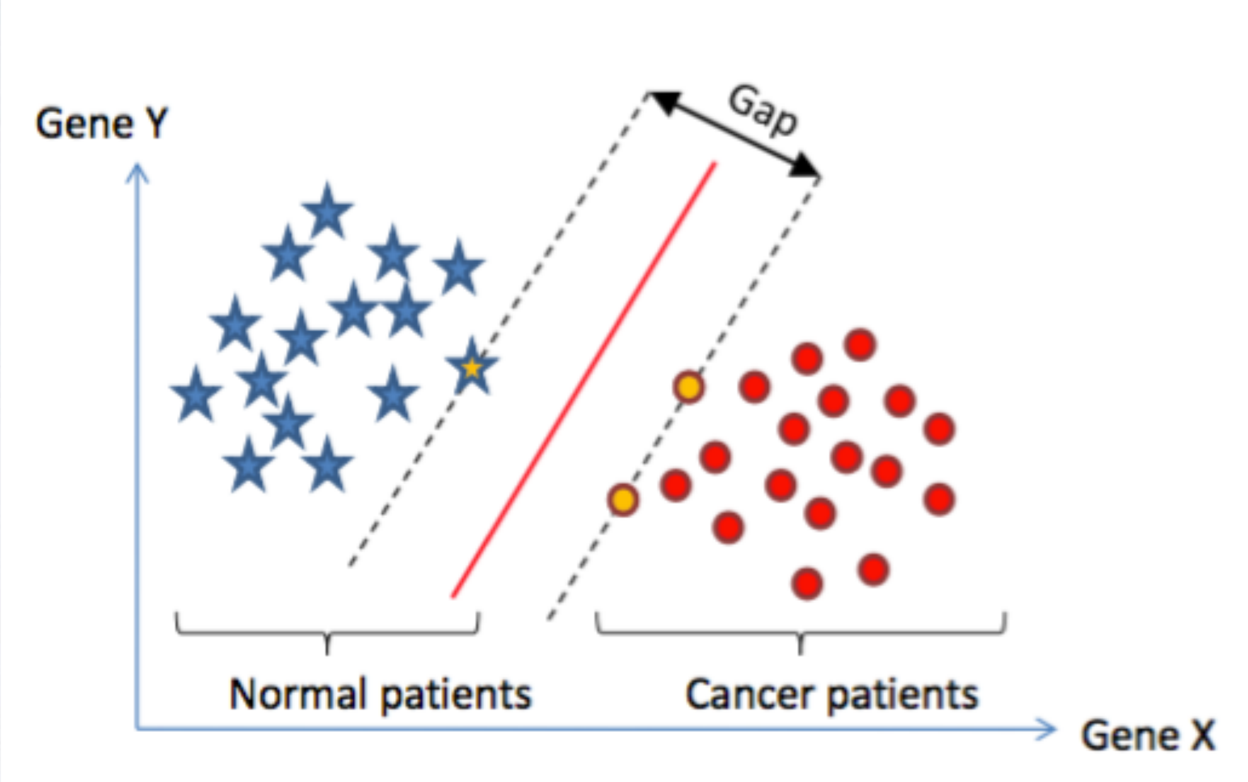
支持向量机－SVM

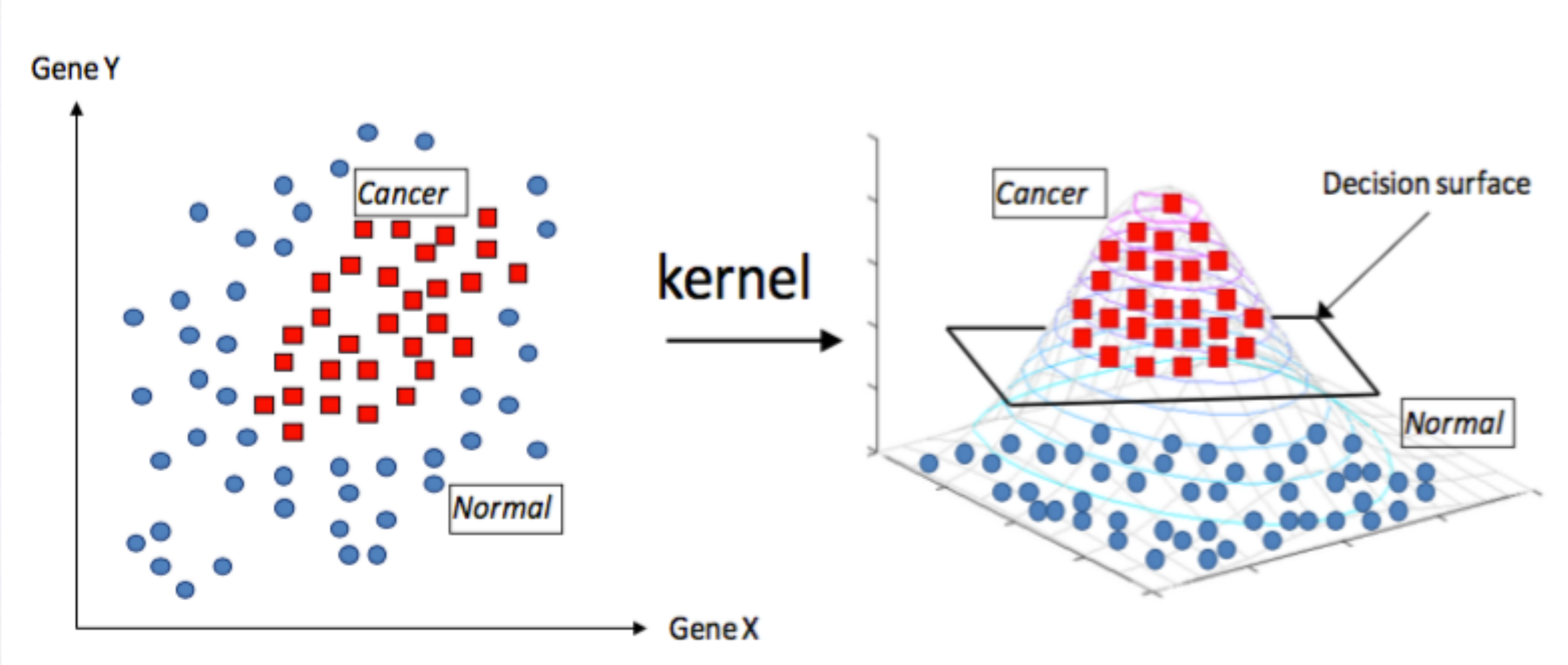
七月算法

SVM中最重要的内容

- 最大“间隔”的方法，寻找支撑向量

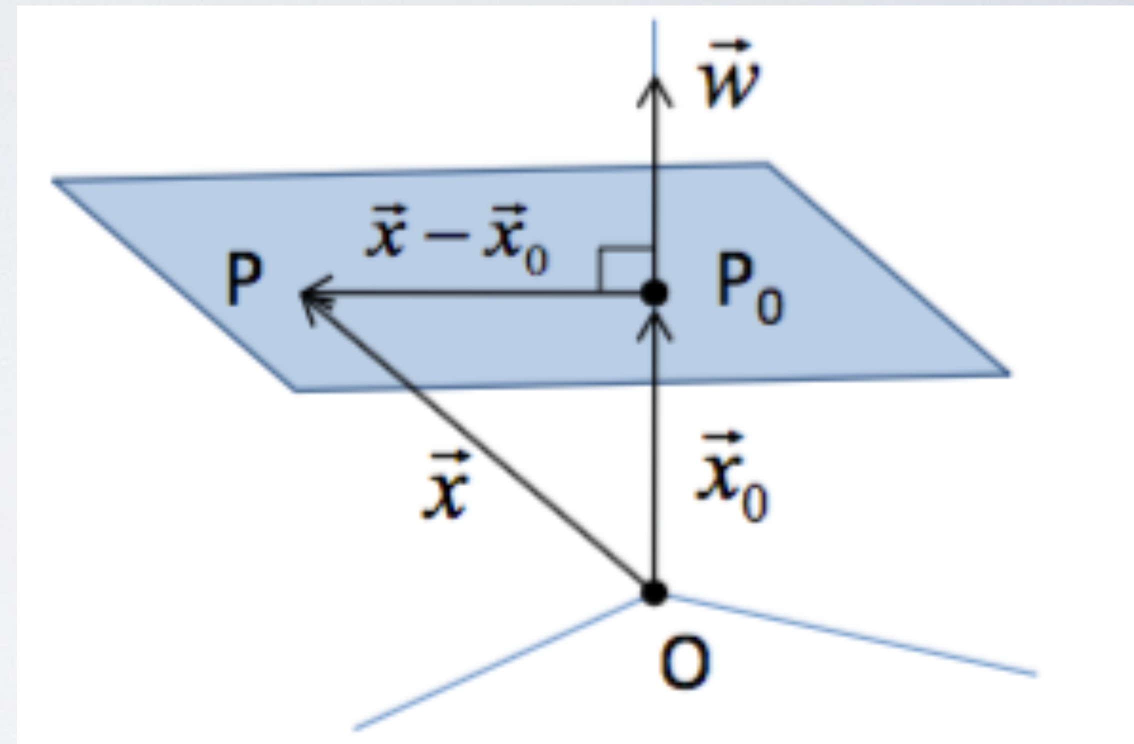


- 引入“核函数”，解决线性不可分



准备一点数学：定义一个平面

- R^3空间里，一个平面可以由平面上一个点P0，以及一个垂直平面的方向W向量确定
- 任意取平面上一个点P，从原点到P，P0，做两个向量x，x0，由于W垂直平面的关系可以得到右边的定义
- 因为P这个点是平面任意指定的点，因此平面可以用wx+b这种形式表达



