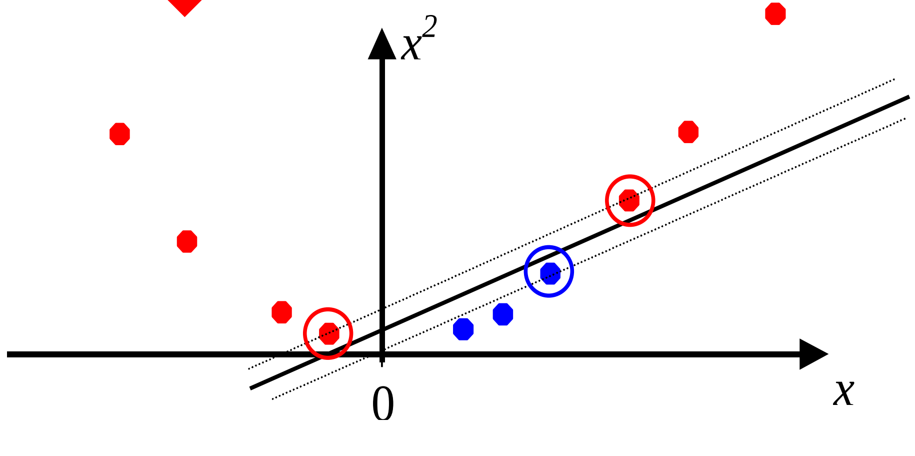
非线性支持向量机原理

非线性SVM的基本思想

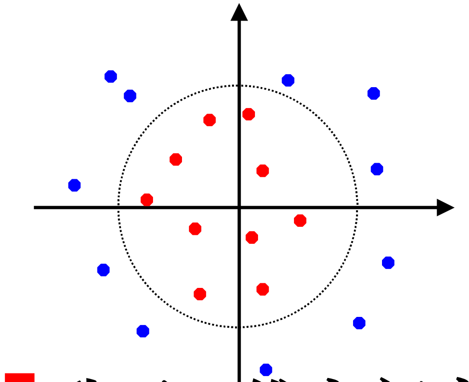
(1) 一维线性不可分问题



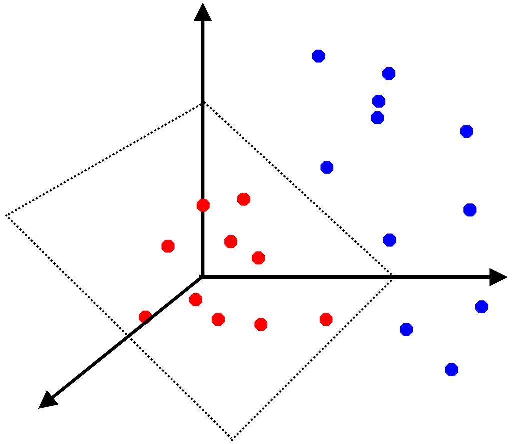
$\Phi : x \rightarrow \phi(x)$ 升至二维空间实现线性可分



(2) 二维线性不可分问题



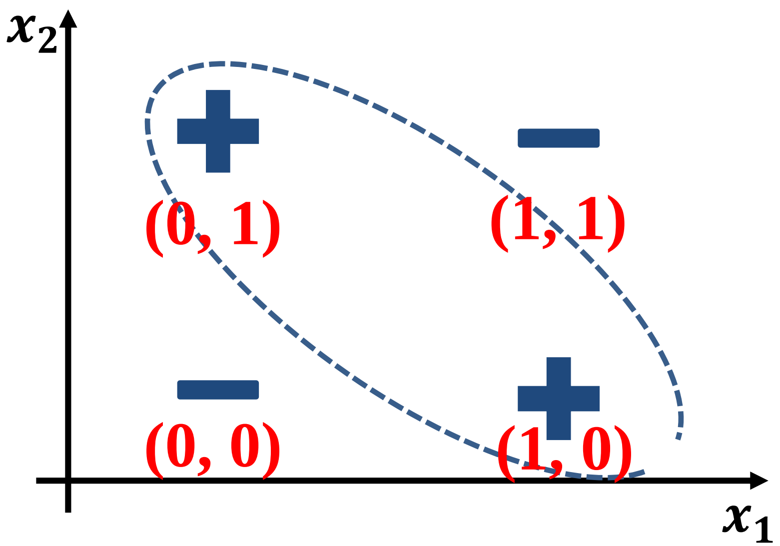
$\Phi : x \rightarrow \phi(x)$ 升至三维空间实现线性可分



非线性支持向量机原理

● 非线性SVM的基本思想

(3) 异或问题

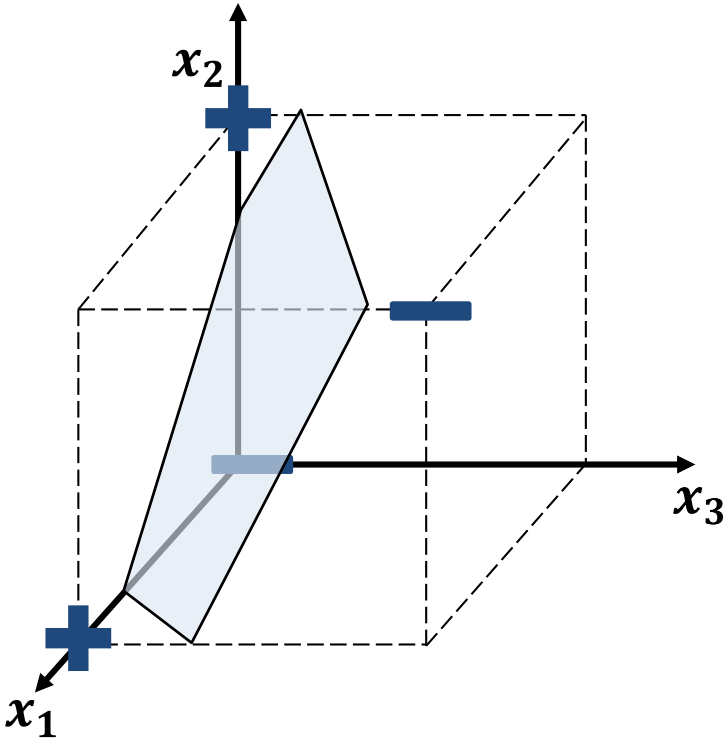


$$\boldsymbol{\Phi}: \boldsymbol{x} \rightarrow \boldsymbol{\phi}(\boldsymbol{x})$$

