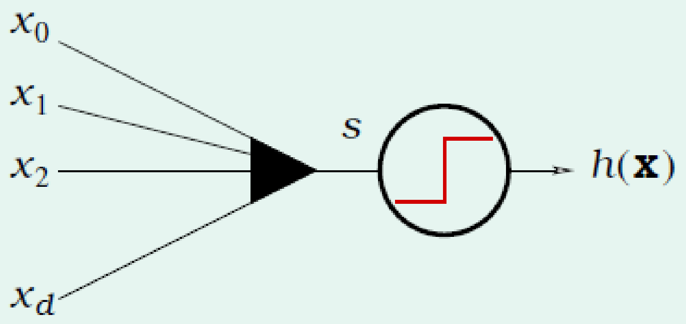
如何选择一条线呢？

$$h(\mathbf{x}) = \text{sign}(s)$$



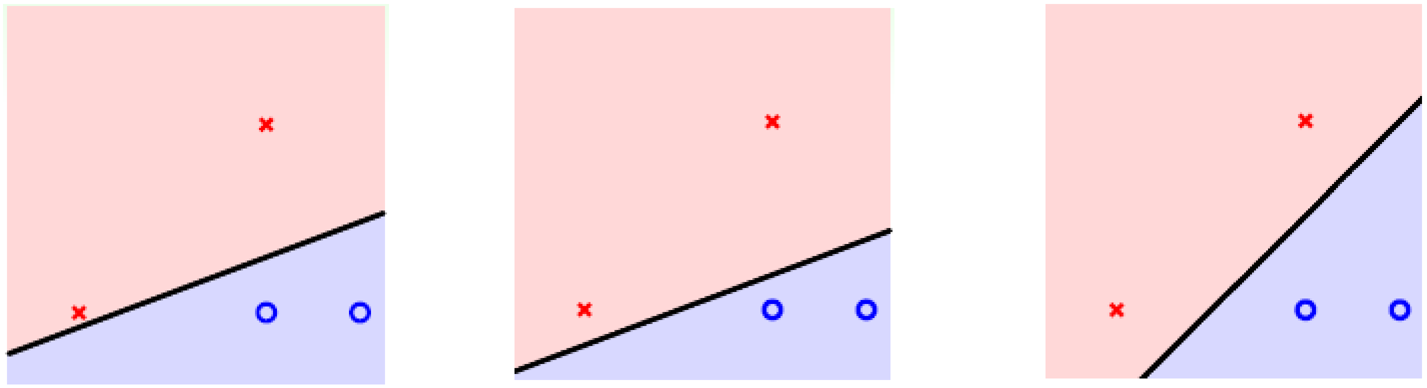
plausible err = 0/1

(small flipping noise)

