

SVM-2



慧科集

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 $x_1 = (3, 3)^T$ ， $x_2 = (4, 3)^T$ ，负例点是 $x_3 = (1, 1)^T$ ，试求最大间隔分离超平面。

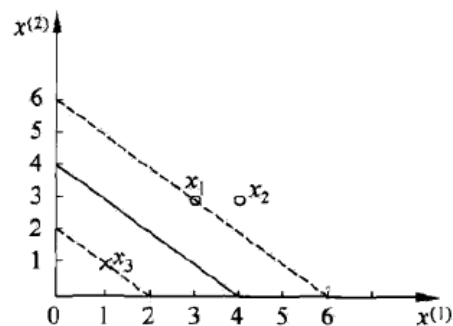


图 7.4 间隔最大分离超平面示例

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{s.t.}, \quad & \alpha_i \geq 0, i = 1, \dots, n \\ & \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$

解 按照算法 7.1，根据训练数据集构造约束最优化问题：

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

$$-\frac{1}{2} (18\alpha_1^2 + 25\alpha_2^2 + 2\alpha_3^2 + 42\alpha_1\alpha_2 - 12\alpha_1\alpha_3 - 14\alpha_2\alpha_3) + \alpha_1 + \alpha_2 + \alpha_3$$

$$\alpha_1 + \alpha_2 - \alpha_3 = 0$$

$$\alpha_i \geq 0, \quad i = 1, 2, 3$$

将 $\alpha_1 + \alpha_2 = \alpha_3$
代入目标函数, 得到关于 α_1, α_2 的函数:

$$-4\alpha_1^2 - \frac{13}{2}\alpha_2^2 - 10\alpha_1\alpha_2 + 2\alpha_1 + 2\alpha_2$$

$$\alpha_i \geq 0, \quad i=1,2,3$$

$$\alpha_1 = 1/4$$

$$\alpha_2 = 0$$

$$\alpha_3 = 1/4$$

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

KKT 条件

首先给出形式化的不等式约束优化问题：

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s. t.} \quad & h_i(x) = 0, \quad i = 1, 2, \dots, m \\ & g_j(x) \leq 0, \quad j = 1, 2, \dots, n \end{aligned}$$

列出 Lagrangian 得到无约束优化问题：

$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^m \alpha_i h_i(x) + \sum_{j=1}^n \beta_j g_j(x)$$