$$\boldsymbol{\phi}(\boldsymbol{x}) = (x_1, x_2, x_1 x_2)$$

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

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



非线性支持向量机原理

- **高维映射引发维度灾难**

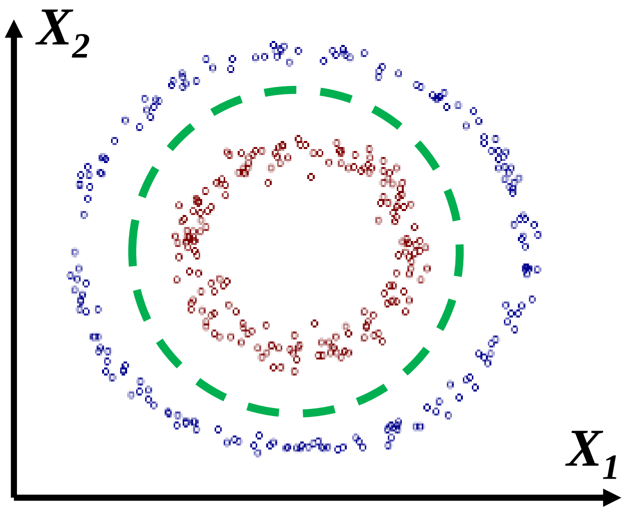
$$a_1X_1 + a_2X_1^2 + a_3X_2 + a_4X_2^2 + a_5X_1X_2 + a_6 = 0$$

一种映射：

$$Z_1 = X_1 ; Z_2 = X_1^2 ; Z_3 = X_2 ; Z_4 = X_2^2 ; Z_5 = X_1X_2$$

$$\Rightarrow \sum_{i=1}^{5} a_iZ_i + a_6 = 0 \qquad \phi : \mathbb{R}^2 \to \mathbb{R}^5$$

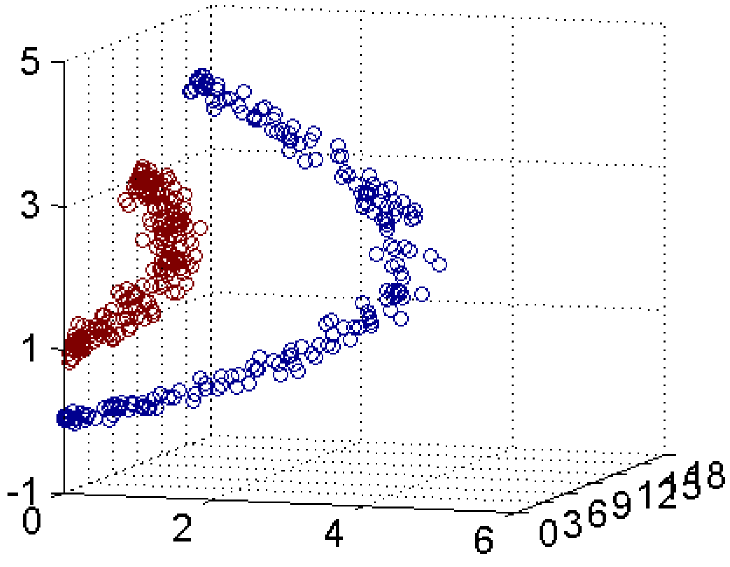
另一种映射：



$$a_1X_1^2 + a_2(X_2 - c)^2 + a_3 = 0$$

$$Z_1 = X_1^2 ; Z_2 = X_2^2 ; Z_3 = X_2$$

$$\phi : \mathbb{R}^2 \to \mathbb{R}^3$$



非线性支持向量机原理

- 高维映射引发维度灾难

原始样本分布：$a_1X_1 + a_2X_1^2 + a_3X_2 + a_4X_2^2 + a_5X_1X_2 + a_6 = 0$

$$\sum_{i=1}^{5} a_iZ_i + a_6 = 0 \qquad \phi: \boxed{\mathbb{R}^2 \to \mathbb{R}^5}$$

原始样本增到三维：

$$a_1X_1^3 + a_2X_2^3 + a_3X_3^3 + a_4X_1^2X_2 + a_5X_1^2X_3 + a_6X_2^2X_1 + a_7X_2^2X_3 + a_8X_3^2X_1$$
$$+a_9X_3^2X_2 + a_{10}X_1X_2X_3 + a_{11}X_1^2 + a_{12}X_2^2 + a_{13}X_3^2 + a_{14}X_1X_2 + a_{15}X_2X_3$$
$$+ a_{16}X_1X_3 + a_{17}X_1 + a_{18}X_2 + a_{19}X_3 + a_{20} = 0$$

$$\Rightarrow \sum_{i=1}^{19} a_iZ_i + a_{20} = 0 \quad \phi: \boxed{\mathbb{R}^3 \to \mathbb{R}^{19}} \Rightarrow$$

维数大大增加
计算变得非常困难

非线性SVM是如何解决高维映射引发的维度灾难?