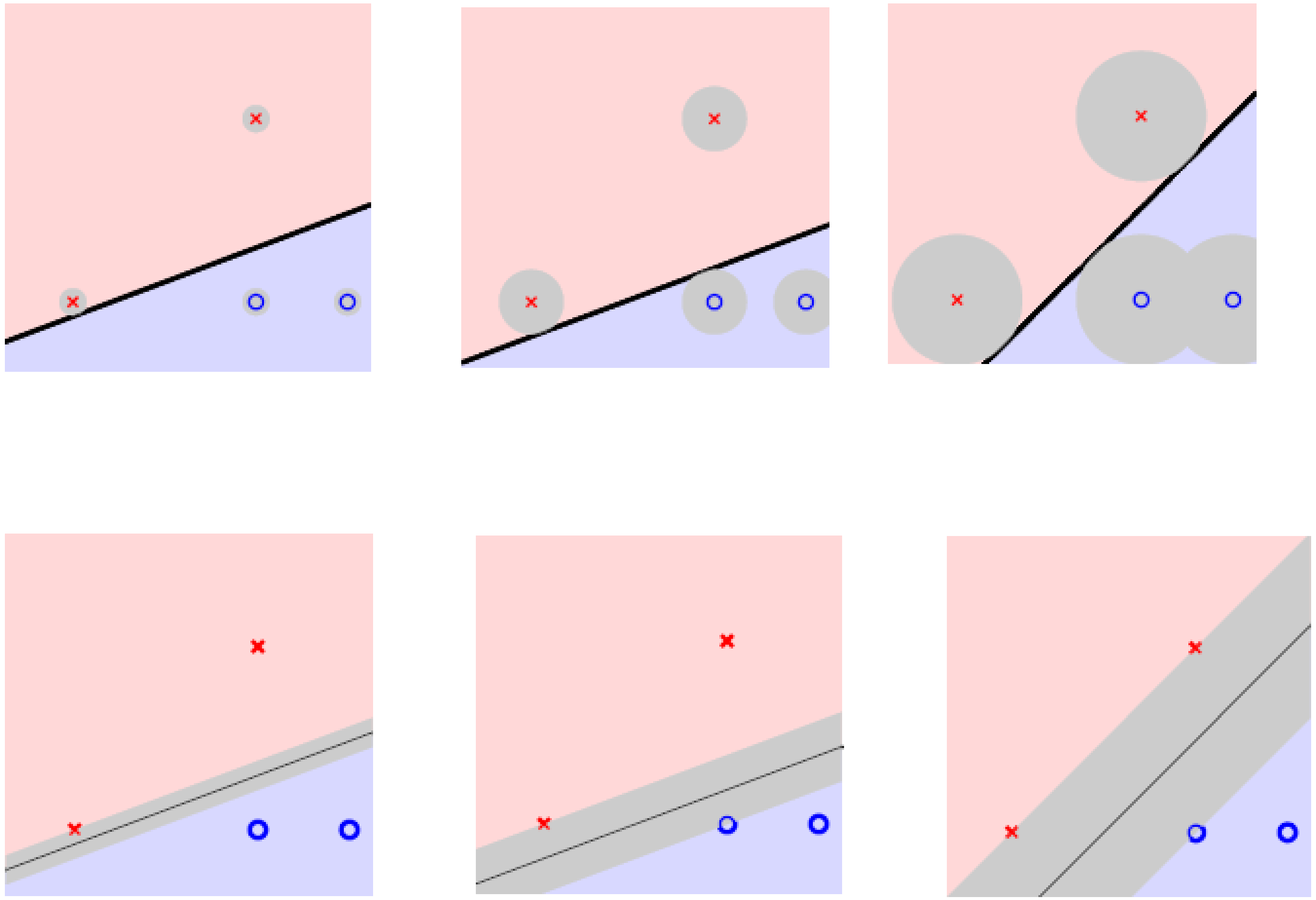
minimize **specially**

linear (hyperplane) classifiers:

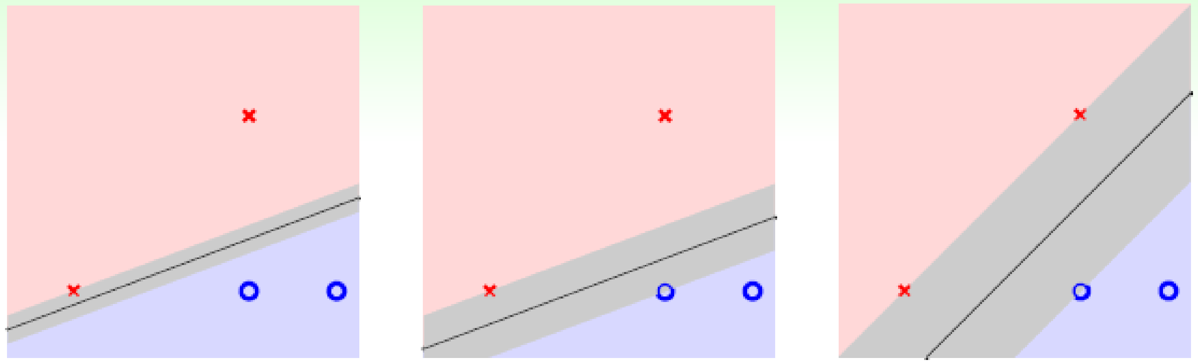
$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x})$$



越胖越好



目的



$$\max_{\mathbf{w}} \quad \text{fatness}(\mathbf{w})$$

subject to $\quad \mathbf{w}$ classifies every $(\mathbf{x}_n, y_n)$ correctly

$$\text{fatness}(\mathbf{w}) = \min_{n=1,\ldots,N} \text{distance}(\mathbf{x}_n, \mathbf{w})$$

