SVM

回顾LR

支持向量机其基本原理，通俗来讲，它是一种二类分类模型，其基本模型定义为特征空间上的间隔最大的线性分类器，其学习策略便是间隔最大化线性分类器使用超平面类型的边界，非线性分类器使用超曲面。

数据：线性可分&线性不可分

两种情况

- **线性可分**
- **线性不可分**

情况1：样本本质上是非线性可分的

解决方法：核函数

情况2：本质上线性，非线性由噪音导致

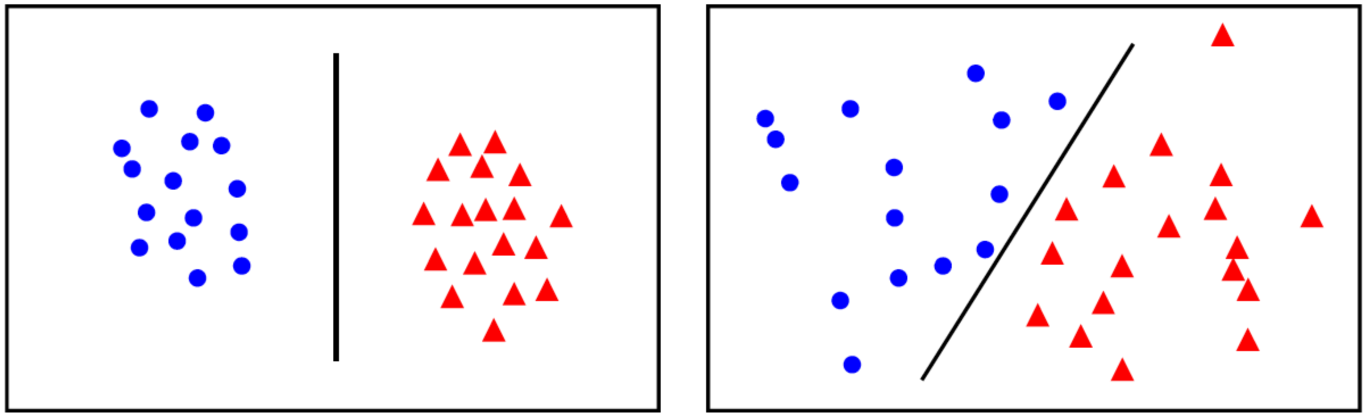
强制使用非线性函数，会导致过拟合

解决方法：软间隔

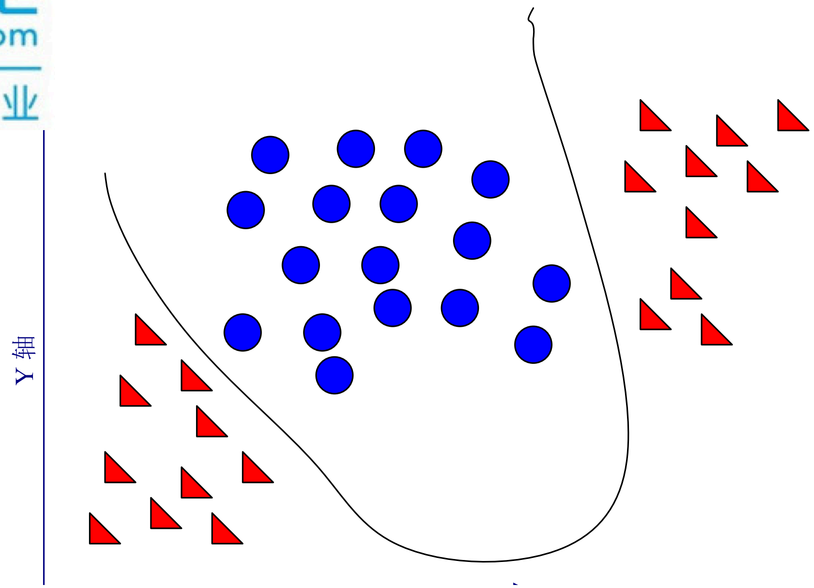
线性可分

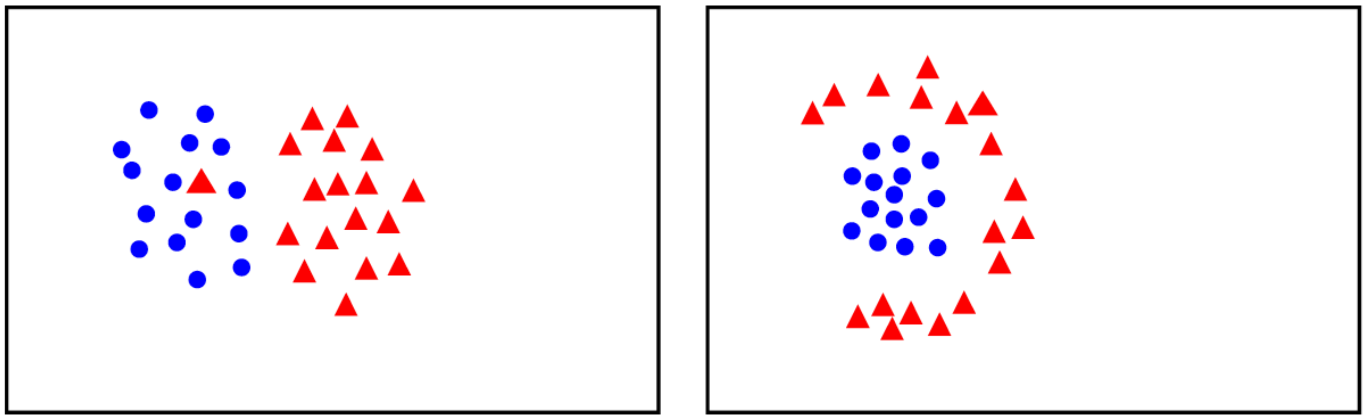
定义：