核函数原理

非线性支持向量机

- 设样本集$\{x_n, t_n\}$，样本x_n映射后的向量为$\phi(x_n)$, 分类超平面为$y(x) = w^T\phi(x) + b$
- 优化目标为：

$$\min_{w,b} \frac{1}{2}||w||^2$$
$$s.t. \quad t_n\left(w^T\phi(x_n) + b\right) \geq 1, \quad n = 1, 2, \ldots, N.$$

- 对偶问题为：

$$\max_{a} \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m \phi(x_n)^T\phi(x_m)$$
$$s.t. \sum_{n=1}^{N} a_n t_n = 0, \quad a_n \geq 0, \quad i = 1, 2, \ldots, N.$$

- 分类超平面为：$y(x) = w^T\phi(x) + b = \sum_{n=1}^{N} a_n t_n \boxed{\phi(x_n)^T \phi(x)} + b$

分类超平面包含高维映射的内积

核函数原理

非线性支持向量机

- 分类超平面为：

$$y(x) = w^T\phi(x) + b = \sum_{n=1}^{N} a_n t_n \boxed{\phi(x_n)^T \phi(x)} + b$$

构造如下函数：

$$k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle = \phi(x_i)^T \phi(x_j)$$

此时，分类超平面表示为：

$$y(x) = w^T\phi(x) + b = \sum_{n=1}^{N} a_n t_n k(x, x_n) + b$$

这里 $k(x, x_n)$就是**核函数，**至此，在处理高维的非线性问题时，利用核函数直接在原来的低维空间中进行计算不需要显式地写出映射后的结果, 避免了先映射到高维空间中然后再根据内积的公式进行计算

核函数原理

常见核函数

- 根据问题和数据的不同，选择不同的核函数。一些常用的核函数有：

线性核： $k(x_1, x_2) = x_1^T x_2$

多项式核： $k(x_1, x_2) = (< x_1, x_2 > +R)^d$

高斯核： $k(x_1, x_2) = \exp\left\{-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right\}$

Sigmoid核： $k(x_1, x_2) = \tanh\left(\beta_0 x_1^T x_2 + \beta_1\right)$

如何判断一个函数是否可以作为核函数?

Mercer定理：
$\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ 上的映射k是一个有效核函数(也称Mercer核函数)，当且仅当对于训练样本其相应的核函数矩阵是对称半正定的,即对于任何平方可积函数 $g(x)$有

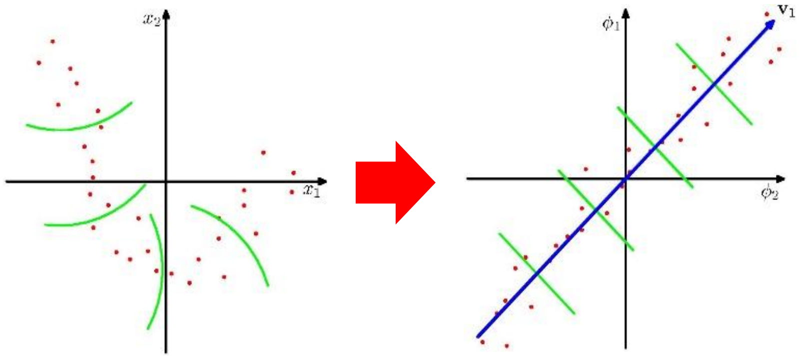
$$\iint k(x, y) g(x) g(y) dx dy \geq 0$$

核技巧

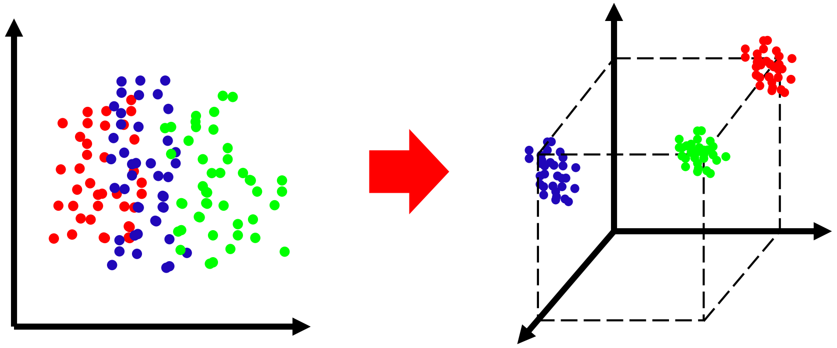
● 核技巧的应用

➢ 在SVM解决非线问题中，在不显式定义高维映射$\phi(x)$的情况下，这种通过设计核函数来隐式地在特征空间进行计算的技巧称为**核技巧**。这种核技巧不仅应用于非线性支持向量机中，之后课程中的主成分分析和聚类方法也使用核技巧来解决非线性问题中的维度灾难。

主成分分析



聚类

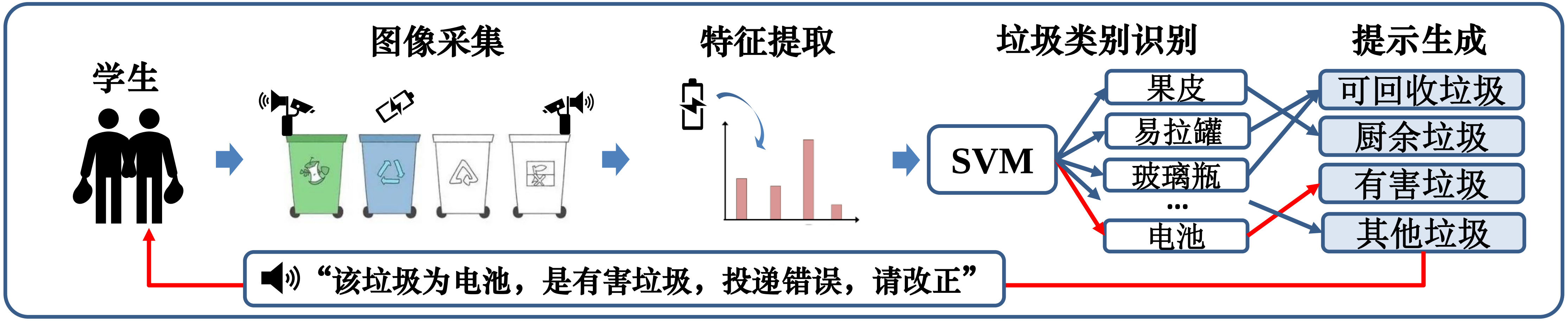


支持向量机应用

● 使用SVM实现自动垃圾分类

➢ 应用背景：我国自2019年起在多个城市强制实施垃圾分类，垃圾分类已成为城市管理的重要组成部分。但在实际生活中，许多居民和游客仍然难以正确分类垃圾，不仅影响资源的回收利用，还增加了垃圾处理难度。某高校学生通过在垃圾桶上安装摄像头，实现在同学们投放垃圾时给予提醒。由于只获取了少量的垃圾样本，所以采用在样本受限条件下依旧性能较好的SVM来构建一个垃圾类别提示系统