- fatness: formally called **margin**
- correctness: $y_n = \text{sign}(\mathbf{w}^T\mathbf{x}_n)$

goal: find **largest-margin**
separating hyperplane

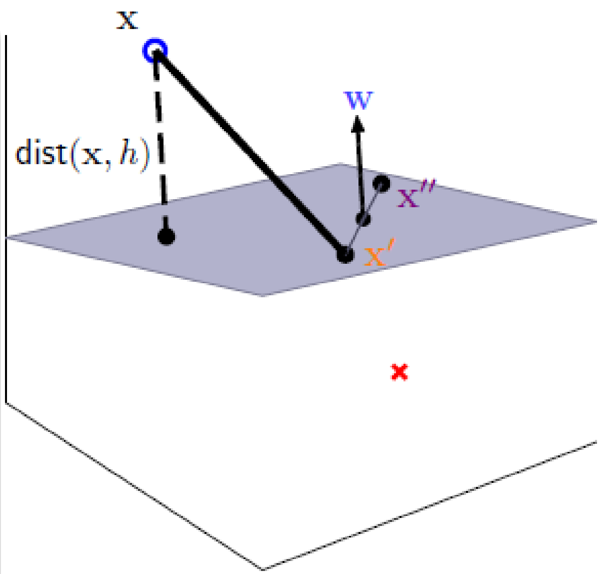
want: distance$(\mathbf{x}, b, \mathbf{w})$, with hyperplane $\mathbf{w}^T\mathbf{x}' + b = 0$

consider $\mathbf{x}'$, $\mathbf{x}''$ on hyperplane

1. $\mathbf{w}^T\mathbf{x}' = -b$, $\mathbf{w}^T\mathbf{x}'' = -b$
2. $\mathbf{w} \perp$ hyperplane:

$$\left(\mathbf{w}^T \underbrace{(\mathbf{x}'' - \mathbf{x}')}_{\text{vector on hyperplane}}\right) = 0$$

3. distance = project $(\mathbf{x} - \mathbf{x}')$ to $\perp$ hyperplane



$$\text{distance}(\mathbf{x}, b, \mathbf{w}) = \left|\frac{\mathbf{w}^T}{\|\mathbf{w}\|}(\mathbf{x} - \mathbf{x}')\right| \overset{1}{=} \frac{1}{\|\mathbf{w}\|}|\mathbf{w}^T\mathbf{x} + b|$$

数据集(X1,Y1)(X2,Y2)到(Xn,Yn)

