给定数据集

$$D=\left\{\left(\overrightarrow{x_1},y_1\right),\left(\overrightarrow{x_2},y_2\right),\ldots,\left(\overrightarrow{x_m},y_m\right)\right\},y_i\in\{-1,+1\}$$

划分超平面：

$$\vec{w}^T\vec{x}+b=0 \tag{6.1}$$

正确分类的划分超平面的约束：

$$\begin{cases}\vec{w}^T\vec{x}+b\geqslant +1, & y_i=+1;\\ \vec{w}^T\vec{x}+b\leqslant -1, & y_i=-1\end{cases} \tag{6.3}$$

异类支持向量的“间隔”margin：

$$\gamma=\frac{2}{\|\vec{w}\|} \tag{6.4}$$

要使间隔最大：

$$\max_{\vec{w},b}=\frac{1}{2}\left\|\vec{w}\right\|^2 \tag{6.5}$$

$$s.t.\ y_i\left(\vec{w}\vec{x}_i+b\right)\geq 1,\ i=1,2,\ldots,m$$

等价于(SVM基本型）：

$$\min_{\vec{w},b}\frac{1}{2}\|\vec{w}\|^2 \tag{6.6}$$

$$s.t.\ y_i\left(\vec{w}^T\vec{x}+b\right)\geqslant 1,\ i=1,2,..,m$$

解得超平面模型：

$$f(x)=\vec{w}^T\vec{x}+b \tag{6.7}$$

SVM基本型的拉格朗日对偶问题：

$$SVM\equiv\min_{\vec{w},b}(\max_{\alpha>0}\ Ł\left(\vec{w},b,\vec{\alpha}\right)$$

$$Ł\left(\vec{w},b,\vec{\alpha}\right)=\frac{1}{2}\|\vec{w}\|^2+\sum_{i=1}^{m}\alpha_i\left(1-y_i\left(\vec{w}^T\vec{x}+b\right)\right) \tag{6.8}$$

求偏导：

$$\vec{w}=\sum\alpha_i y_i x_i \tag{6.9}$$

$$0=\sum\alpha_i y_i \tag{6.10}$$

代入6.8得到对偶问题：

$$\max_{\vec{\alpha}}\sum\alpha_i - \frac{1}{2}\sum\sum\alpha_i\alpha_j\, y_i y_j\, x_i^T x_j \tag{6.11}$$
$$s.t.\ \sum\alpha_i y_i = 0,$$
$$\alpha_i \geqslant 0,\ i = 1, 2,\ldots, m$$

原样本空间非线性可分，于是将样本映射到高维空间，高维特征空间中的划分超平面模型：

$$f\left(\vec{x}\right) = \vec{w}^{\,T}\phi\left(\vec{x}\right) + b \tag{6.19}$$

特征空间中SVM基本型：

$$\min_{\vec{w},\, b} = \frac{1}{2}\|\vec{w}\|^2 \tag{6.20}$$
$$s.t.\ \ y_i f\left(\vec{x}\right) = y_i\left(\vec{w}^T\phi\left(\vec{x}\right) + b\right) \geqslant 1,\ i = 1, 2,\ldots, m$$

其对偶问题为：

$$\max_{\vec{\alpha}}\ \sum\alpha_i - \frac{1}{2}\sum\sum a_i a_j\, y_i y_j\, \phi\left(\vec{x_i}\right)^T\phi\left(\vec{x_j}\right) \tag{6.21}$$
$$s.t.\ \sum a_i y_i = 0,$$
$$\alpha_i \geqslant 0, i = 1, 2,\ldots, m$$

定义核函数，其计算在低维空间$\mathcal{X}$：

$$\kappa\left(\vec{x_i}, \vec{x_j}\right) = \phi\left(\vec{x_i}\right)^T\phi\left(\vec{x_j}\right)$$

于是，式（6.21）可写为：

$$\max_{\vec{\alpha}} = \sum\alpha_i - \frac{1}{2}\sum\sum\alpha_i\alpha_j\, y_i y_j\, \kappa(\vec{x_i}, \vec{x_j}) \tag{6.23}$$
$$s.t.\ \sum\alpha_j y_j = 0,$$
$$\alpha_i \geqslant 0,\ i = 1, 2,\ldots, m$$