➢ 垃圾类别提示系统



支持向量机应用

使用SVM实现自动垃圾分类

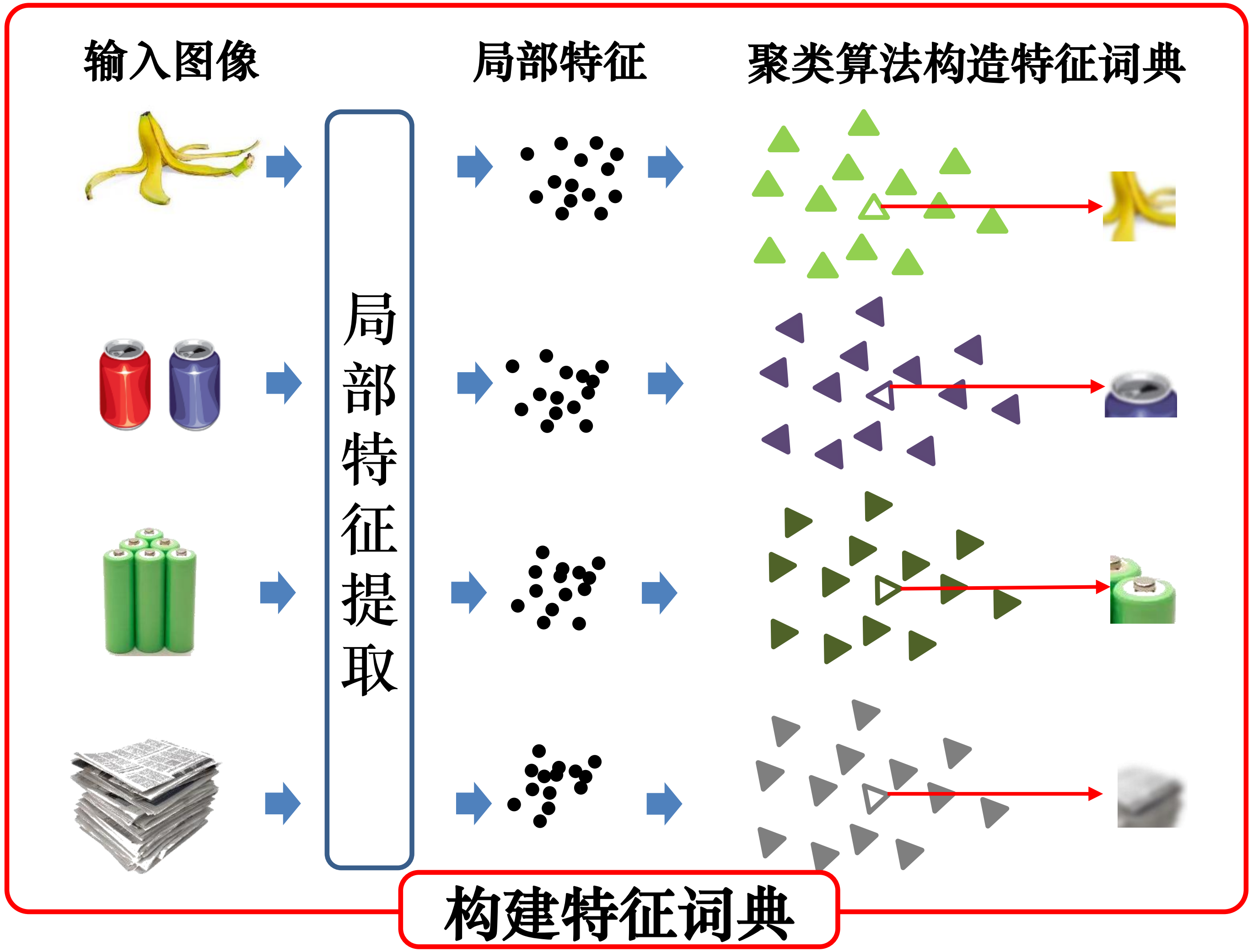
- **特征提取算法选择:**

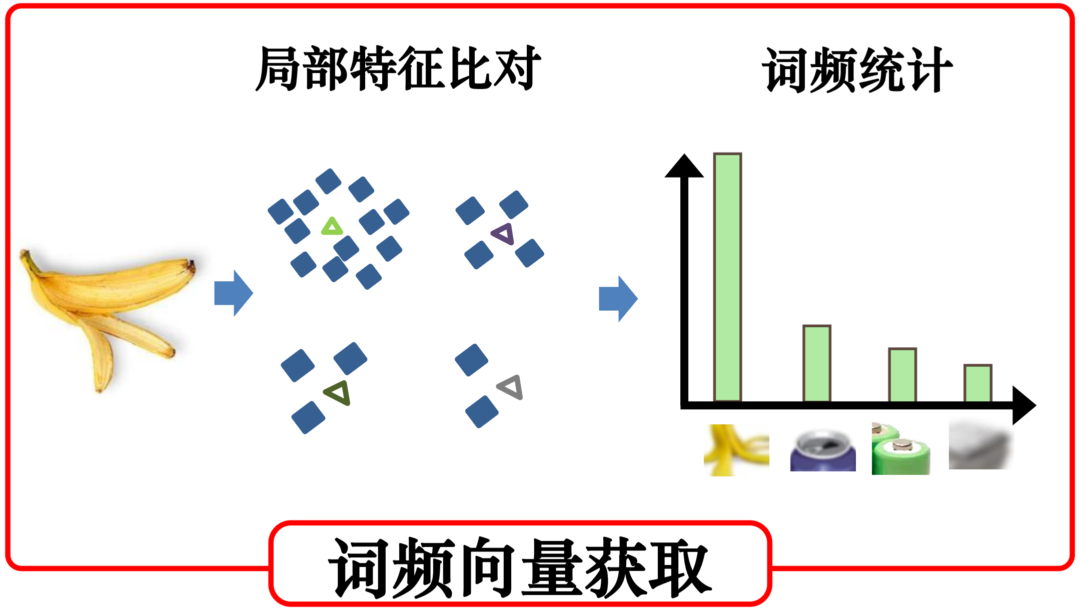
Bag of Words (BOW) 是一种经典的特征提取方法，通过统计数据中各类特征（例如图像中的局部特征）的出现频率，将图像转换为特征向量。由于垃圾图像表现出类间差异大、类内差异小的特点，使用BOW可有效捕捉同类垃圾的特征，并将其转化为可用于分类的词频向量。此外，BOW方法受垃圾图像中的位置、大小和旋转等因素影响较小，提取的垃圾图像特征更加鲁棒。

