第七章

支持向量机

张志忠 华东师范大学

1、支持向量机回顾

2、支持向量机对偶形式

3、软间隔支持向量机

4、非线性支持向量机

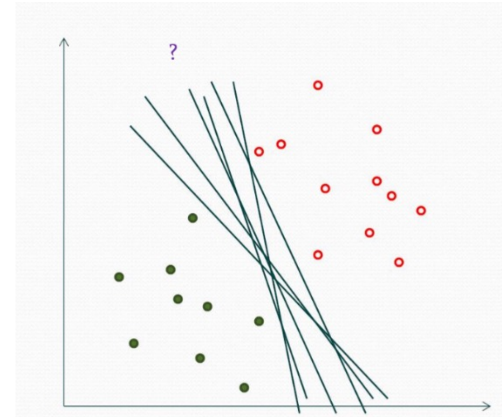
5、作业

1、支持向量机回顾

- 支持向量机

函数间隔: $\hat{\gamma}_i = y_i(w \cdot x_i + b)$

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



$$\max_{w,b} \quad \gamma \qquad \text{几何间隔}$$

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

$$\max_{w,b} \quad \frac{\hat{\gamma}}{\|w\|} \qquad \text{函数间隔}$$

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

取 $\gamma = 1$

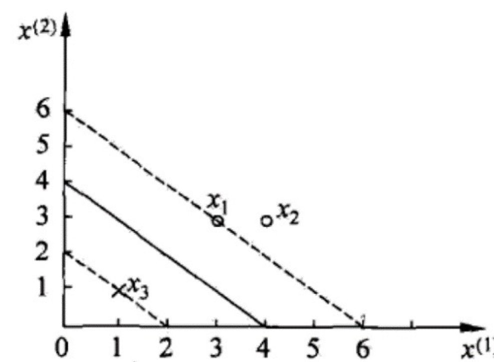


图 7.4　间隔最大分离超平面示例

线性可分支持向量机学习的最优化问题

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

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

• 支持向量机

例 7.1 数据与例 2.1 相同．已知一个如图 7.4 所示的训练数据集，其正例点是 $x_1=(3,3)^{\mathrm{T}}$，$x_2=(4,3)^{\mathrm{T}}$，负例点是 $x_3=(1,1)^{\mathrm{T}}$，试求最大间隔分离超平面．

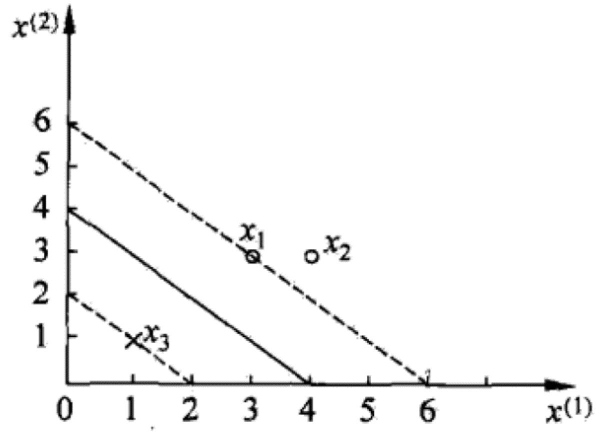


图 7.4 间隔最大分离超平面示例

$$
\begin{aligned}
\min_{w,b} \quad & \frac{1}{2}(w_1^2+w_2^2) \\
\text{s.t.} \quad & 3w_1+3w_2+b \geqslant 1 \\
& 4w_1+3w_2+b \geqslant 1 \\
& -w_1-w_2-b \geqslant 1
\end{aligned}
$$

凸优化问题:

对于约束优化问题

$$
\begin{aligned}
& \min f(x) \\
s.t. \quad & g_j(x) \le 0 \quad j=1,2,\ldots,m
\end{aligned}
$$

若 $f(x)$ $g_j(x)$都为凸函数，则此问题为凸规划。

凸规划的任何局部最优解就是全局最优解

2、支持向量机对偶形式

拉格朗日对偶

- 如何求解：(7.13) ~ (7.14) 1
- 在约束最优化问题中，常常利用拉格朗日对偶性(Lagrange duality)将原始问题转换为对偶问题，通过解对偶问题得到原始问题的解