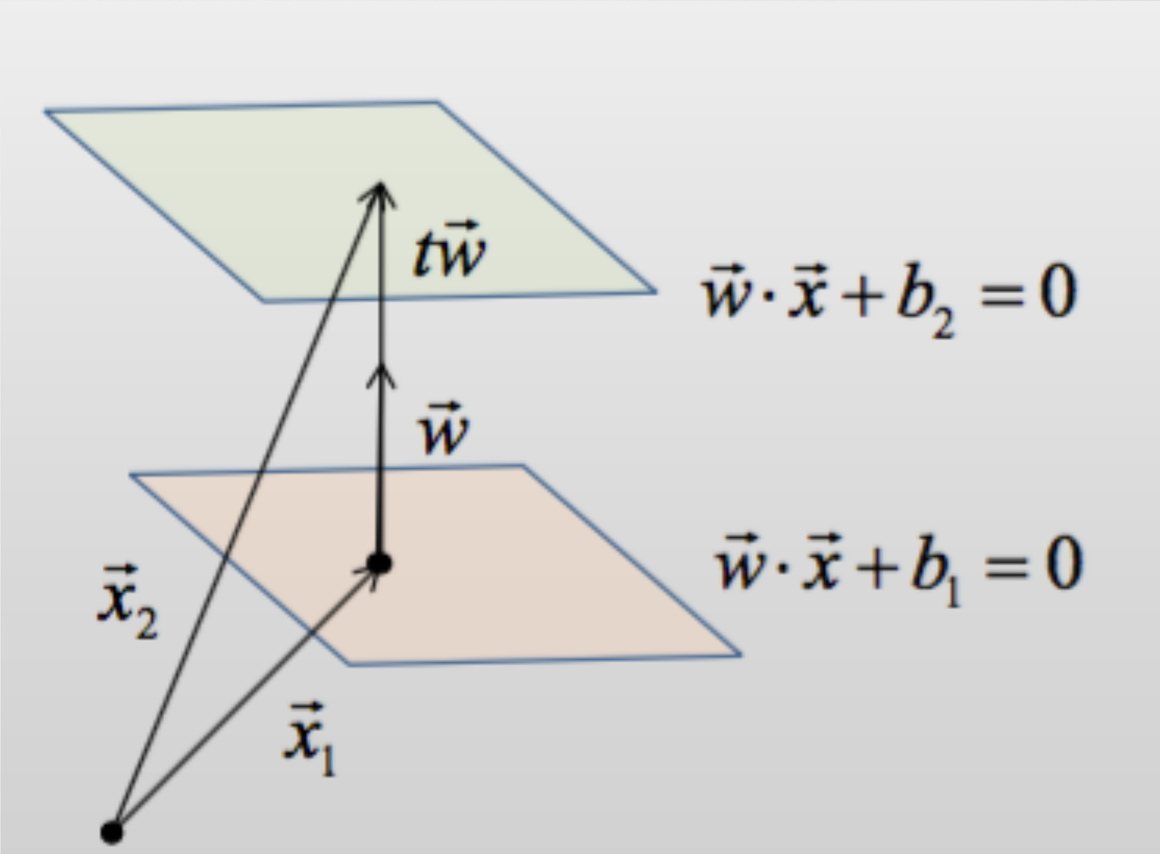
$\vec{w}\cdot(\vec{x}-\vec{x}_0)=0$ or

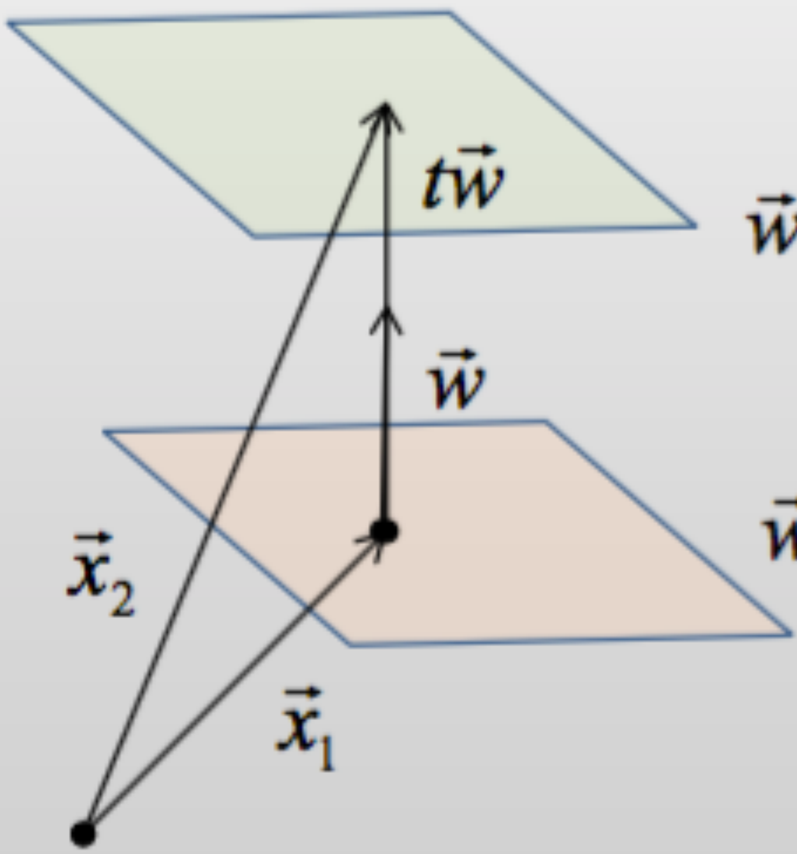
$\vec{w}\cdot\vec{x}-\vec{w}\cdot\vec{x}_0=0$ define $b=-\vec{w}\cdot\vec{x}_0$

$\boxed{\vec{w}\cdot\vec{x}+b=0}$

两个平行平面的距离

- 在平面1上任意上指定一个点x1，我们按w方向找到x2, x2-x1垂直与两个平面
- 我们设定x2-x1=tw，因为x2-x1方向就是w, 不过长度需要重新计算，因此两个平面的距离为|tw|
- 按下一页计算可以得到，使用b来表示的两个平面的距离公式