$$y(x) = w^T \Phi(x) + b$$

Y为样本的类别：

当X为正例时候 Y = +1

当X为负例时候 Y = -1

$$y(x_i) > 0 \Leftrightarrow y_i = +1$$

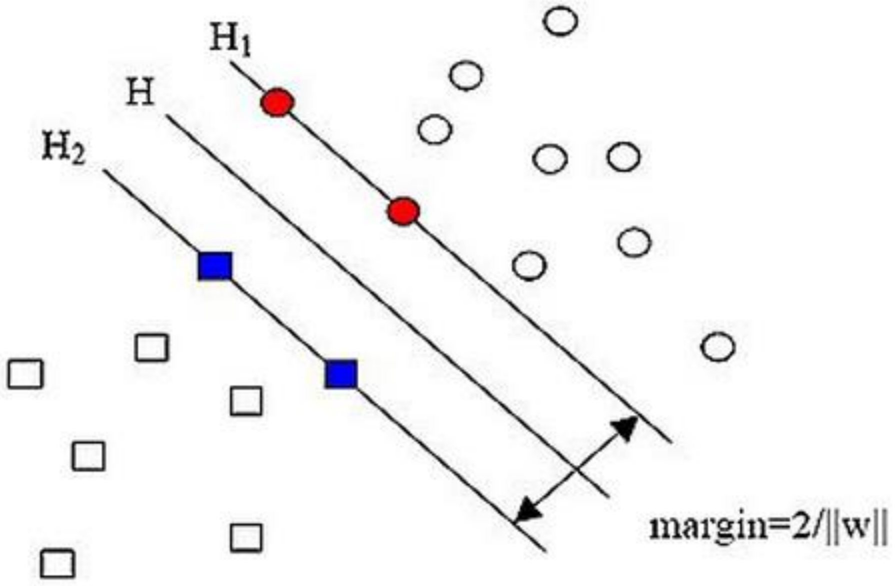
$$y(x_i) < 0 \Leftrightarrow y_i = -1$$

可推出 $y_i \cdot y(x_i) > 0$

找到一个条线（w和b），使得离该线最近的点能够最远

argmax(w,b)使得min(最近的点到该线的距离)

$$\frac{y_i \cdot \left(w^T \cdot \Phi(x_i) + b\right)}{\|w\|}$$



对于线（w,b）可以通过放缩使得其结果值|Y|>= 1

$$y_i \cdot \left(w^T \cdot \Phi(x_i) + b\right) \geq 1$$

$$\arg\max_{w,b} \left\{ \frac{1}{\|w\|} \min_i \left[y_i \cdot \left(w^T \cdot \Phi(x_i) + b\right) \right] \right\}$$

$$\arg\max_{w,b} \frac{1}{\|w\|}$$ (搞定目标函数)

目标函数：$max_{w,b}\ \frac{1}{||W||}$且$y_i\left(w^T\cdot\Phi(x_i)+b\right)\geq 1$

转换成求最小值$min_{w,b}\ \frac{1}{2}w^2$且 $y_i\left(w^T\cdot\Phi(x_i)+b\right)\geq 1$

拉格朗日乘子法标准格式：

$$\begin{aligned}&\min && f(x)\\ &\text{s.t.} && g_i(x)\leq 0, \quad i=1,\dots,m\end{aligned}$$

拉格朗日乘子法

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

对偶问题：

$$\min_{w,b}\max_{\alpha}L(w,b,\alpha)\ \text{->}\ \max_{\alpha}\min_{w,b}L(w,b,\alpha)$$

分别对w和b求偏导,分别得到两个条件

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n} \alpha_i y_i \Phi(x_n)$$

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

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

$$=\frac{1}{2}w^Tw-w^T\sum_{i=1}^{n}\alpha_i y_i\Phi(x_i)-b\sum_{i=1}^{n}\alpha_i y_i+\sum_{i=1}^{n}\alpha_i$$

$$=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}(\sum_{i=1}^{n}\alpha_i y_i\Phi(x_i))^T\sum_{i=1}^{n}\alpha_i y_i\Phi(x_i)$$

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

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

$\Rightarrow$ 完成了第一步求解 $\min_{w,b} L(w,b,\alpha)$

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

继续对 α 求极大值 $\max\limits_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \left(\Phi(x_i)\cdot\Phi(x_j)\right)$