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

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

$$\min_{x \in \mathbf{R}^n} \ f(x)$$

$$\text{s.t.} \quad c_i(x) \leqslant 0, \quad i = 1, 2, \cdots, k$$

$$h_j(x) = 0, \quad j = 1, 2, \cdots, l$$

- 引进拉格朗日函数 α_i, β_j 为乘子 $\alpha_i \geqslant 0$

$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^{k} \alpha_i c_i(x) + \sum_{j=1}^{l} \beta_j h_j(x)$$

【数之道】支持向量机SVM是什么，八分钟直觉理解其本质_哔哩哔哩_bilibili

拉格朗日对偶

$$\min_{x\in R^n} f(x)$$
$$\text{s.t.}\quad c_i(x)\leqslant 0,\quad i=1,2,\cdots,k$$
$$h_j(x)=0,\quad j=1,2,\cdots,l$$

原始问题 对偶问题

$$\underline{\min_{x}\ \max_{\alpha,\beta:\alpha_i\geqslant 0}}\ L(x,\alpha,\beta) \iff \underline{\max_{\alpha,\beta:\alpha_i\geqslant 0}\ \min_{x}}\ L(x,\alpha,\beta)$$

极小极大问题 极大极小问题

x*，和α*，β*分别是原始问题和对偶问题的解的充分必要条件是x*，和α*，β*满足karush-Kuhn-Tucker(KKT)条件：

$$\nabla_x L(x^*,\alpha^*,\beta^*)=0$$
$$\nabla_\alpha L(x^*,\alpha^*,\beta^*)=0$$
$$\nabla_\beta L(x^*,\alpha^*,\beta^*)=0$$
$$\underline{\alpha_i^* c_i(x^*)=0,\quad i=1,2,\cdots,k}$$
$$c_i(x^*)\leqslant 0,\quad i=1,2,\cdots,k$$
$$\alpha_i^*\geqslant 0,\quad i=1,2,\cdots,k$$
$$h_j(x^*)=0\quad j=1,2,\cdots,l$$

互补条件

$$\min f(x)$$
$$s.t. g_1(x)=a-x\le 0$$
$$g_2(x)=x-b\le 0$$

$$\Rightarrow \boxed{\begin{array}{l} g_1(x)=a-x\le 0 \\ g_2(x)=x-b\le 0 \end{array}} \rightarrow \boxed{\begin{array}{l} h_1(x,a_1)=g_1(x)+a_1^2=a-x+a_1^2=0 \\ h_2(x,b_1)=g_2(x)+b_1^2=x-b+b_1^2=0 \end{array}}$$

$$L(x,a_1,b_1,\mu_1,\mu_2)=f(x)+\mu_1(a-x+a_1^2)+\mu_2(x-b+b_1^2)$$

$$\min f(x)$$
$$s.t.\, g_1(x) = a - x \le 0$$
$$g_2(x) = x - b \le 0$$

$$\boxed{\begin{aligned} g_1(x) &= a - x \le 0 \\ g_2(x) &= x - b \le 0 \end{aligned}} \Rightarrow \boxed{\begin{aligned} h_1(x, a_1) &= g_1(x) + a_1^2 = a - x + a_1^2 = 0 \\ h_2(x, b_1) &= g_2(x) + b_1^2 = x - b + b_1^2 = 0 \end{aligned}}$$

$$L(x, a_1, b_1, \mu_1, \mu_2) = f(x) + \mu_1(a - x + a_1^2) + \mu_2(x - b + b_1^2)$$

$$\frac{\partial F}{\partial a_1} = 2\mu_1 a_1 = 0, \qquad \frac{\partial F}{\partial b_1} = 2\mu_2 b_1 = 0,$$

互补松弛条件：

$$\alpha_i^* c_i(x^*) = 0, \quad i = 1, 2, \cdots, k$$

对于 $\mu_1 a_1 = 0$，我们有两种情况：

情形1： $\mu_1 = 0, a_1 \neq 0$

此时由于乘子 $\mu_1 = 0$，因此 g_1 与其相乘为零，可以理解为**约束 g_1 不起作用，且有** $g_1(x) = a - x < 0$.