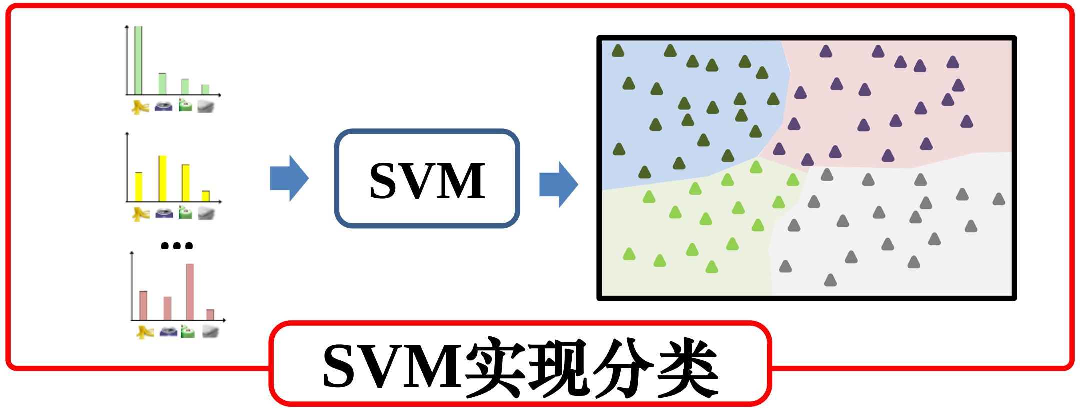
Scale-Invariant Feature Transform (SIFT)： SIFT通过在高斯差分空间检测关键点，并利用梯度方向直方图对关键点进行方向归一化，所提取的局部特征具有尺度和旋转不变性，适合提取多角度、多尺度的垃圾图像。同时SIFT对噪声具有较强的抗干扰能力，适合复杂环境条件下的垃圾图像特征提取。

支持向量机应用

- 使用SVM实现自动垃圾分类







5.5 序列最小优化算法

- 什么是序列最小优化算法?
- 序列最小优化算法

什么是序列最小优化算法?

- 支持向量机的学习问题可以形式化为求解具有全局最优解的凸二次规划问题。许多方法可以用于求解这一问题，但当训练样本容量很大时，这些算法往往效率较低，以致无法使用。

【1999年J. C. Platt提出】

- 序列最小优化算法（ Sequential Minimal Optimization, SMO ）
  - 启发式算法
  - 先前可用的SVM使用复杂的训练方法。SMO算法较好地避免了这一问题，用于解决支持向量机训练目标函数优化问题（凸二次规划问题）的算法

什么是序列最小优化算法?

- 序列最小优化算法（ Sequential Minimal Optimization, SMO ）
  - 基本假设：如果所有变量的解都满足优化问题的KKT条件，那么优化问题的解就得到了，因为KKT条件是该最优化问题有解的充要条件。
  - 基本思想：不断地将原二次规划问题分解为只有两个变量的二次规划问题，并对子问题进行解析求解，直到所有变量都满足KKT条件为止。因为子问题解析解存在，所以每次计算子问题都很快，虽然子问题次数很多，但是总体上还是高效的。

序列最小优化算法

● SMO算法解凸二次规划问题

➢ 优化目标：

$$\min_{a} \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m t_n t_m k(x_m, x_n) - \sum_{n=1}^{N} a_n$$

$$\text{s.t.} \quad 0 \le a_n \le C, n = 1, 2, \cdots, N, \qquad \sum_{n=1}^{N} a_n t_n = 0$$

➢ SMO是解决支持向量机优化问题的迭代算法。由于目标函数为凸函数，一般的优化算法都通过梯度方法一次优化一个变量求解二次规划问题的最大值，但由于限制条件$\sum_{n=1}^{N} a_n t_n = 0$，当某个$a_{\mathrm{i}}$从$a_i^{old}$更新至$a_i^{new}$时等式约束不成立，上述限制条件被打破。为了满足约束条件，SMO采用一次更新两个变量的方法。

序列最小优化算法

SMO算法解凸二次规划问题

- 在每个迭代步骤中，算法首先选取两个待更新的向量a_1和a_2，分别计算它们的误差项，并求解出a_1^{new}, a_2^{new}。最后再根据SVM的定义计算偏移量b。对于误差项而言，可以根据a_1^{new}, a_2^{new}, b的增量进行调整，无需每次重新计算。