对于来自两类的一组模式能用一个线性判别函数正确分类，则称他们是线性可分的。



线性不可分

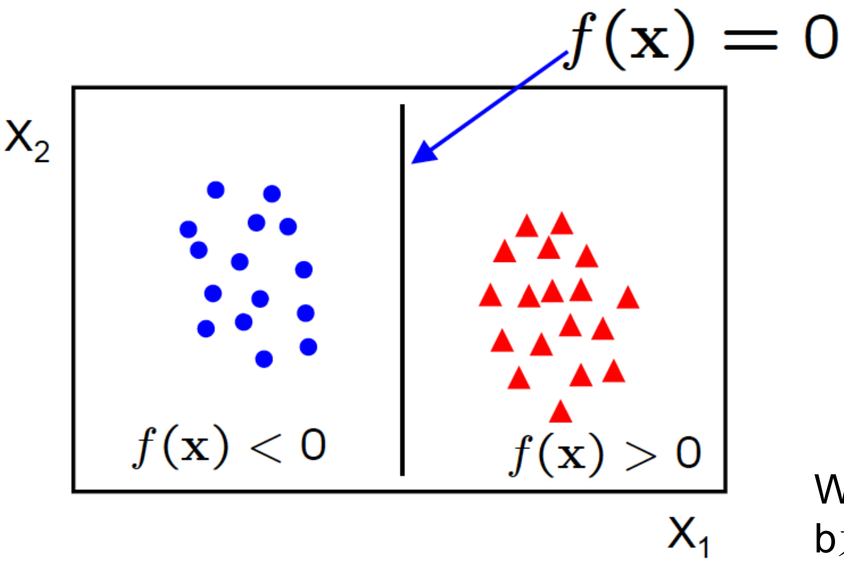




线性可分情况

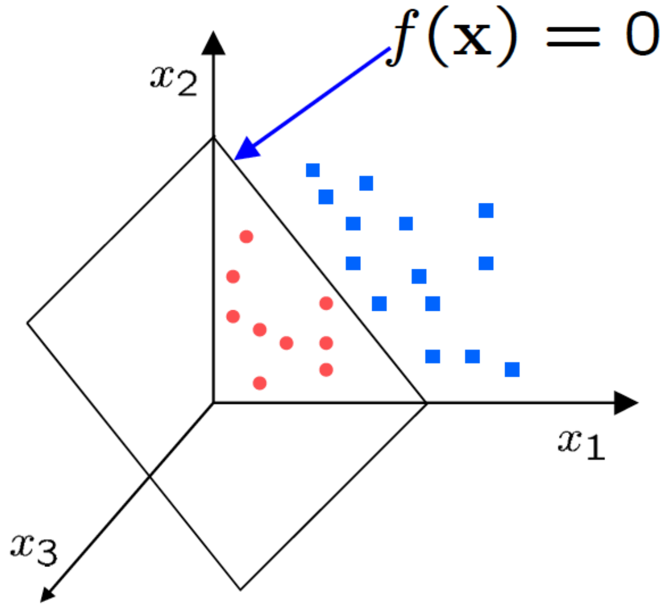
- 我们怎样才能取得一个最优的划分直线f(x)呢？

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$$