情形2： $\mu_1 \geq 0, a_1 = 0$

此时 $g_1(x) = a - x = 0$ 且 $\mu_1 > 0$，可以理解为**约束 g_1 起作用，且有 $g_1(x) = 0$**.

合并情形1和情形2得： $\mu_1 g_1 = 0$，且在约束起作用时 $\mu_1 > 0$， $g_1(x) = 0$；约束不起作用时 $\mu_1 = 0$， $g_1(x) < 0$.

学习的对偶算法

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

- 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

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

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

拉格朗日目标函数:

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

2 $$\min_{w,b} L(w,b,\alpha)$$ ➡ 求偏导

3 $$\nabla_w L(w,b,\alpha)=w-\sum_{i=1}^{N}\alpha_i y_i x_i=0$$

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

➡

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

$$\sum_{i=1}^{N}\alpha_i y_i=0$$ ➡ 新的约束条件

$\|w\|^2$ $\alpha_i y_i(wx_i+b)$

4 $$L(w,b,\alpha)=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)-\sum_{i=1}^{N}\alpha_i y_i\left(\left(\sum_{j=1}^{N}\alpha_j y_j x_j\right)\cdot x_i+b\right)+\sum_{i=1}^{N}\alpha_i$$

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

学习的对偶算法

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

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

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

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

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

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

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

对偶形式能够便于我们求解非线性问题

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

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

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

定理：设 $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_N^*)$ 是上述最优化问题的最优解，存在下标j，使得 α_j^* 并可按下式求解原始问题的最优解 $\mathbf{w}^*, \mathbf{b}^*$

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

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

证明:

由 KKT 条件:

$$\nabla_w L(w^*,b^*,\alpha^*) = w^* - \sum_{i=1}^{N} \alpha_i^* y_i x_i = 0 \qquad (7.27)$$

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

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

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

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

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

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：$\alpha^* = 0$ 由 (7.27) 可知w*=o，

因为数据集有正类和负类点，所以（w,b)=(0,b)不是最优化可行解

但这不是原始优化问题的解，产生矛盾

互补松弛条件

$$\alpha_i^* c_i(x^*) = 0, \quad i = 1,2,\cdots,k$$

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

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

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

学习的对偶算法

- 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：

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

- 这就是说，分类决策函数只依赖于输入x和训练样本输入的内积，上式称为线性可分支持向量机的对偶形式.

线性可分支持向量机学习算法

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

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

- 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：$\alpha^* = (\alpha_1^*,\alpha_2^*,\cdots,\alpha_N^*)^{\mathrm{T}}$

线性可分支持向量机学习算法

- 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

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

支持向量

- 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 $x_i \in \mathbf{R}^n$
- 称为支持向量

- 支持向量一定在分割边界上，由KKT互补条件：

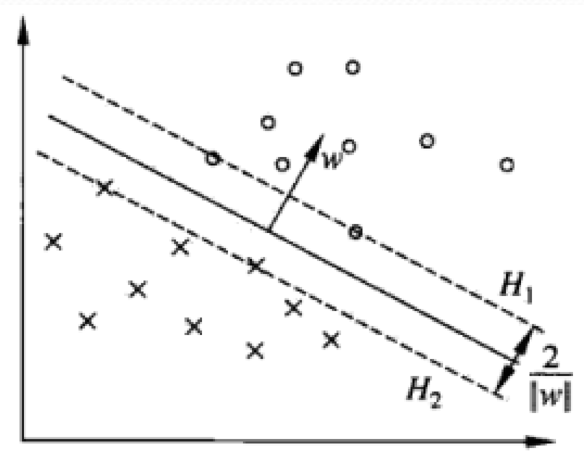
$$\alpha_i^*(y_i(w^* \cdot x_i + b^*) - 1) = 0\,, \quad i = 1, 2, \cdots, N$$

- 对应于 $\alpha_j^* > 0$ 的样本 x_i

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

- 或

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



例 7.1 数据与例 2.1 相同．已知一个如图 7.4 所示的训练数据集，其正例点是 $x_1=(3,3)^{\mathrm{T}}$，$x_2=(4,3)^{\mathrm{T}}$，负例点是 $x_3=(1,1)^{\mathrm{T}}$，试求最大间隔分离超平面．