$\vec{w}\cdot\vec{x}+b_2=0$

$\vec{w}\cdot\vec{x}+b_1=0$

$$\vec{x}_2=\vec{x}_1+t\vec{w}$$

$$D=\|t\vec{w}\|=|t|\|\vec{w}\|$$

$$\vec{w}\cdot\vec{x}_2+b_2=0$$

$$\vec{w}\cdot(\vec{x}_1+t\vec{w})+b_2=0$$

$$\vec{w}\cdot\vec{x}_1+t\|\vec{w}\|^2+b_2=0$$

$$(\vec{w}\cdot\vec{x}_1+b_1)-b_1+t\|\vec{w}\|^2+b_2=0$$

$$-b_1+t\|\vec{w}\|^2+b_2=0$$

$$t=(b_1-b_2)/\|\vec{w}\|^2$$

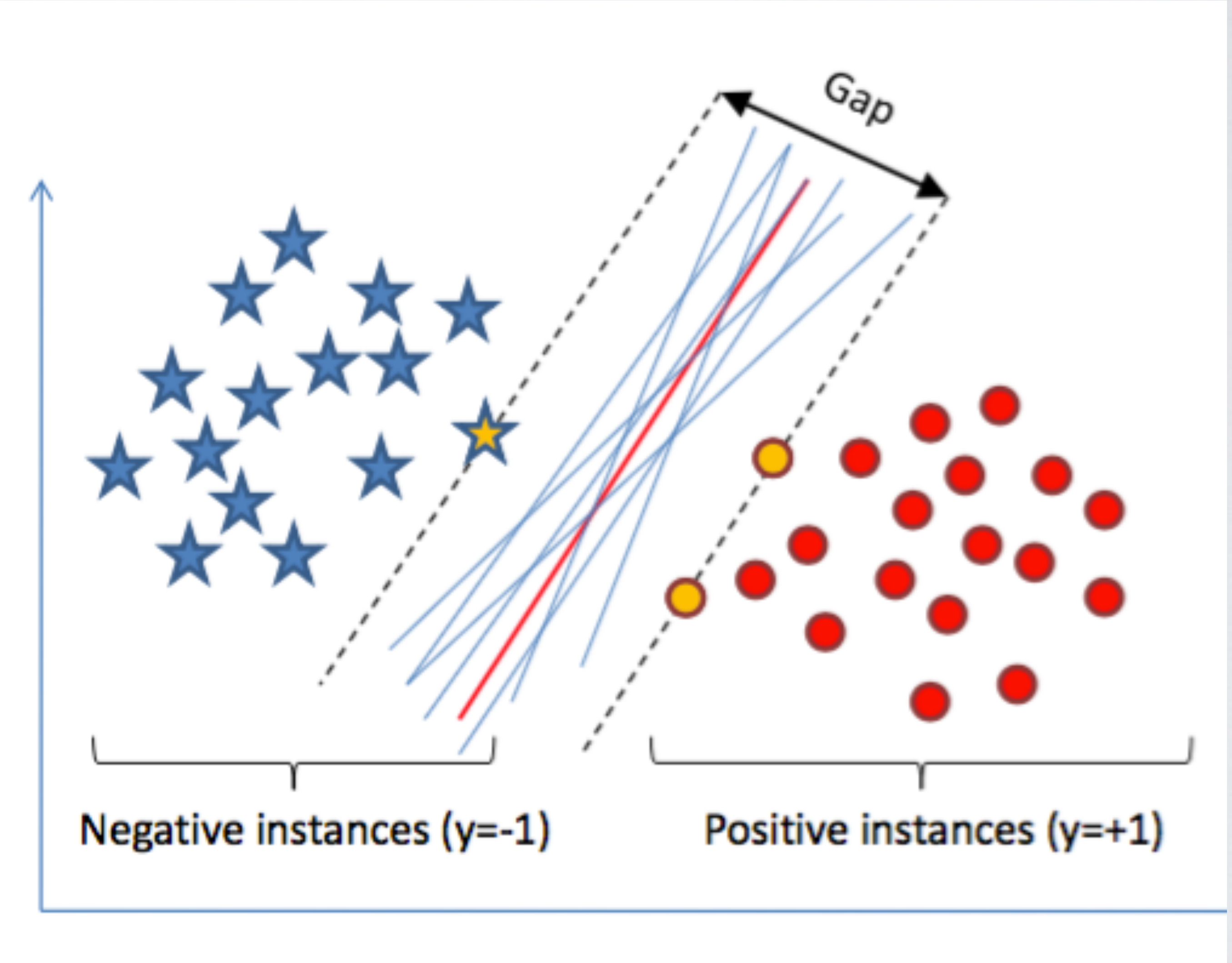
$$\boxed{\Rightarrow D=|t|\|\vec{w}\|=|b_1-b_2|/\|\vec{w}\|}$$