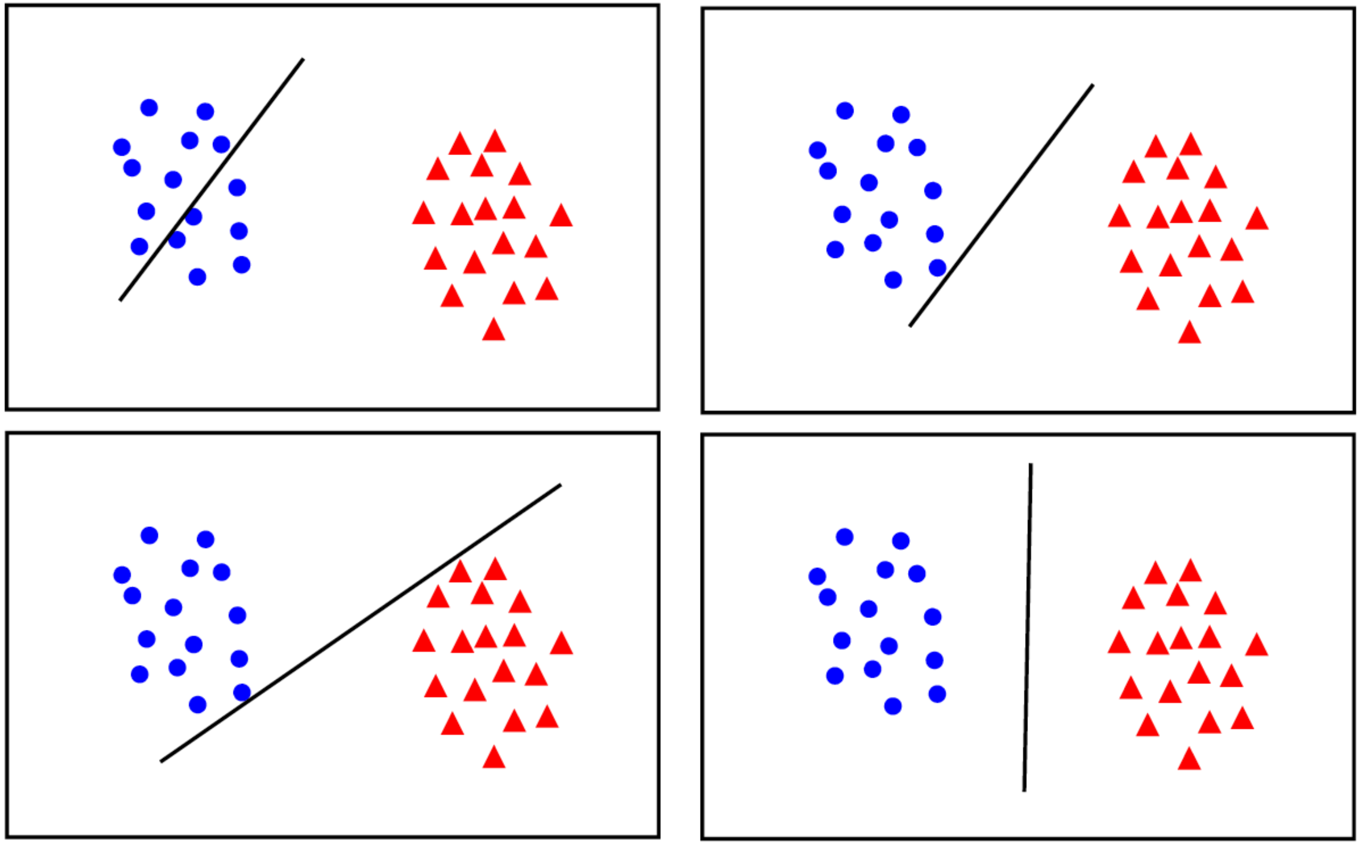
W权重向量，需要学习的参数
b为偏差，需要学习的参数

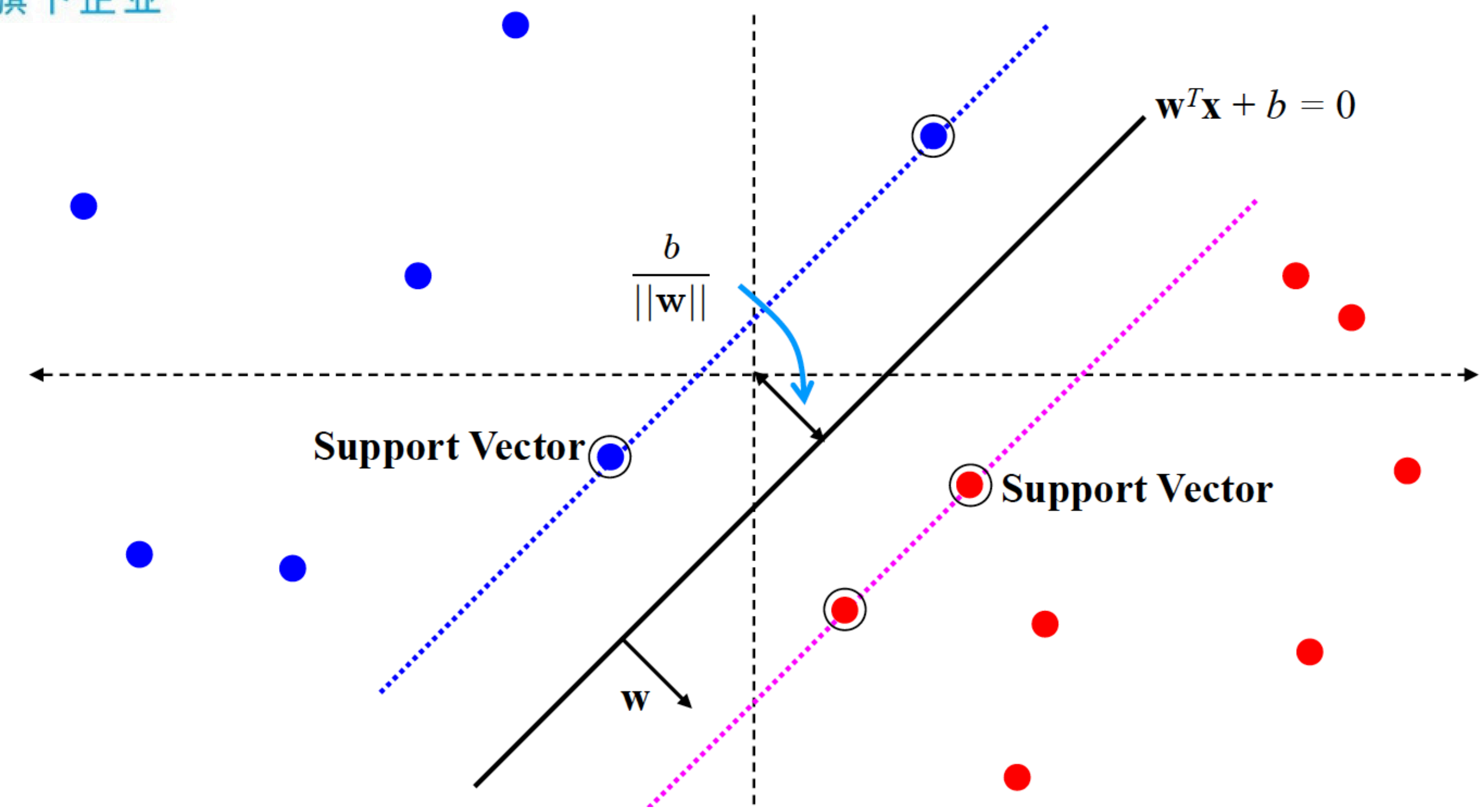
$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$$