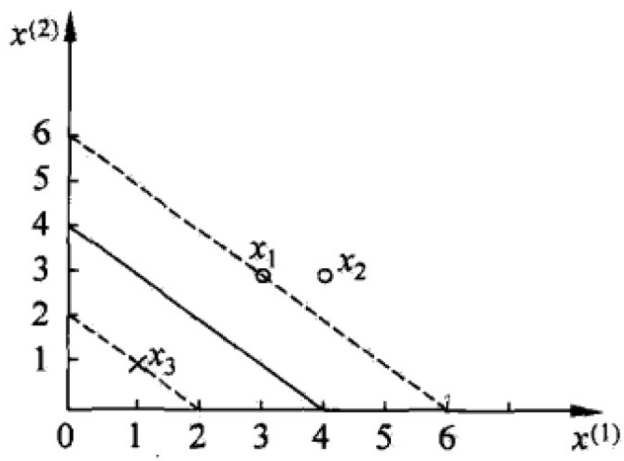


图 7.4 间隔最大分离超平面示例

$$
\begin{aligned}
\min_{w,b}\quad & \frac{1}{2}(w_1^2+w_2^2) \\
\text{s.t.}\quad & 3w_1+3w_2+b\geqslant 1 \\
& 4w_1+3w_2+b\geqslant 1 \\
& -w_1-w_2-b\geqslant 1
\end{aligned}
$$

请写出对偶形式，并求解

$$
\begin{aligned}
\min_{\alpha}\quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)-\sum_{i=1}^{N}\alpha_i \\
\text{s.t.}\quad & \sum_{i=1}^{N}\alpha_i y_i=0 \\
& \alpha_i\geqslant 0,\quad i=1,2,\cdots,N
\end{aligned}
$$

例子：

- 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

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

例子：

- 对α_1,α_2 求偏导数，并令其为o，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

例子：

- 计算得：

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

$$b^* = -2$$

- 分离超平面为：

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

- 分类决策函数为：

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

软间隔支持向量机

二、线性支持向量机与软间隔最大化

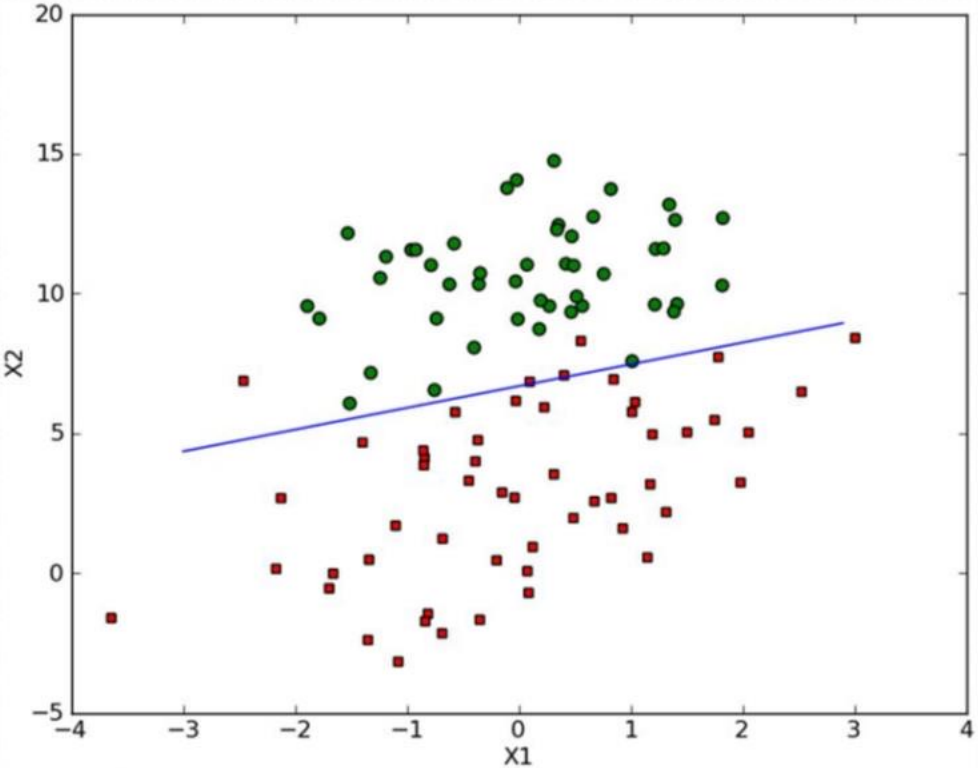
- 训练数据中有一些特异点（outlier），不能满足函数间隔大于等于1的约束条件。
- 解决方法：对每个样本点 (x_i, y_i) 引进一个松弛变量 $\xi_i \geqslant 0$
- 使得函数间隔加上松弛变量大于等于1，约束条件变为