经过之前的分析，便得知加上不等式约束后可行解 x 需要满足的就是以下的 KKT 条件：

$$\nabla_x L(x, \alpha, \beta) = 0 \tag{1}$$

$$\beta_j g_j(x) = 0, \quad j = 1, 2, \dots, n \tag{2}$$

$$h_i(x) = 0, \quad i = 1, 2, \dots, m \tag{3}$$

$$g_j(x) \leq 0, \quad j = 1, 2, \dots, n \tag{4}$$

$$\beta_j \geq 0, \quad j = 1, 2, \dots, n \tag{5}$$



慧科集团旗下企业

线性不可分情况下

根据线性可分情况下的结论：

$$w = \sum_{i=1}^n \alpha_i y^{(i)} x^{(i)}$$

将分类函数变形得最终分类函数，为：

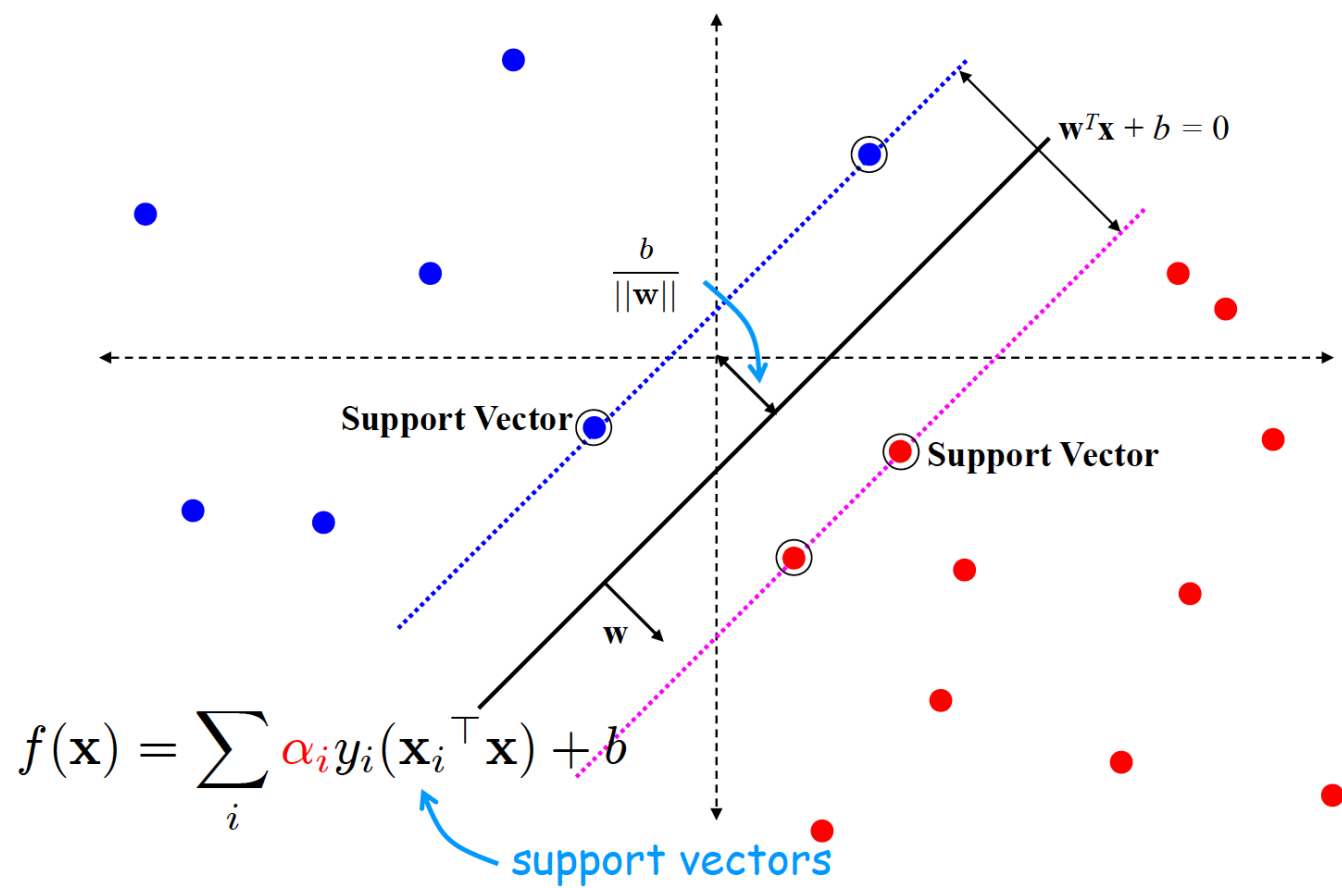
$$f(x) = \left(\sum_{j=1}^N \alpha_j y_j \mathbf{x}_j \right)^{\top} \mathbf{x} + b = \sum_{j=1}^N \alpha_j y_j (\mathbf{x}_j^{\top} \mathbf{x}) + b$$

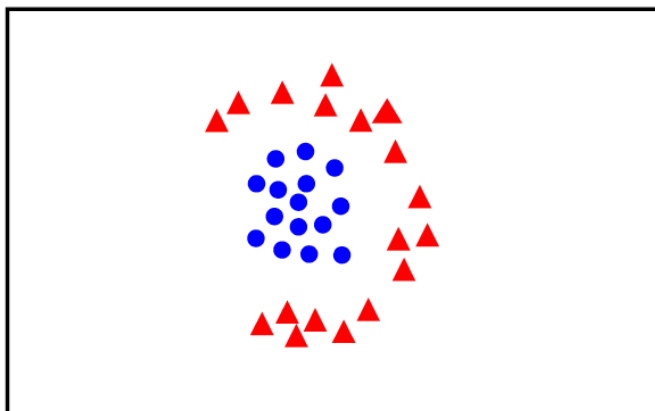
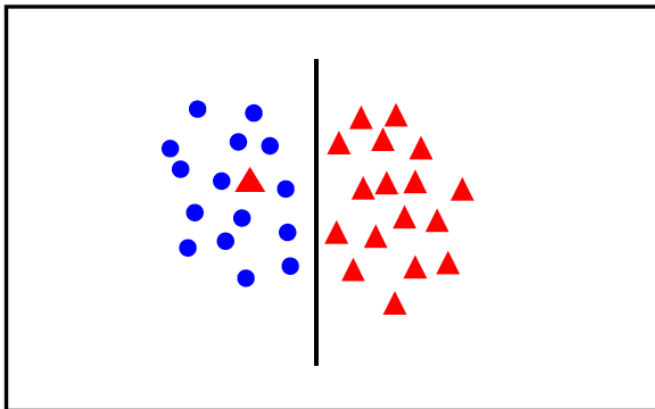
- 损失函数:

$$\|\mathbf{w}\|^2 = \left\{ \sum_j \alpha_j y_j \mathbf{x}_j \right\}^\top \left\{ \sum_k \alpha_k y_k \mathbf{x}_k \right\} = \sum_{jk} \alpha_j \alpha_k y_j y_k (\mathbf{x}_j^\top \mathbf{x}_k)$$

约束条件:

$$y_i \left(\sum_{j=1}^N \alpha_j y_j (\mathbf{x}_j^\top \mathbf{x}_i) + b \right) \geq 1$$





软件隔

$$\min_{\mathbf{w} \in \mathbb{R}^d, \xi_i \in \mathbb{R}^+} \|\mathbf{w}\|^2 + C \sum_i^N \xi_i$$

subject to

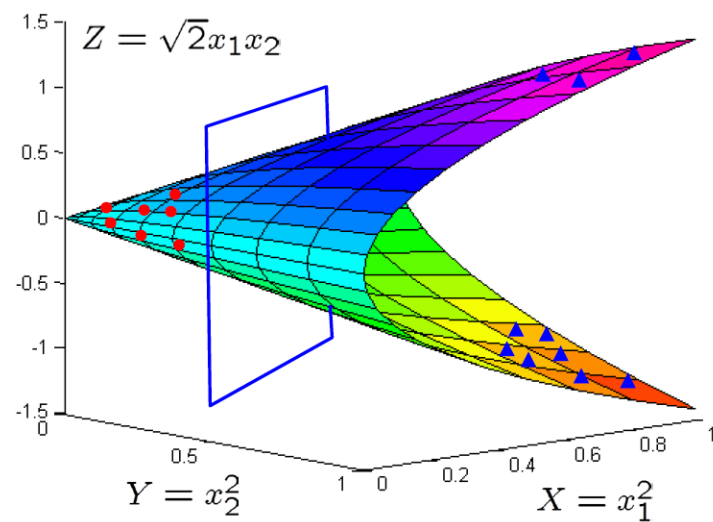
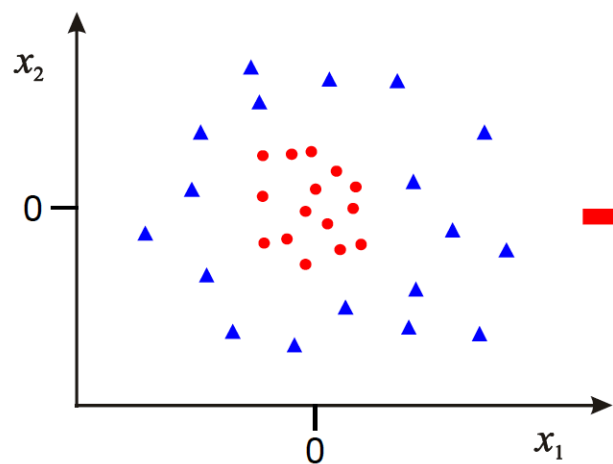
$$y_i (\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i \text{ for } i = 1 \dots N$$

这类问题：

如何解决？

$$\Phi : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{pmatrix} \quad \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

(0,0)
(1,-1)
(1,1)
(-1,1)
(-1,-1)



分类问题:

$$f(\mathbf{x}) = \sum_i^N \alpha_i y_i \mathbf{x}_i^\top \mathbf{x} + b$$
$$\rightarrow f(\mathbf{x}) = \sum_i^N \alpha_i y_i \Phi(\mathbf{x}_i)^\top \Phi(\mathbf{x}) + b$$

学习问题:

$$\max_{\alpha_i \geq 0} \sum_i \alpha_i - \frac{1}{2} \sum_{jk} \alpha_j \alpha_k y_j y_k \mathbf{x}_j^\top \mathbf{x}_k$$
$$\rightarrow \max_{\alpha_i \geq 0} \sum_i \alpha_i - \frac{1}{2} \sum_{jk} \alpha_j \alpha_k y_j y_k \Phi(\mathbf{x}_j)^\top \Phi(\mathbf{x}_k)$$