SMO算法流程：

步骤1：随机数初始化向量权重a_i，计算偏移量b

步骤2：初始化误差项E_i

步骤3：选取两个向量a_1和a_2作为待更新向量

步骤4：求解a_1^{new}, a_2^{new}

步骤5：利用求解的a_1^{new}, a_2^{new}更新E_i和b

步骤6：重复步骤3-5，直至达到终止条件

选择变量的启发式方法

两个变量二次规划的解析方法

序列最小优化算法

- **步骤4. 两个变量二次规划的解析方法**

  - 不失一般性，对**两个变量a_1和a_2**进行更新，固定其他变量$a_i (i = 3, 4, \cdots, N)$，原最优化问题可以表示为：

$$\min_{a_1, a_2} W(a_1, a_2) = \frac{1}{2} K_{11} a_1^2 + \frac{1}{2} K_{22} a_2^2 + t_1 t_2 K_{12} a_1 a_2 - (a_1 + a_2)$$

$$+ t_1 a_1 \sum_{n=3}^{N} t_n a_n K_{n1} + t_2 a_2 \sum_{n=3}^{N} t_n a_n K_{n2} + const$$

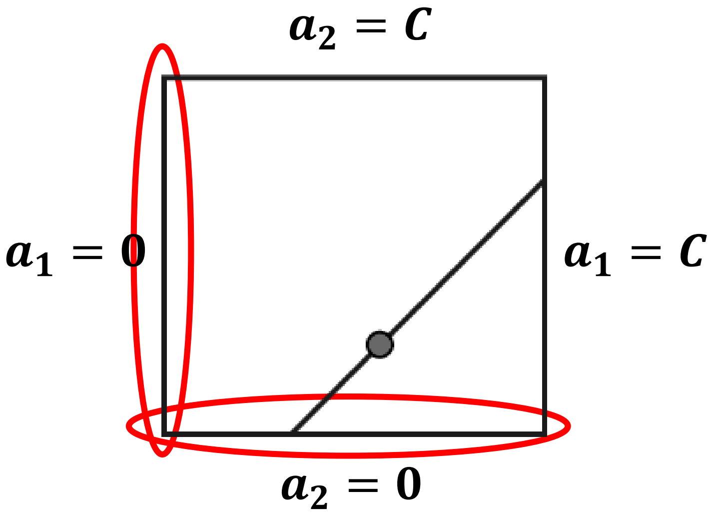
$$\text{s.t.} \quad a_1 t_1 + a_2 t_2 = -\sum_{n=3}^{N} a_n t_n = \zeta, \quad 0 \le a_i \le C, i = 1, 2$$

其中核函数$K_{mn} = K(x_m, x_n), m, n = 1, 2, \cdots, N, \zeta$是常数，上式中省略了不含$a_1$和$a_2$的常数项。

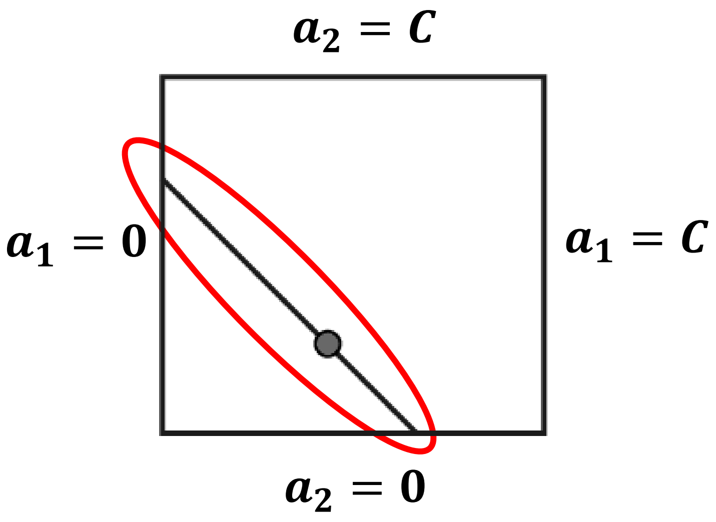
序列最小优化算法

- **两个变量二次规划的解析方法**
  - 两个变量(a_1, a_2)的约束条件可用二维空间的图形表示

$$a_1 t_1 + a_2 t_2 = -\sum_{n=3}^{N} a_n t_n = \zeta, \quad 0 \le a_i \le C, i = 1,2$$



$t_1 \ne t_2 \Rightarrow a_1 - a_2 = k$



$t_1 = t_2 \Rightarrow a_1 + a_2 = k$

目标函数在一条平行于对角线的线段上取得最优解，**两个变量的最优化问题转化为单变量最优化问题。**

序列最小优化算法

- **两个变量二次规划的解析方法**
  - 考虑**变量为a_2的最优化问题**，假设问题的初始可行解为a_1^{old}, a_2^{old}，最优解为a_1^{new}, a_2^{new}，并且在沿着约束方向未经剪辑(未考虑不等式约束$0 \le a_n \le C, n = 1, 2$)时$a_2$的最优解为$a_2^{new,unc}$。
  - 由于a_2^{new}需要满足不等式约束，所以其取值范围须满足$L \le a_n^{new} \le H$, 其中，L与H是a_2^{new}所在的对角线段端点的界。

如果$t_1 \neq t_2$，则$L = \max\left(0, a_2^{old} - a_1^{old}\right)$，$H = \min(C, C + a_2^{old} - a_1^{old})$

如果$t_1 = t_2$，则$L = \max\left(0, a_2^{old} + a_1^{old} - C\right)$，$H = \min(C, a_2^{old} + a_1^{old})$

序列最小优化算法

- 两个变量二次规划的解析方法
  - 记$g(x) = \sum_{n=1}^{N} a_n y_n K(x_n, x) + b$，描述输入变量$x_n$的预测值;
    令$E_n = g(x_n) - y_n = (\sum_{m=1}^{N} a_m y_m K(x_m, x_n) + b) - y_n$， $n = 1, 2$
    即$n = 1, 2$时， E_n为函数$g(x)$对输入x_n的预测值与真实输出y_n之差。
  - 那么未经剪辑的a_2 的解$a_2^{new,unc} = a_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$，
    其中$\eta = K_{11} + K_{22} - 2K_{12} = \| \phi(x_1) - \phi(x_2) \|^2$。
  - 经剪辑后a_2 的解是$a_2^{new} = \begin{cases} H, & a_2^{new,unc} > H \\ a_2^{new,unc}, & L \le a_2^{new,unc} \le H \\ L, & a_2^{new,unc} < L \end{cases}$
  - 由a_2^{new}求得$a_1^{new} = a_1^{old} + y_1 y_2 (a_2^{old} - a_2^{new})$