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

目标函数变为：

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

C > 0 为惩罚参数



线性支持向量机与软间隔最大化

- 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

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

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

线性支持向量机与软间隔最大化

- 原始问题 3 的拉格朗日函数：

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

- 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：

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

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

—致的

- 带入

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

新的条件

线性支持向量机与软间隔最大化

- 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：

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

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

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

$$\alpha_i\geqslant 0$$

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

线性支持向量机与软间隔最大化

- 原始问题 3 的对偶问题：

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

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

4

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

- 定理：
- 设 $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_N^*)^{\mathrm{T}}$ 是对偶问题 4 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*

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

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

线性支持向量机学习算法

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

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

- 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： $\alpha^* = (\alpha_1^*,\alpha_2^*,\cdots,\alpha_N^*)^{\mathrm{T}}$

线性支持向量机学习算法

- 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

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

线性不可分的情况下

$\alpha_i^* > 0$ 的样本点 (x_i, y_i) 的实例 x_i 称为支持向量（软间隔的支持向量）

KKT 条件:

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

$$\mu_i^* \xi_i^* = 0$$

C-a*-u*=0

若 $\alpha_i^* < C$，则 $\xi_i = 0$

若 $\alpha_i^* = C$，$0 < \xi_i < 1$

若 $\alpha_i^* = C$，$\xi_i = 1$

若 $\alpha_i^* = C$，$\xi_i > 1$

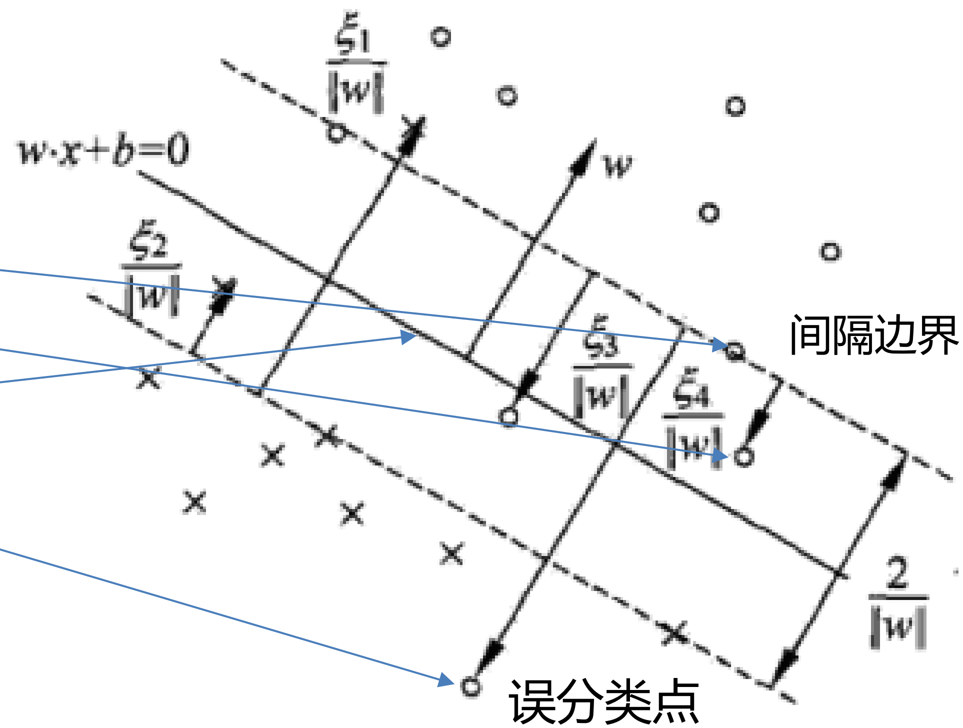


图 7.5 软间隔的支持向量

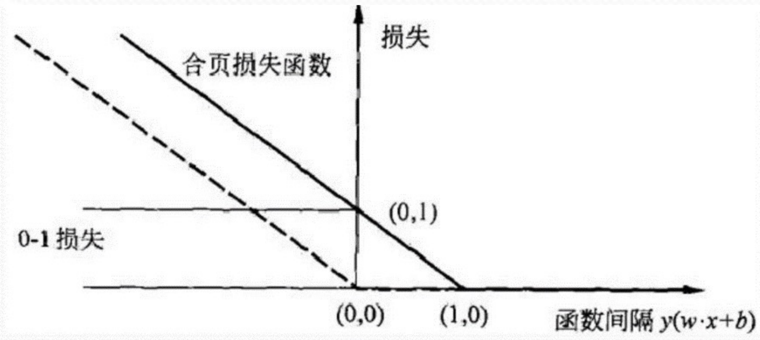
合页损失函数hinge loss function

- 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

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



合页损失函数hinge loss function

- 线性支持向量机原始最优化问题:

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

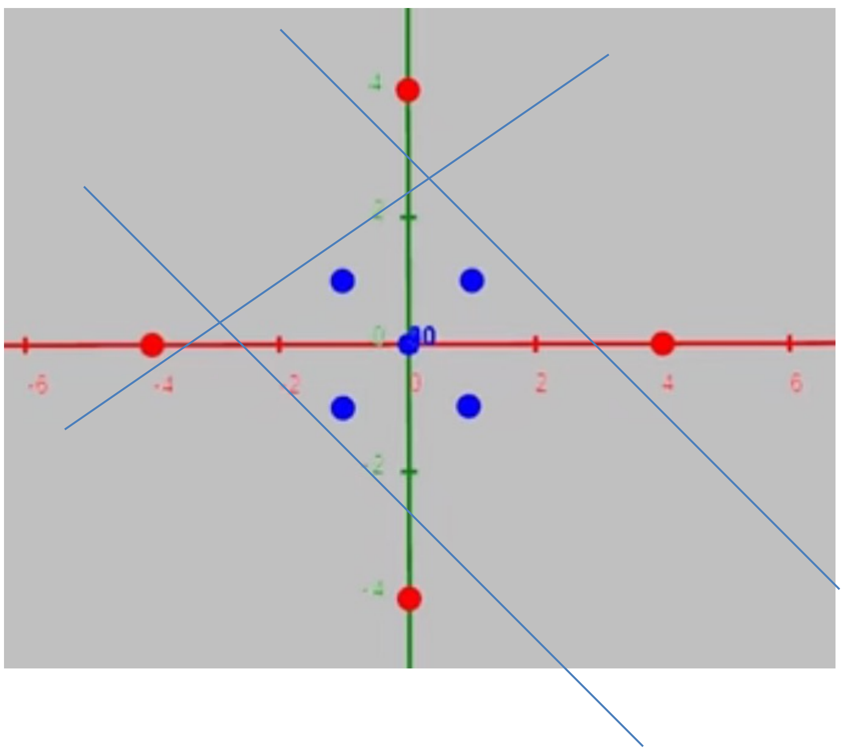
$$\text{s.t.} \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1,2,\cdots,N$$

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

- 等价于:

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

线性SVM： 处理线性可分问题



非线性可分问题

- 非线性问题往往不好求解，所以希望能用解**线性分类问题的方法**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：

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

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

$$\vec{x}=(x_1,x_2) \xrightarrow{T} T(\vec{x})=(1,\sqrt{2}x_1,\sqrt{2}x_2,x_1^2,x_2^2,\sqrt{2}x_1x_2)$$

二维转六维

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

$$\vec{x} = (x_1, x_2) \xrightarrow{T} T(\vec{x}) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)$$ 二维转六维

大家计算一下 <T(X1) ,T(X2)>?