求解后即可得到模型（支持向量展式）：

$$f\left(\vec{x}\right) = \vec{w}^{\,T}\phi\left(\vec{x}\right) + b$$

$$= \sum\alpha_i\, y_i\, \phi\left(\vec{x_i}\right)^T\phi\left(\vec{x_j}\right) + b$$

$$= \sum\alpha_i\, y_i\, \kappa\left(\vec{x_i}, \vec{x_j}\right) + b \tag{6.24}$$

常用核函数的选择：

$$线性核：\kappa\left(\vec{x_i}, \vec{x_j}\right) = \vec{x_i}\vec{x_j}$$

$$多项式核：\kappa\left(\vec{x_i}, \vec{x_j}\right) = \left(\vec{x_i}\vec{x_j}\right)^d, d \geqslant 1$$

$$高斯核 / RBF核：exp\left(-\frac{\|\vec{x_i} - \vec{x_j}\|^2}{2\sigma^2}\right), \ \sigma > 0$$

$$拉普拉斯核：exp\left(-\frac{\|\vec{x_i} - \vec{x_j}\|}{\sigma}\right), \ \sigma > 0$$

软间隔SVM，允许某些样本不满足约束$y_i\left(\vec{w}^T\vec{x_i} + b\right) \geqslant 1$，优化目标可写为：

$$\min_{\vec{w}, b} \frac{1}{2}\|\vec{w}\|^2 + C\sum \ell_{0/1}\left(y_i\left(\vec{w}\vec{x_i} + b\right) - 1\right) \tag{6.29}$$

$C > 0$是正则化常数，$\ell_{0/1}$是“0 / 1损失函数”：$\ell_{0/1}(z) = \begin{cases} 1, & if\ z < 0; \\ 0, & otherwise \end{cases}$

当C无穷大时，若要实现目标使最小化（间隔最大化），只有所有样本均满足约束，此时6.29等价于6.6；所以，若要允许一些样本不满足约束，C必须取有限值。

损失函数的“替代损失”：

$$\begin{gathered} hinge损失：\ell_{hinge}(z) = max(0, 1 - z); \\ 指数损失：\ell_{exp}(z) = exp(-z); \\ logistic\ loss: \ell_{log}(z) = log(1 + exp(-z)) \end{gathered}$$

若采用hinge损失，式（6.29）变为：

$$\min_{\vec{w}, b} = \frac{1}{2}\|\vec{w}\|^2 + C\sum max\left(0,\ 1 - y_i\left(\vec{w}\vec{x_i} + b\right)\right) \tag{6.34}$$

将损失部分定义为“松弛变量(slack variables)” $\xi_i \geqslant 0$,
软间隔SVM可重写为：

$$\begin{gathered} \min_{\vec{w}, b} = \frac{1}{2}\|\vec{w}\|^2 + C\sum \xi_i \\ s.t.\ \ y_i\left(\vec{w}^T\vec{x_i} + b\right) \geqslant 1 - \xi_i \\ \xi_i \geqslant 0,\ i = 1, 2, \ldots, m \end{gathered} \tag{6.35}$$

每个样本都有一个对应的松弛变量ξ_i，用以表征该样本不满足约束的程度。
若样本被正确划分，即满足约束，那么该样本对应的$\xi_i = 0$，上式等于SVM基本型(6.6)

（6.35）的的拉格朗日函数：

$$Ł\left(\vec{w}, b, \vec{\alpha}, \vec{\xi}, \vec{\mu}\right) = \frac{1}{2}\|\vec{w}\|^2 + C\sum\xi_i + \sum\alpha_i\left(1 - \xi_i - y_i\left(\vec{w}^T\vec{x_i} + b\right)\right) + \sum\mu_i\,\xi_i \tag{6.36}$$

求偏导代入，得到（6.35）的对偶问题：

$$\max_{\vec{\alpha}} \sum\alpha_i - \frac{1}{2}\sum\sum\alpha_i\alpha_j\, y_i y_j\, \vec{x_i}^T\vec{x_j} \tag{6.40}$$

$$s.t. \sum\alpha_i\, y_i = 0,$$
$$0 \leqslant \alpha_i \leqslant C,\ i = 1, 2, \ldots, m$$

与（6.11）唯一差别是约束不同