由于不知道特征映射的形式，
也就是不知道原本特征的样子，所
以我们并不知道经过怎样的映射，也就是选择怎样
的核函数，我们并不知道怎样的核函数是合适的。

III

软间隔

允许某些样本不满足约束：$y_i(w^Tx_i+b)\geq 1$

优化目标：$\min\limits_{w,b} \frac{1}{2}\|w\|^2 + C\sum\limits_{i=1}^{m} l_{0/1}\left(y_i(w^Tx_i+b)-1\right)$

当 C 无穷大，所有样本都应满足 $y_i(w^Tx_i+b)\geq 1$

当 C 取有限值，允许一些样本不满足约束

$$l_{0/1}(z) = \begin{cases} 1, & z<0 \\ 0, & z\geq 0 \end{cases}$$

$l_{0/1}$ 非凸，不连续，数学性质不好

hinge 损失：$l_{hinge}(z) = \max(0, 1-z)$

指数损失：$l_{exp}(z) = \exp(-z)$

对率损失：$l_{log}(z) = \log(1+\exp(-z))$

采用 hinge 损失：$\min\limits_{w,b} \frac{1}{2}\|w\|^2 + C\sum\limits_{i=1}^{m} \max(0, 1-y_i(w^Tx_i+b))$

引入"松弛变量" $\varepsilon_i \geq 0$ s.t. $\varepsilon_i \geq 1-y_i(w^Tx_i+b)$

使用拉格朗日函数：

$f(x) = \frac{1}{2}\|w\|^2 + C\sum\limits_{i=1}^{m}\varepsilon_i$

$g_1(x) = 1-y_i(w^Tx_i+b)-\varepsilon_i \leq 0$

$g_2(x) = \varepsilon_i \geq 0 \quad -\varepsilon_i \leq 0$

$L(w,b,\alpha,\varepsilon,\mu) = f(x) + \alpha_i g_1(x) + \mu_i g_2(x)$

$= \frac{1}{2}\|w\|^2 + C\sum\limits_{i=1}^{m}\varepsilon_i + \sum\limits_{i=1}^{m}\alpha_i(1-y_i(w^Tx_i+b)-\varepsilon_i) - \sum\limits_{i=1}^{m}\mu_i\varepsilon_i$

其中 $\alpha_i \geq 0$，$\mu_i \geq 0$ 是拉格朗日乘子

令 $L(w,b,\alpha,\varepsilon,\mu)$ 对 w，b，ε_i 的偏导为 0

$w = \sum\limits_{i=1}^{m}\alpha_i y_i x_i$

$0 = \sum\limits_{i=1}^{m}\alpha_i y_i$

$C = \alpha_i + \mu_i$

对偶问题，代入得：

$$\begin{cases} \max\limits_{\alpha} \sum\limits_{i=1}^{m}\alpha_i - \frac{1}{2}\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j \\ \text{s.t.} \sum\limits_{i=1}^{m}\alpha_i y_i = 0 \end{cases}$$