慧科集团旗下企业

原来在二维空间中一个线性不可分的问题，映射到高维空间后，变成了线性可分的。因此，这也形成了我们最初想解决线性不可分问题的基本思路---向高维空间转化，使其变得线性可分。

而转化的关键的部分在于找到 x 到 y 的映射方法。

如何找到这个映射没有系统的方法，此外，在数据维度较大时，计算困难。

- 维度爆炸问题:

$$\Phi : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{pmatrix} \quad \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

$$\Phi(\mathbf{x})^\top \Phi(\mathbf{z}) = (x_1^2, x_2^2, \sqrt{2}x_1x_2) \begin{pmatrix} z_1^2 \\ z_2^2 \\ \sqrt{2}z_1z_2 \end{pmatrix}$$

$$= x_1^2z_1^2 + x_2^2z_2^2 + 2x_1x_2z_1z_2$$

计算次数: 11次乘法和2次加法

$$\begin{aligned}\Phi(\mathbf{x}^\top \mathbf{z}) &= (\mathbf{x}^\top \mathbf{z})^2 \\ &= (x_1 z_1 + x_2 z_2)^2 \\ &= x_1^2 z_1^2 + x_2^2 z_2^2 + 2x_1 x_2 z_1 z_2\end{aligned}$$

计算次数：3次乘法和1次加法

核函数：对所有 \mathbf{x}, \mathbf{z} 属于 \mathcal{X} ，满足

$$k(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}) \cdot \phi(\mathbf{z}) \rangle$$

可以在特征空间中直接计算内积 $\langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}) \rangle$ ，就像在原始输入点的函数中一样

分类函数为:

$$\sum_{i=1}^n \alpha_i y_i \kappa(\mathbf{x}_i, \mathbf{x}) + b$$

优化问题的表达式:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \kappa(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t.}, \quad & \alpha_i \geq 0, i = 1, \dots, n \\ & \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$

常见核函数

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + c)^d$$

- 线性核

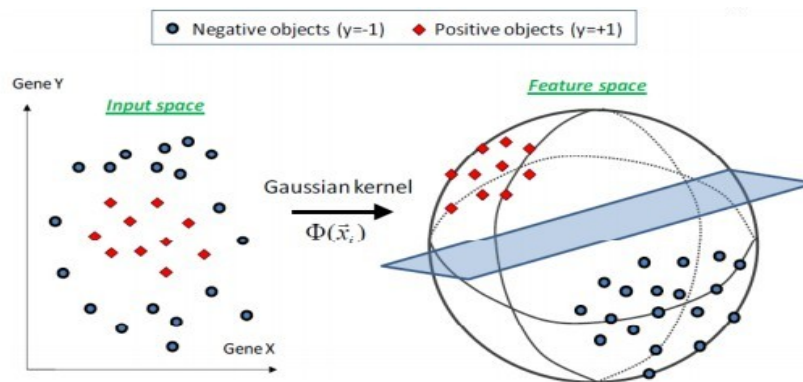
$$k(x, y) = \langle x \cdot y \rangle$$

- 高斯径向基函数核

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right)$$

- Sigmoid核

$$k(\mathbf{x}, \mathbf{y}) = \tanh(\kappa(\mathbf{x} \cdot \mathbf{y}) + \Theta)$$





慧科集团旗下企业

对于核函数的选择，现在还缺乏指导原则。各种实验的观察结果表明，某些问题用某些核函数效果很好，用另一些很差，但一般来讲，径向基核函数是不会出现太大偏差的一种，一般作为首选。



慧科集团旗下企业

- 无穷维-高斯核函数

$$\begin{aligned}k(x, y) &= \exp(-\|x - y\|^2) \\&= \exp(-(x_1 - y_1)^2 - (x_2 - y_2)^2) \\&= \exp(-x_1^2 + 2x_1y_1 - y_1^2 - x_2^2 + 2x_2y_2 - y_2^2) \\&= \exp(-\|x\|^2) \exp(-\|y\|^2) \exp(2x^T y)\end{aligned}$$

$$k(x, y) = \exp(-\|x\|^2) \exp(-\|y\|^2) \sum_{n=0}^{\infty} \frac{(2x^T y)^n}{n!}$$



慧科集团旗下企业

- 高斯核函数的理解