多维超平面

最大间隔

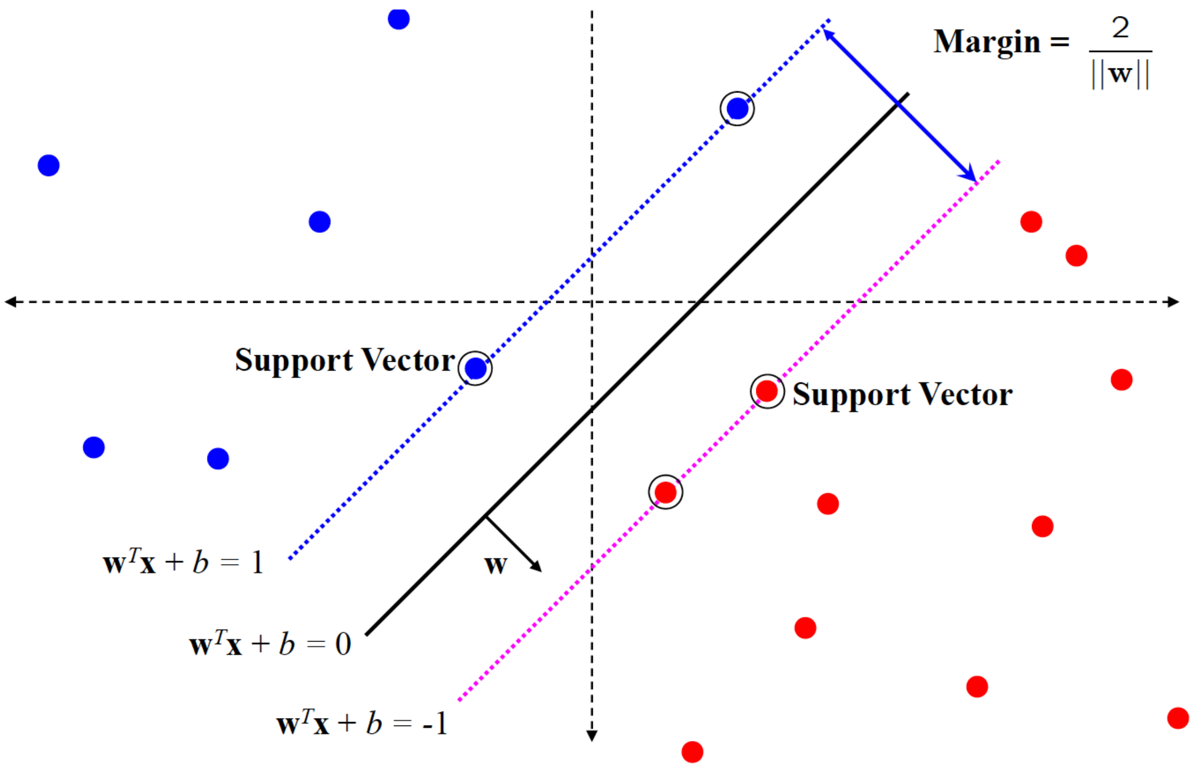




$\mathbf{w}^T\mathbf{x} + b = 0$
$\frac{b}{||\mathbf{w}||}$
Support Vector
Support Vector
$\mathbf{w}$

y取{-1，+1}

$$\hat{\gamma} = y(w^T x + b) = yf(x)$$



几何间隔

- 点到直线：

$$\tilde{\gamma} = \frac{\hat{\gamma}}{\|w\|} = \frac{w^T x}{\|w\|} + \frac{b}{\|w\|} = \frac{f(x)}{\|w\|}$$