条件： $\sum_{i=1}^{n} \alpha_i y_i = 0$

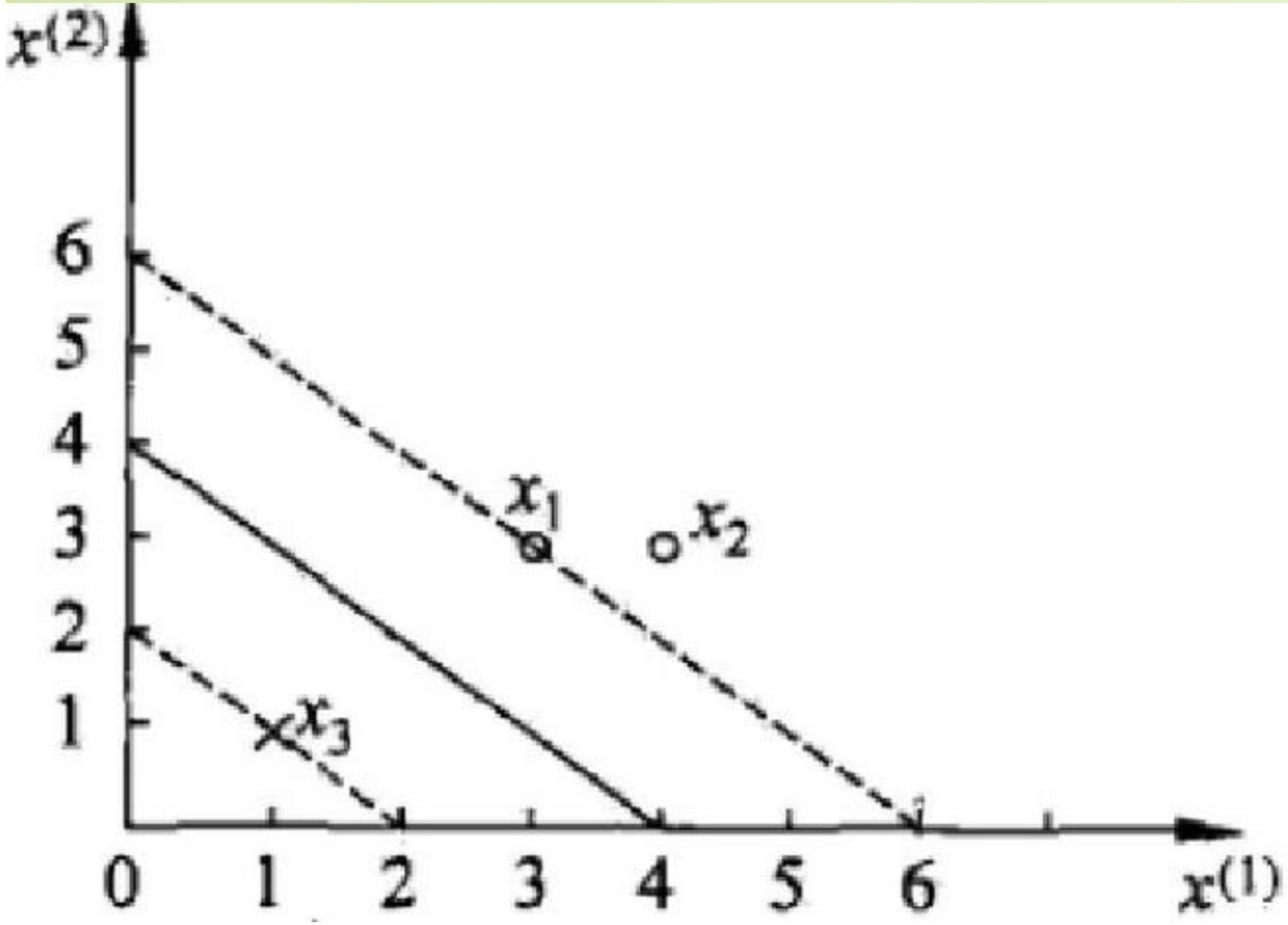
$\alpha_i \geq 0$

极大值转换成求极小值 ➡ $\min\limits_{\alpha} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \left(\Phi(x_i)\cdot\Phi(x_j)\right) - \sum_{i=1}^{n} \alpha_i$

条件： $\sum_{i=1}^{n} \alpha_i y_i = 0$

$\alpha_i \geq 0$

例1. 已知一个如图所示的训练数据集，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。



样本：X1(3,3,1) X2(4,3,1) X3(1,1-1)

求解：
$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right)-\sum_{i=1}^{n}\alpha_i$$
$$\alpha_1+\alpha_2-\alpha_3=0$$
$$\alpha_i\geq 0,\quad i=1,2,3$$

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

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

$$\alpha_1+\alpha_2=\alpha_3 \Rightarrow 4\alpha_1^2+\frac{13}{2}\alpha_2^2+10\alpha_1\alpha_2-2\alpha_1-2\alpha_2$$

分别对参数进行求导得： α_1 =1.5

α_2 =-1 ⇒ 不满足条件 $\alpha_i \geq 0,\quad i=1,2,3$

最终的解应该为边界上的点 ⇒ α_1 =0

α_2 =-2/13 ⇒ 带入原式=-0.153

α_1 =0.25

α_2 =0 ⇒ 带入原式=-0.25

最小值在(0.25,0,0.25)处取得

对于α值(0.25,0,0.25)

$$w^* = \sum_{i=1}^{N} \alpha_i^* y_i \Phi(x_i)$$

$$b^* = y_i - \sum_{i=1}^{N} \alpha_i^* y_i \left(\Phi(x_i) \cdot \Phi(x_j)\right)$$