序列最小优化算法

步骤3. 变量的选择方法

- (1)第一个变量a_1的选择：外层循环

外层循环在训练样本中选取违反KKT条件最严重的样本点，并将其对应的变量作为第一个变量。检验是否满足KKT条件：

$$a_i = 0 \Leftrightarrow y_i g(x_i) \geq 1$$
$$0 < a_i < C \Leftrightarrow y_i g(x_i) = 1$$
$$a_i = C \Leftrightarrow y_i g(x_i) \leq 1$$

外循环首先遍历满足条件$0 < a_i < C$的样本点，就是在边界上的支持向量点，如果都满足KKT条件，再遍历整个训练集，从整个数据集中选取违反KKT条件最严重的点。

序列最小优化算法

- **变量的选择方法**

➢ (2)第二个变量a_2的选择：内层循环

假设在外层循环已经找到第一个变量a_1，那么内层循环要找到第二个变量a_2，希望使a_2有足够大的变化。

$$a_2^{new}=\begin{cases}H, & a_2^{new,unc}>H\\ a_2^{new,unc}, & L\le a_2^{new,unc}\le H\\ L, & a_2^{new,unc}<L\end{cases}\qquad a_2^{new,unc}=a_2^{old}+\frac{y_2(E_1-E_2)}{\eta}$$

a_2^{new}是依赖于的$|E_1-E_2|$的，简单做法是选择a_2使$|E_1-E_2|$最大：a_1已定，E_1确定，如果E_1为正，选择最小的E_i作为E_2；如果E_1为负，选择最大的E_i作为E_2。

序列最小优化算法

- **变量的选择方法**

  - **(2)第二个变量a_2的选择：内层循环**

    如果通过$|E_1-E_2|$最大化的方法无法使目标函数有足够的变化，那就遍历间隔边界上的支持向量点，依次将其对应的乘子作为 a_2 ，直到目标函数有足够的下降。

    如果间隔边界上的支持向量仍未使得目标函数有足够的下降，那么遍历整个数据集，如果数据集还未使得目标函数有足够的下降，则更换a_1，重新找到新的a_1进行优化。

序列最小优化算法

- **步骤5. 更新b和差值E_{i}**
  - 在每次完成两个变量的优化后，都要重新计算偏移量b。当$0 < a_1^{new}, a_2^{new} < C$时，由KKT条件$0 < a_i < C \Leftrightarrow y_i g(x_i) = 1$可计算：

$$b_1^{new} = -E_1 - y_1 K_{11}\left(a_1^{new} - a_1^{old}\right) - y_2 K_{21}\left(a_2^{new} - a_2^{old}\right) + b^{\mathrm{old}}$$
$$b_2^{new} = -E_2 - y_1 K_{12}\left(a_1^{new} - a_1^{old}\right) - y_2 K_{22}\left(a_2^{new} - a_2^{old}\right) + b^{\mathrm{old}}$$

  - 如果a_1^{new}, a_2^{new}同时满足条件$0 < a_i^{new} < C, \mathrm{i} = 1, 2$，那么$b_1^{new} = b_2^{new}$。如果$a_1^{new}, a_2^{new}$是0或者$C$，那么$b_1^{new}, b_2^{new}$以及他们之间的数都是符合KKT条件的阈值，这时取它们的中点作为b^{new}，之后更新E_{i}。

$$E_i = \sum_S y_j a_j K\left(x_i, x_j\right) + b^{new} - y_i$$

序列最小优化算法

SMO算法

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

输出：近似解$\hat{a}$。

（1）取初值 $a^{(0)}=0$，令 $k=0$；

（2）选取优化变量a_1^k,a_2^k，解析求解两个变量的最优化问题，求得最优解a_1^{k+1},a_2^{k+1}，更新a 为 a^{k+1} 。

（3）若在精度ϵ范围内满足停止条件$\sum_{i=1}^{N}a_iy_i=0\,,0\le a_i\le C,i=1,2,\cdots,N$，其中$g(x_i)=\sum_{j=1}^{N}a_jy_jK(x_j,x_i)+b$，$y_ig(x_i)=\begin{cases}\ge 1, & \{x_i|a_i=0\}\\ =1, & \{x_i|0<a_i<C\}\\ \le 1, & \{x_i|a_i=C\}\end{cases}$，则转至（4），否则令$k=k+1$，转至（2）；

（4）取$\hat{a}=a^{(k+1)}$。

附录：线性支持向量机模型推导

● 对偶问题和KKT条件

➢ 为了不失一般性，设存在不等式约束的函数极值问题，优化目标为：

$$\min_{w} f(w)$$

$$s.t.\ g_i(w) \le 0, i = 1, \dots, k.$$

$$h_i(w) = 0, i = 1, \dots, l.$$

求取拉格朗日函数：

$$L(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

定义 $\theta_P(w) = \max_{\alpha,\beta\,;\,\alpha_i \ge 0} L(w, \alpha, \beta)$ ，当违背约束条件时，$g_i(w) > 0$ ，$h_i(w) \ne 0$ ，$L(w, \alpha, \beta)$ 的最大值为无穷大

附录：线性支持向量机模型推导

● 对偶问题和KKT条件

当满足约束条件时，有 $g_i(w) \le 0, h_i(w) = 0$，此时 $L(w,\alpha,\beta)$ 有最大值，即 $f(w)$，则有：

$$\theta_P(w) = \begin{cases} f(w), \text{如果 } w \text{ 满足约束条件} \\ +\infty, \text{否则} \end{cases}$$