- 这时如果成比例的改变w和b，几何间隔的值不会发生改变。

因此，我们的最大间隔分类的目标函数可以定义为：

$$\max \frac{\hat{\gamma}}{\|w\|}$$

$$s.t.\ y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}$$

我们改变优化问题的表述方式。

此时，优化问题的表达式为：

$$\min \frac{1}{2}\|w\|^2$$
$$s.t.\; y^{(i)}(w^T x^{(i)}+b)\geq 1$$

我们的优化问题转变成了一个凸优化问题

拉格朗日乘子法

- minf(w)
- s.t. gi(w)≤0 i=1,2,...,k

  hi(w)=0 i=1,2,...,l （这里0指的是零向量）

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

定义：

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

当所有约束条件都满足时有 $\theta_p = f(w)$

$$\mathcal{L}(w,b,\alpha)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{n}\alpha_i\left(y_i(w^T x_i+b)-1\right)$$

对偶问题

$$p^* = \min_{w,b} \theta(w) = \min_{w,b} \max_{\alpha_i \geq 0} L(w,b,a)$$

$$d^* = \max_{\alpha_i \geq 0} \min_{w,b} L(w,b,a)$$

一般有 $d^* \leq p^*$ ，但是在某些特定条件下（KKT），这两个最优化问题会取相同的值。

1.首先固定α，要让L关于w和b最小化，我们分别对w,b偏导并令其等于零，得到

$$\frac{\partial L}{\partial w}=0 \Rightarrow w=\sum_{i=1}^{n}\alpha_i y_i x_i$$

$$\frac{\partial L}{\partial b}=0 \Rightarrow \sum_{i=1}^{n}\alpha_i y_i=0$$

带回 $L(w,b,a)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{n}\alpha_i(y_i(w^T x_i+b)-1)$得到：

$$\begin{aligned}\mathcal{L}(w,b,\alpha) &= \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j-\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j-b\sum_{i=1}^{n}\alpha_i y_i+\sum_{i=1}^{n}\alpha_i \\ &= \sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j\end{aligned}$$

问题转换为：

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$s.t., \alpha_i \geq 0, i = 1, \ldots, n$$

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

由凸二次规划的性质能保证这样最优的向量a是存在的

- 2.求对α的极大，即是关于对偶变量的优化问题（SMO优化算法--序列最小最优化算法）
- 然后根据

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

$$b^* = -\frac{\max_{i:y_i=-1} w^{*T} x_i + \min_{i:y_i=1} w^{*T} x_i}{2}$$

- 可求出最优的w和b，即最优超平面。

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．

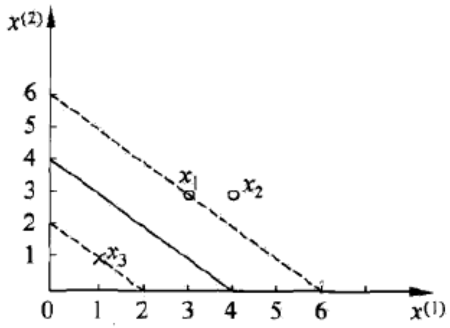


图 7.4　间隔最大分离超平面示例

$$\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$\geq 0, i=1,\ldots,n$$

$$\alpha_i y_i=0$$

解　按照算法 7.1，根据训练数据集构造约束最优化问题：

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

$$\begin{aligned}\text{s.t.}\quad & 3w_1+3w_2+b\geqslant 1\\ & 4w_1+3w_2+b\geqslant 1\\ & -w_1-w_2-b\geqslant 1\end{aligned}$$

