SVM的options求解：

两个目标：样本分对；最大化Margin（最小化 w乘以w的转置 ）

样本是两类：+1，-1（标签），+1的样本必须wx+b>=1，才是将样本分对。

- Correctly classify all data points:

$$w\cdot x_i + b \geq 1 \quad if\ y_i = +1$$
$$w\cdot x_i + b \leq -1 \quad if\ y_i = -1$$
$$y_i(w\cdot x_i + b) - 1 \geq 0$$

- Maximize the margin:

$$\max M = \frac{2}{\|w\|} \Rightarrow \min \frac{1}{2} w^T w$$

- Quadratic Optimization Problem
  - Minimize $\Phi(w) = \frac{1}{2} w^t w$
  - Subject to $y_i(w\cdot x_i + b) \geq 1$

最核心的优化函数

拉格朗日乘数法：（拉格朗日系数：α）

$$L_P \equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l}\alpha_i y_i(w\cdot x_i + b) + \sum_{i=1}^{l}\alpha_i$$

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

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

Dual Problem

$$L_D \equiv \sum_i \alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i \cdot x_j$$

$$\equiv \sum_i \alpha_i - \frac{1}{2}\alpha^T H\alpha \ where\ H_{ij} = y_i y_j x_i \cdot x_j$$

Quadratic problem again!

$$subject\ to: \sum_i \alpha_i y_i = 0\ \ \&\ \ \alpha_i \geq 0$$

一个不等式，引入一个拉格朗日函数，α

分别对w、b求导，结果很重要。将结果带回，得到新的目标函数（与上面的函数是对偶问题）。原问题和对偶问题一般情况下是不等价的，但在SVM情况下满足一些条件，所以是等价的。所以转为求

L_D

的问题，其是由α组成的（有条件约束），简化了问题。

解方程，得到很多α的值。很多α都是=0的，只有少数是不等于0的，这些不等于0的是Support Vectors。因为α=0的不对w作任何contribution。

Support Vectors : Samples with positive α

$$y_s(x_s \cdot w + b) = 1$$

$$y_s(\sum_{m\in S} \alpha_m y_m x_m \cdot x_s + b) = 1$$

$$y_s^2(\sum_{m\in S} \alpha_m y_m x_m \cdot x_s + b) = y_s$$

$$b = y_s - \sum_{m\in S} \alpha_m y_m x_m \cdot x_s$$

$$b = \frac{1}{N_s}\sum_{s\in S}(y_s - \sum_{m\in S} \alpha_m y_m x_m \cdot x_s)$$

$$g(x) = \sum_{i=1}^{l} \alpha_i y_i x_i \cdot x + b$$

inner product

随便挑选一个support vector就可以将b求出来。用多个support vectors也可以，求解完累加，再除上个数就可以。

注意

$x_i . x$

是两个向量做内积，这也是SVM的精华所在，很关键。将里面主要的运算都做成向量的内积运算。

随便取了一个support vectors

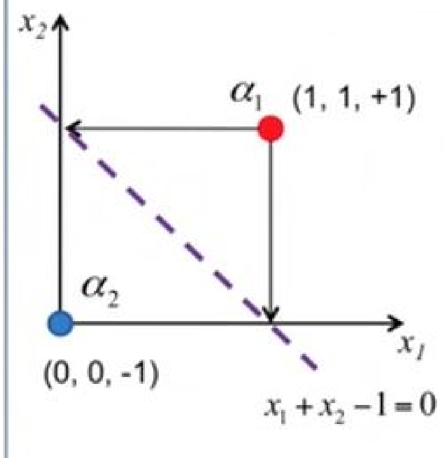
x_s

，将w替换，将方程展开，两边同时乘上

y_s

（为+1或者-1），进行化简。

举例：



$$\sum_{i=1}^{2}\alpha_i y_i = 0 \Rightarrow \alpha_1-\alpha_2=0 \Rightarrow \alpha_1=\alpha_2$$

$$H=\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}=\begin{bmatrix} y_1y_1x_1\cdot x_1 & y_1y_2x_1\cdot x_2 \\ y_2y_1x_2\cdot x_1 & y_2y_2x_2\cdot x_2 \end{bmatrix}=\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$L_D \equiv \sum_{i=1}^{2}\alpha_i-\frac{1}{2}[\alpha_1,\alpha_2]H\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}=2\alpha_1-\alpha_1^2$$

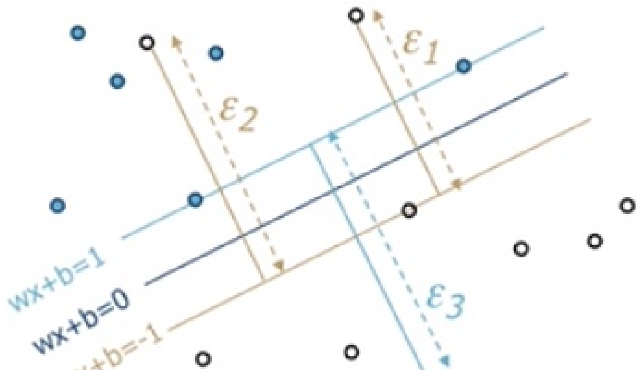
$$w=\sum_{i=1}^{2}\alpha_i y_i x_i = 1\times 1\times[1,1]+1\times(-1)\times[0,0]=[1,1] \qquad \boxed{\alpha_1=1;\ \alpha_2=1}$$

$$b=-wx_1+1=-2+1=-1$$

$$g(x)=wx+b=x_1+x_2-1 \qquad M=\frac{2}{\|w\|}=\frac{2}{\sqrt{2}}=\sqrt{2}$$

Soft Margin（处理噪点问题，并不是处理实际上的线性不可分）

原本的情况是假设上平面就为+1，下平面就为-1。但很多情况无法做到完美分割，给你加一个允错范围。当加上它之后你可以都大于0或者小于0，就近似看为全部分对。但由于引入了允错范围，所以还要加上一个惩罚量。



$$y_i(wx_i+b)-1+\xi_i \ge 0$$

$$\Phi(w)=\frac{1}{2}w'w+C\sum_i \xi_i$$

$$\xi_i \ge 0$$

$$L_P \equiv \frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}\xi_i-\sum_{i=1}^{l}\alpha_i[y_i(w\cdot x_i+b)-1+\xi_i]-\sum_{i=1}^{l}\mu_i\xi_i$$

两组不等式引入两个拉格朗日函数，α和μ

仍按原来的方式求解：

引入了一个soft margin，但最终结果并没有很复杂。

$$\left.\begin{aligned}\frac{\partial L_p}{\partial w} &= 0 \Rightarrow w = \sum_{i=1}^{l} \alpha_i y_i x_i \\ \frac{\partial L_p}{\partial b} &= 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0\end{aligned}\right\} \text{Same as before}$$

$$\frac{\partial L_p}{\partial \xi_i} = 0 \Rightarrow C = \alpha_i + \mu_i$$

$$L_P \equiv \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l} \alpha_i[y_i(w \cdot x_i + b) - 1 + \xi_i] - \sum_{i=1}^{l} \mu_i \xi_i$$

$$\boxed{L_D \equiv \sum_i \alpha_i - \frac{1}{2}\alpha^T H \alpha \quad s.t.\, 0 \le \alpha_i \le C \quad and \quad \sum_i \alpha_i y_i = 0}$$

发现与原来的类似，只有<=C不同