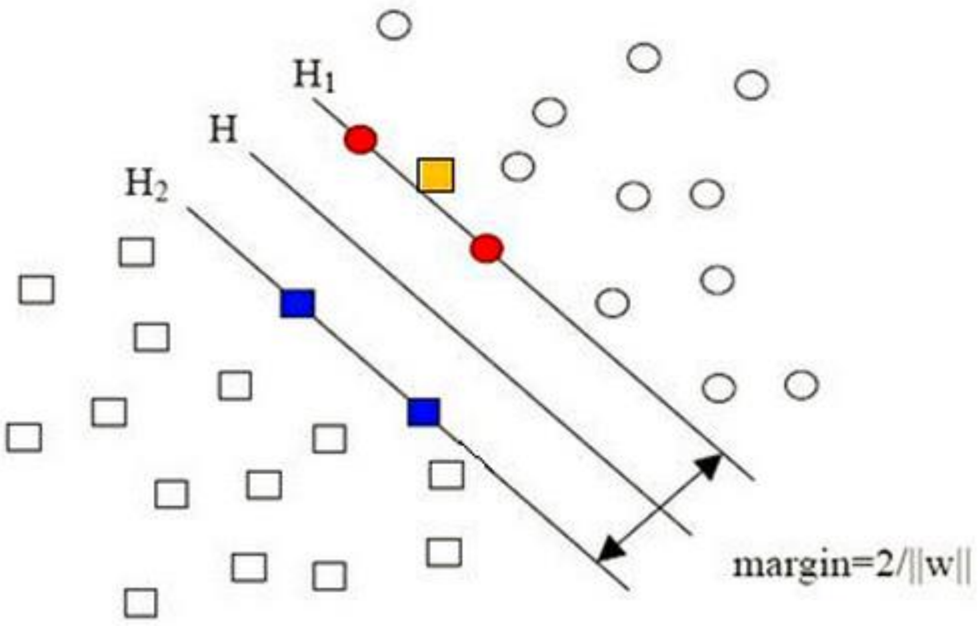
$\Rightarrow$

$w_1 = w_2 = 0.5$

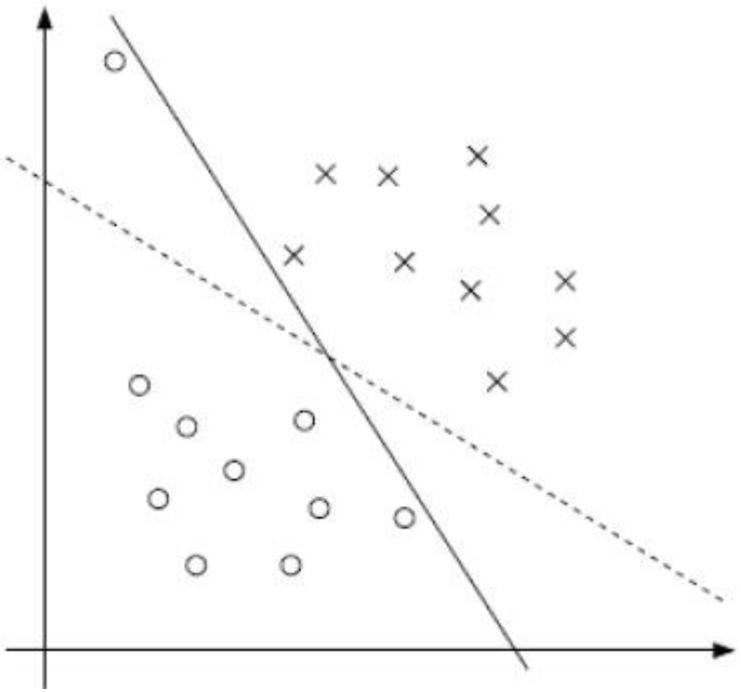
$b = -2$

$\Rightarrow$

$0.5X1 + 0.5X2 - 2 = 0$



H_1
H
H_2
margin=2/||w||



为了解决该问题，引入松弛因子

当C趋近于无穷大时：意味着分类严格不能有错误
当C趋近于很小的时：意味着可以有更大的错误容忍

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i$$

目标函数：$\min \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i$

$$L(w,b,\xi,\alpha,\mu) \equiv \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i(y_i(w \cdot x_i + b) - 1 + \xi_i) - \sum_{i=1}^{n}\mu_i\xi_i$$