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

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

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

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

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

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

α_1=1/4

α_2=0

α_3=1/4

代入公式：

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

$$b^* = -\frac{\max\limits_{i:y_i=-1} w^{*T} x_i + \min\limits_{i:y_i=1} w^{*T} x_i}{2}$$

求得此最优化问题的解 $w_1 = w_2 = \frac{1}{2}$， $b = -2$．于是最大间隔分离超平面为

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

其中， $x_1 = (3,3)^{\mathrm{T}}$ 与 $x_3 = (1,1)^{\mathrm{T}}$ 为支持向量．

首先给出形式化的不等式约束优化问题：

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

列出 Lagrangian 得到无约束优化问题：

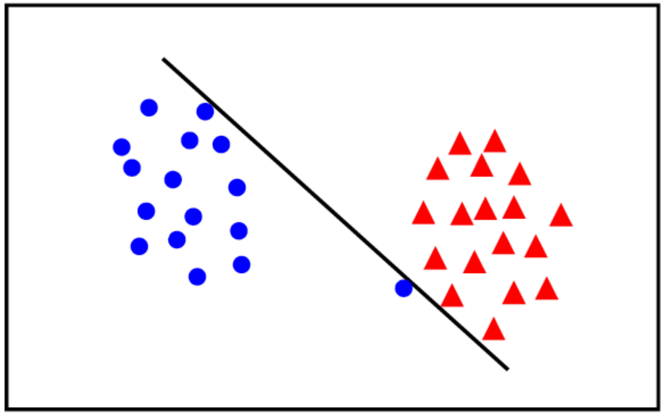
$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^{m} \alpha_i h_i(x) + \sum_{j=1}^{n} \beta_i g_i(x)$$

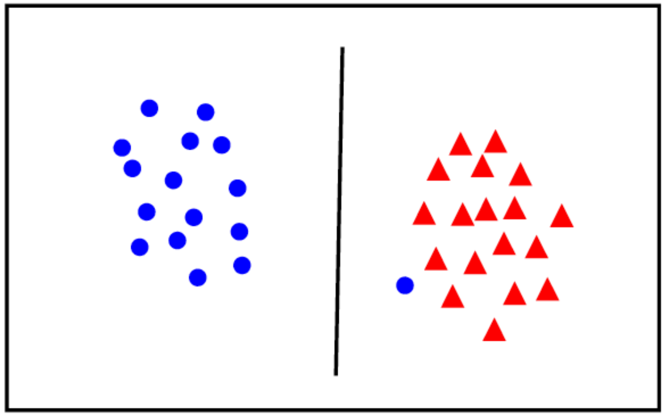
经过之前的分析，便得知加上不等式约束后可行解 x 需要满足的就是以下的 KKT 条件：

$$
\begin{aligned}
\nabla_x L(x, \alpha, \beta) &= 0 && (1) \\
\beta_j g_j(x) &= 0, \ j = 1, 2, \ldots, n && (2) \\
h_i(x) &= 0, \ i = 1, 2, \ldots, m && (3) \\
g_j(x) &\leq 0, \ j = 1, 2, \ldots, n && (4) \\
\beta_j &\geq 0, \ j = 1, 2, \ldots, n && (5)
\end{aligned}
$$