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

维度转换函数T

$$\vec{x}_i = (x_{i1}, x_{i2}) \xrightarrow{T} T(\vec{x}_i) = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in})$$

$$\vec{x}_j = (x_{j1}, x_{j2}) \qquad T(\vec{x}_j) = (x_{j1}, x_{j2}, x_{j3}, \ldots, x_{jn})$$

$$T(\vec{x}_i) \cdot T(\vec{x}_j) \xrightarrow{\text{点积}} \sum_{m=1}^{n} x_{im} x_{jm}$$

核函数$K(\vec{x}_i, \vec{x}_j)$

$$T(\vec{x}_i) \cdot T(\vec{x}_j) = K(\vec{x}_i, \vec{x}_j)$$

直接计算
点积结果

$$\vec{x} = (x_1, x_2) \xrightarrow{T} T(\vec{x}) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)$$ 二维转六维

大家计算一下 <T(X1) ,T(X2)>?

方法1

$$\vec{x} = (x_1, x_2) \xrightarrow{T} T(\vec{x}) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)$$ 二维转六维

$$T(\vec{x}_i) \cdot T(\vec{x}_j) = (1, \sqrt{2}x_{i1}, \sqrt{2}x_{i2}, x_{i1}^2, x_{i2}^2, \sqrt{2}x_{i1}x_{i2}) \cdot (1, \sqrt{2}x_{j1}, \sqrt{2}x_{j2}, x_{j1}^2, x_{j2}^2, \sqrt{2}x_{j1}x_{j2})$$ 六维向量点积

$$= \boxed{1 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} + x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}}$$

相同

方法2

$$K(\vec{x}_i, \vec{x}_j) = (1 + \vec{x}_i \cdot \vec{x}_j)^2 = (1 + x_{i1}x_{j1} + x_{i2}x_{j2})^2 = \boxed{1 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} + x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}}$$

直接使用二维向量点积结果计算

多项式核

非线性支持向量机与核函数

- 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.

非线性支持向量机与核函数

- 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

非线性支持向量机与核函数

- 核技巧的想法是：
- 在学习与预测中只定义核函数K(x,z)，而不显式地定义映射函数，通常，直接计算K(x,z)比较容易，而通过$\phi(x)$和$\phi(z)$计算K(x, z)并不容易。
- 注意：φ是 输入空间R^n到特征空间H的映射，特征空间H一般是高维，映射可以不同。

非线性支持向量机与核函数

- 例：
- 假设输入空间是R²，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 $\phi(x):\mathbf{R}^2\rightarrow\mathcal{H}$
- 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： $\phi(x)\cdot\phi(z)=(x\cdot z)^2=K(x,z)$

非线性支持向量机与核函数

- 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$\phi(x):\mathbf{R}^2\rightarrow\mathcal{H}$
- 解：
- 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。

核函数在支持向量机的应用

- 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：

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

正定核

- 问题：
  - 已知映射函数φ，可以通过 $\phi(x)$ 和 $\phi(z)$ 的内积求得核函数K(x,z).
  - 不用构造映射φ, 能否直接判断一个给定的函数K(x,z)是不是核函数?
  - 或者说，函数K(x,z)满足什么条件才能成为核函数?
- 假设K(x,z)是定义在X*X上的对称函数，并且对任意的

$$x_1, x_2, \cdots, x_m \in \mathcal{X}$$

- K(x,z)关于 $x_1, x_2, \cdots, x_m$ 的Gram矩阵是半正定的，可以依据函数K(x,z)，构成一个希尔伯特空间(Hilbert space)； **判断的依据**
- 其步骤是首先定义映射φ，并构成向量空间S，然后在S上定义内积构成内积空间; 最后将S完备化构成希尔伯特空间.

常用核函数

- 1、多项式核函数（Polynomial kernel function)

$$K(x,z)=(x\cdot z+1)^p$$

- 对应的支持向量机为P次多项式分类器，分类决策函数：

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

- 2、高斯核函数 （Gaussian Kernel Function)

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

- 决策函数：

$$f(x)=\mathrm{sign}\left(\sum_{i=1}^{N_s}a_i^* y_i \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)+b^*\right)$$

$$K_d(\vec{\boldsymbol{x}}_i, \vec{\boldsymbol{x}}_j) = (c + \vec{\boldsymbol{x}}_i \cdot \vec{\boldsymbol{x}}_j)^d$$