$$\Rightarrow$$

$$w = \sum_{i=1}^{n}\alpha_i y_i \phi(x_n)$$

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

$$C - \alpha_i - \mu_i = 0$$

带入原式：

$$-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right)+\sum_{i=1}^{n}\alpha_i$$

仍然求对偶问题

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

$$C-\alpha_i-\mu_i=0$$

$$\alpha_i\geq 0$$

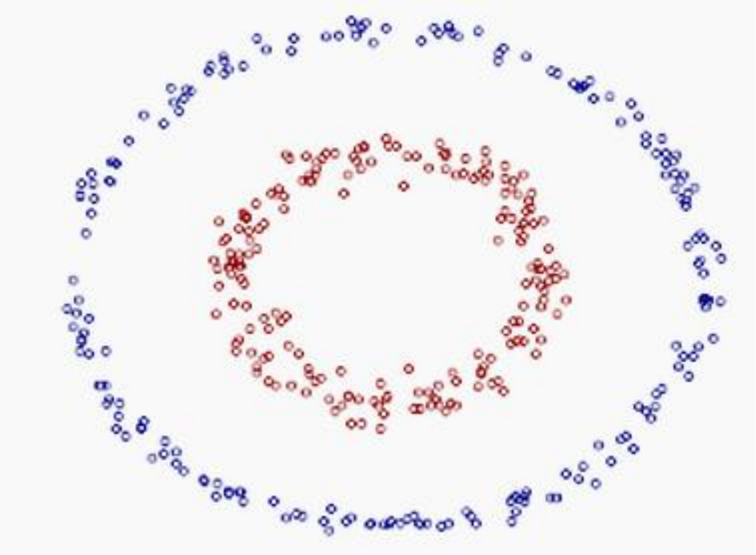
$$\mu_i\geq 0$$

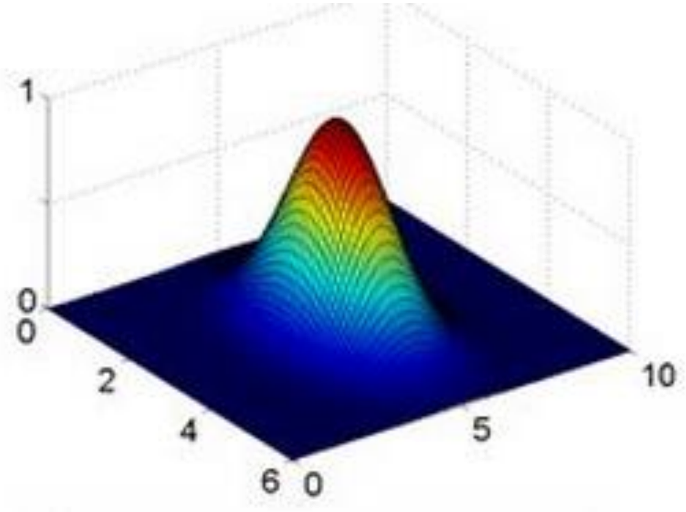
$$0\leq\alpha_i\leq C$$

$$\min_{\alpha}\ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right)-\sum_{i=1}^{n}\alpha_i$$

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

$$0\leq\alpha_i\leq C$$




1
0
0
2
4
6 0
5
10

还是先从一个小例子来阐述问题。假设我们有俩个数据，x = (x1, x2, x3); y = (y1, y2, y3)，此时在3D空间已经不能对其经行线性划分了，那么我们通过一个函数将数据映射到更高维的空间，比如9维的话，那么 f(x) = (x1x1, x1x2, x1x3, x2x1, x2x2, x2x3, x3x1, x3x2, x3x3)，由于需要计算内积，所以在新的数据在9维空间，需要计算<f(x),f(y)>的内积，需要花费O（n^2）。