软间隔

哪个划分更好？

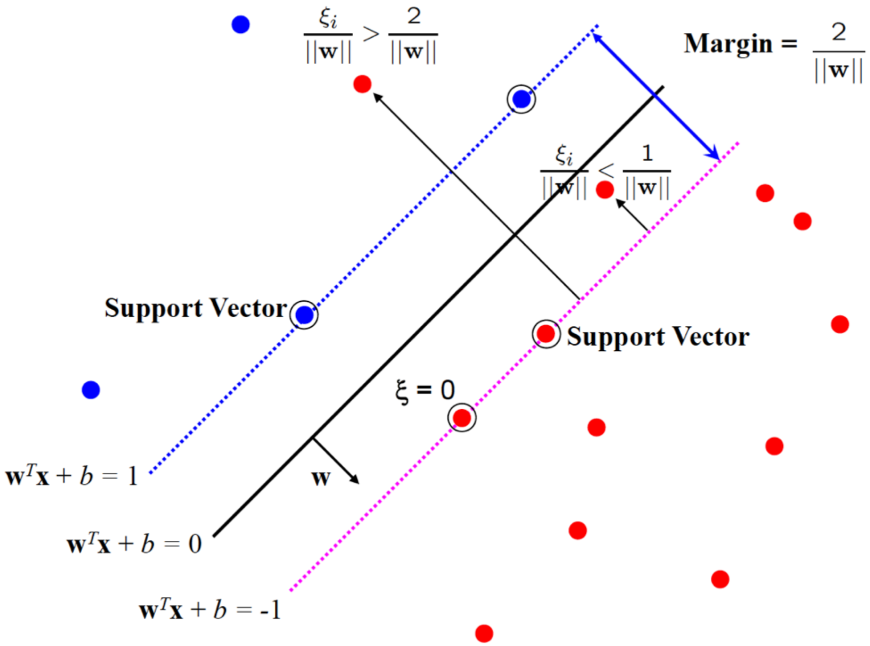




允许一定程度犯错

$$\min_{\mathbf{w}\in\mathbb{R}^d, \xi_i \in \mathbb{R}^+} ||\mathbf{w}||^2 + C \sum_i^N \xi_i \qquad y_i \left(\mathbf{w}^\top \mathbf{x}_i + b \right) \geq 1 - \xi_i \text{ for } i = 1 \ldots N$$

$$\xi_i \geq 0$$

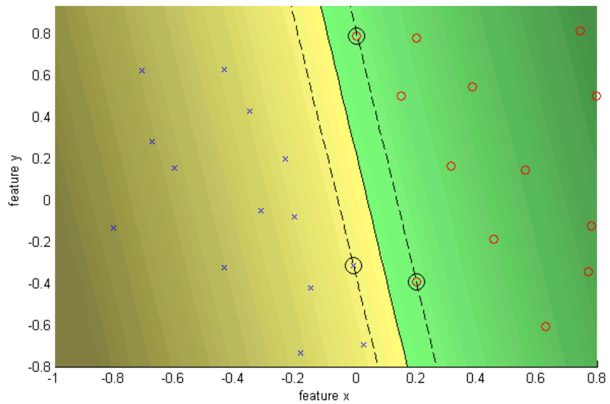


$$y_i f(\mathbf{x}_i) \geq 1 - \xi_i$$

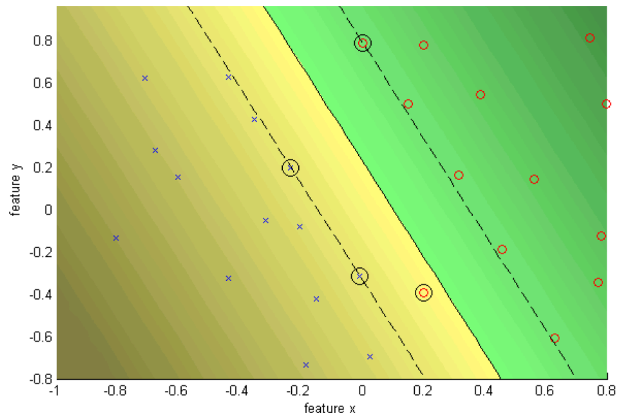
$$\xi_i = \max\left(0, 1 - y_i f(\mathbf{x}_i)\right)$$

$$\min_{\mathbf{w}\in\mathbb{R}^d} \underbrace{||\mathbf{w}||^2}_{\text{regularization}} + C \sum_i^N \underbrace{\max\left(0, 1 - y_i f(\mathbf{x}_i)\right)}_{\text{loss function}}$$

C = Infinity   hard margin



C = 10   soft margin



C的取值和带宽