则原问题 $\min\limits_{w} f(w)$ 为一个求极小极大值问题：

$$\min_{w} \theta_P(w) = \min_{w} \max_{\alpha,\beta\,;\,\alpha_i \ge 0} L(w,\alpha,\beta)$$

这里 $\min\limits_{w} \theta_P(w)$ 记为 p^*

➢ 直接求解困难，进而转向另一个问题： $\theta_D(\alpha,\beta) = \min\limits_{w} L(w,\alpha,\beta)$

附录：线性支持向量机模型推导

对偶问题和KKT条件

这里，先固定α, β，求拉格朗日函数关于w的最小值，再求 $\theta_D(\alpha, \beta)$ 最大值，则有：

$$\max_{\alpha,\beta\,;\,\alpha_i \geq 0} \theta_D(\alpha, \beta) = \max_{\alpha,\beta\,;\,\alpha_i \geq 0} \min_{w} L(w, \alpha, \beta)$$

此问题就是原问题的对偶问题，记为 d^*

- 对比 p^*和 d^*，存在如下的关系：

$$d^* = \max \theta_D\,(\alpha, \beta) = \theta_D(\alpha^*, \beta^*)$$

这里 α^*, β^*是 $\theta_D(\alpha, \beta)$ 取最大时的 α, β 的取值，同时，θ_D是关于w的求最小的拉格朗日函数，则有：

$$d^* = \max \theta_D\,(\alpha, \beta) = \theta_D(\alpha^*, \beta^*) = \min L\,(w, \alpha^*, \beta^*)$$

附录：线性支持向量机模型推导

对偶问题和KKT条件

显然，$\min L(w,\alpha^*,\beta^*)$ 一定小于等于w取任意值的 $L(w,\alpha^*,\beta^*)$，则有：

$$d^* = \max \theta_D(\alpha,\beta) = \theta_D(\alpha^*,\beta^*) = \min L(w,\alpha^*,\beta^*) \le L(w^*,\alpha^*,\beta^*)$$

这里， w取 w^*，即 $L(w,\alpha^*,\beta^*)$ 取最小值时的w值，则有：

$$\begin{aligned} d^* &= \max \theta_D(\alpha,\beta) = \theta_D(\alpha^*,\beta^*) = \min L(w,\alpha^*,\beta^*) \le L(w^*,\alpha^*,\beta^*) \\ &= f(w^*) + \sum_{i=1}^{k} \alpha_i^* g_i(w^*) + \sum_{i=1}^{l} \beta_i^* h_i(w^*) \end{aligned}$$

由于 $g_i(w) \le 0, h_i(w) = 0$，则有：

$$\begin{aligned} d^* &= \max \theta_D(\alpha,\beta) = \theta_D(\alpha^*,\beta^*) = \min L(w,\alpha^*,\beta^*) \le L(w^*,\alpha^*,\beta^*) \\ &= f(w^*) + \sum_{i=1}^{k} \alpha_i^* g_i(w^*) + \sum_{i=1}^{l} \beta_i^* h_i(w^*) \le f(w^*) \end{aligned}$$

附录：线性支持向量机模型推导

对偶问题和KKT条件

之前已知：

$$p^* = \min_{w} \theta_P(w) = f(w^*)$$

则有：

$$\begin{aligned} d^* &= \max \theta_D(\alpha, \beta) = \theta_D(\alpha^*, \beta^*) = \min L(w, \alpha^*, \beta^*) \leq L(w^*, \alpha^*, \beta^*) \\ &= f(w^*) + \sum_{i=1}^{k} \alpha_i^* g_i(w^*) + \sum_{i=1}^{l} \beta_i^* h_i(w^*) \leq f(w^*) = p^* \end{aligned}$$

此时可知，$d^* \leq p^*$，原问题与对偶问题是不等价，只有当上式中的两个$\leq$均取等号时，原问题与对偶问题才等价。

附录：线性支持向量机模型推导

对偶问题和KKT条件

$$d^* = \max \theta_D(\alpha,\beta) = \theta_D(\alpha^*,\beta^*) = \min L(w,\alpha^*,\beta^*) \le L(w^*,\alpha^*,\beta^*)$$

$$= f(w^*) + \sum_{i=1}^{k} \alpha_i^* g_i(w^*) + \sum_{i=1}^{l} \beta_i^* h_i(w^*) \le f(w^*) = p^*$$

当 $w = w^*$ 时，即 $\frac{\partial}{\partial w_i} L(w^*,\alpha^*,\beta^*) = 0$时，第一个 $\le$ 取等号，

由于 $h_j(w) = 0$ ，当$\alpha_i^* g_i(w^*) = 0$时，第二个$\le$取等号。

即满足 $\frac{\partial}{\partial w_i} L(w^*,\alpha^*,\beta^*) = 0$ 和 $\alpha_i^* g_i(w^*) = 0$时，$d^* = p^*$

➢所以，当同时满足$\frac{\partial}{\partial w_i} L(w^*,\alpha^*,\beta^*) = 0$ 和$\alpha_i^* g_i(w^*) = 0$时，原问题与对偶问题等价。

线性支持向量机模型推导

- KKT (Karush-Kuhn-Tucker) 条件

$$\frac{\partial}{\partial w_i} L(w^*, \alpha^*, \beta^*) = 0, i = 1, \dots, n$$

$$\boxed{\alpha_i^* g_i(w^*) = 0, i = 1, \dots, k}$$

KKT互补松弛条件

$$g_i(w^*) \le 0, i = 1, \dots, k$$

$$\alpha^* \ge 0, i = 1, \dots, k$$

如果w^*, α^*, β^*满足上述的KKT条件，那么它们就是原问题和对偶问题的解。

➢ KKT互补松弛条件分析：

补充条件隐含如果 $\alpha^* > 0$，那么 $g_i(w^*) = 0$，即w处于可行域的边界上，是起作用的(Active)约束，而位于可行域内部的点都是不起作用的约束，其 $\alpha^* = 0$。