在具体点，令x = (1, 2, 3); y = (4, 5, 6)，那么f(x) = (1, 2, 3, 2, 4, 6, 3, 6, 9)，f(y) = (16, 20, 24, 20, 25, 36, 24, 30, 36)，

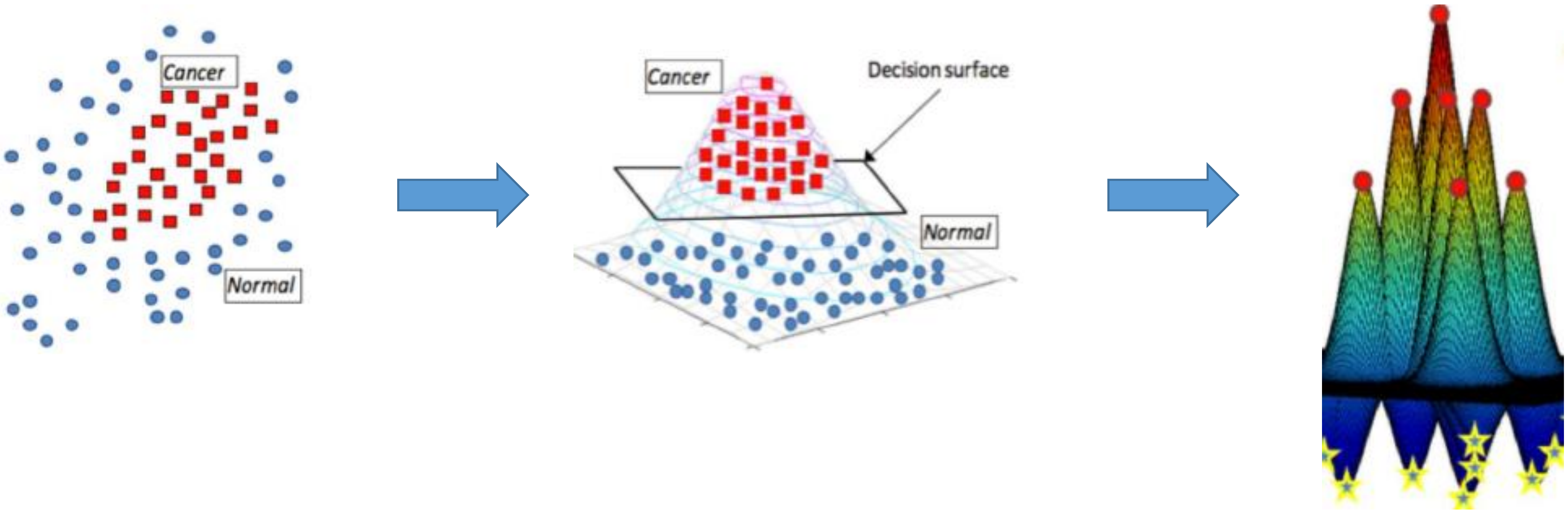
此时<f(x), f(y)> = 16 + 40 + 72 + 40 + 100+ 180 + 72 + 180 + 324 = 1024

似乎还能计算，但是如果将维数扩大到一个非常大数时候，计算起来可就不是一丁点问题了。

但是发现， K(x, y) = (<x, y>)^2

K(x,y)=(4 + 10 + 18) ^2 = 32^2 = 1024

俩者相等，K(x, y) = (<x, y>)^2=<f(x), f(y)>，但是 K(x, y) 计算起来却比<f(x), f(y)>简单的多，也就是说只要用K(x, y)来计算，，效果和<f(x), f(y)>是一样的，但是计算效率却大幅度提高了，如：K(x, y)是O（n），而<f(x),f(y)>是O（n^2）.所以使用核函数的好处就是，可以在一个低维空间去完成高维度（或者无限维度）样本内积的计算，比如K(x,y)=(4 + 10 + 18) ^2 的3D空间对比<f(x), f(y)> = 16 + 40 + 72 + 40 + 100+ 180 + 72 + 180 + 324 的9D空间。



高斯核函数：

$$K(X, Y) = \exp\left\{-\frac{\|X - Y\|^2}{2\sigma^2}\right\}$$