常数项是非常关键的

$$c = 1, d = 2$$

$$K_d(\vec{\boldsymbol{x}}_i, \vec{\boldsymbol{x}}_j) = (1 + \vec{\boldsymbol{x}}_i \cdot \vec{\boldsymbol{x}}_j)^2 = (1 + x_{i1}x_{j1} + x_{i2}x_{j2})^2 = 1 + \underbrace{2x_{i1}x_{j1} + 2x_{i2}x_{j2}}_{\text{一次项点积部分}} + \underbrace{x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}}_{\text{二次项点积部分}}$$

$$K_2'(\vec{\boldsymbol{x}}_i, \vec{\boldsymbol{x}}_j) = (\vec{\boldsymbol{x}}_i \cdot \vec{\boldsymbol{x}}_j)^2 = (x_{i1}x_{j1} + x_{i2}x_{j2})^2 = \underbrace{x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}}_{\text{二次项点积部分}}$$

$$(c = 0, d = 1) + (c = 0, d = 2)$$

$$K_1'(\vec{\boldsymbol{x}}_i, \vec{\boldsymbol{x}}_j) + K_2'(\vec{\boldsymbol{x}}_i, \vec{\boldsymbol{x}}_j) = \vec{\boldsymbol{x}}_i \cdot \vec{\boldsymbol{x}}_j + (\vec{\boldsymbol{x}}_i \cdot \vec{\boldsymbol{x}}_j)^2 = (x_{i1}x_{j1} + x_{i2}x_{j2}) + (x_{i1}x_{j1} + x_{i2}x_{j2})^2$$

$$= \underbrace{x_{i1}x_{j1} + x_{i2}x_{j2}}_{\text{一次项点积部分}} + \underbrace{x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}}_{\text{二次项点积部分}}$$

$$\vec{x} = (x_1, x_2) \xrightarrow{T} T(\vec{x}) = (a_1x_1, a_1x_2, a_1x_1x_2, a_2x_1^2, a_2x_2^2, \dots a_nx_2^n, a_nx_2^n, \dots \infty)$$

二维转无穷维

方法1

$T(\vec{x}_i) \cdot T(\vec{x}_j) = \infty \cdot \infty$ 不可行

方法2

$T(\vec{x}_i) \cdot T(\vec{x}_j) = K(\vec{x}_i, \vec{x}_j)$

$K(\vec{x}_i, \vec{x}_j) = e^{-\gamma \|\vec{x}_i - \vec{x}_j\|^2}$ 高斯核函数 RBF

$$\begin{aligned}
&K(\vec{x}_i, \vec{x}_j) \\
&= e^{-\frac{\|\vec{x}_i - \vec{x}_j\|^2}{2}} \\
&= e^{-\frac{1}{2}(\vec{x}_i - \vec{x}_j)\cdot(\vec{x}_i - \vec{x}_j)} \\
&= e^{-\frac{1}{2}(\vec{x}_i\cdot\vec{x}_i + \vec{x}_j\cdot\vec{x}_j - 2\vec{x}_i\cdot\vec{x}_j)} \\
&= e^{-\frac{1}{2}\left(\|\vec{x}_i\|^2 + \|\vec{x}_j\|^2 - 2\vec{x}_i\cdot\vec{x}_j\right)} \\
&= \underline{e^{-\frac{1}{2}\left(\|\vec{x}_i\|^2 + \|\vec{x}_j\|^2\right)}}\, e^{\vec{x}_i\cdot\vec{x}_j} \\
&= Ce^{\vec{x}_i\cdot\vec{x}_j}
\end{aligned}$$

$$e^{-\frac{1}{2}\left(\|\vec{x}_i\|^2 + \|\vec{x}_j\|^2\right)} = C$$

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

$$K(\vec{x}_i, \vec{x}_j) = Ce^{\vec{x}_i\cdot\vec{x}_j} = C\sum_{n=0}^{\infty} \frac{\vec{x}_i\cdot\vec{x}_j{}^n}{n!} = C\sum_{n=0}^{\infty} \frac{K_{Poly(n)}(\vec{x}_i, \vec{x}_j)}{n!}$$

非线性支持向量机学习算法

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

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

- 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：

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

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

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

5

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

非线性支持向量机学习算法

- 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

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

- 当K(x,z)是正定核函数时， 5 是凸二次规划问题，解是存在的。

Q & A