RECALL

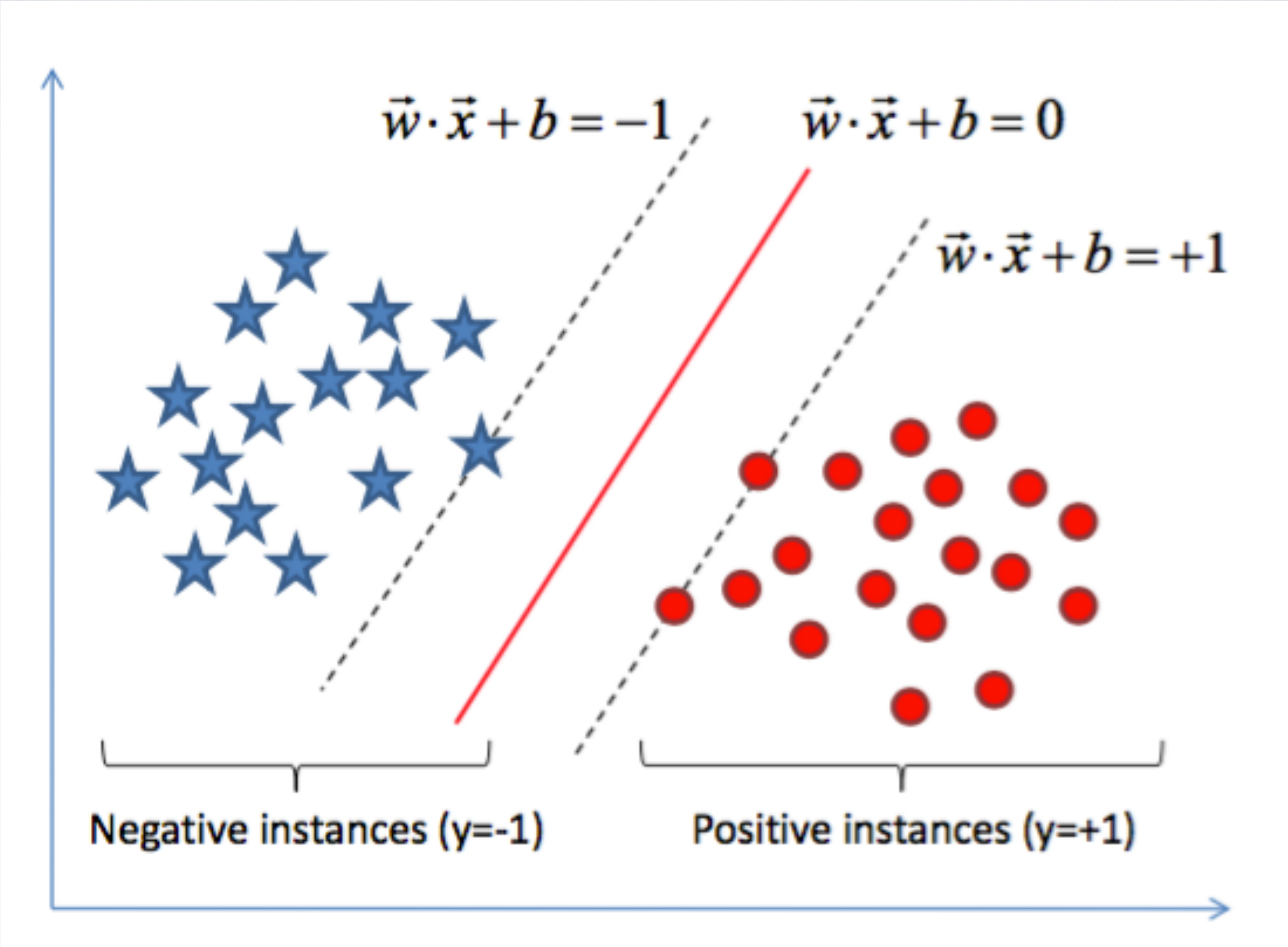
- 已经知道 平面可以用 $wx+b=0$来表示，两个平行平面的距离计算可以用w和b来计算
- 之前我们学习了凸优化

SVM问题1：线性可分的数据

- 线性可分的情况又叫做，hard-margin linear SVM
- 样本分为+1及-1两类样本
- 寻找间隔最大的平面

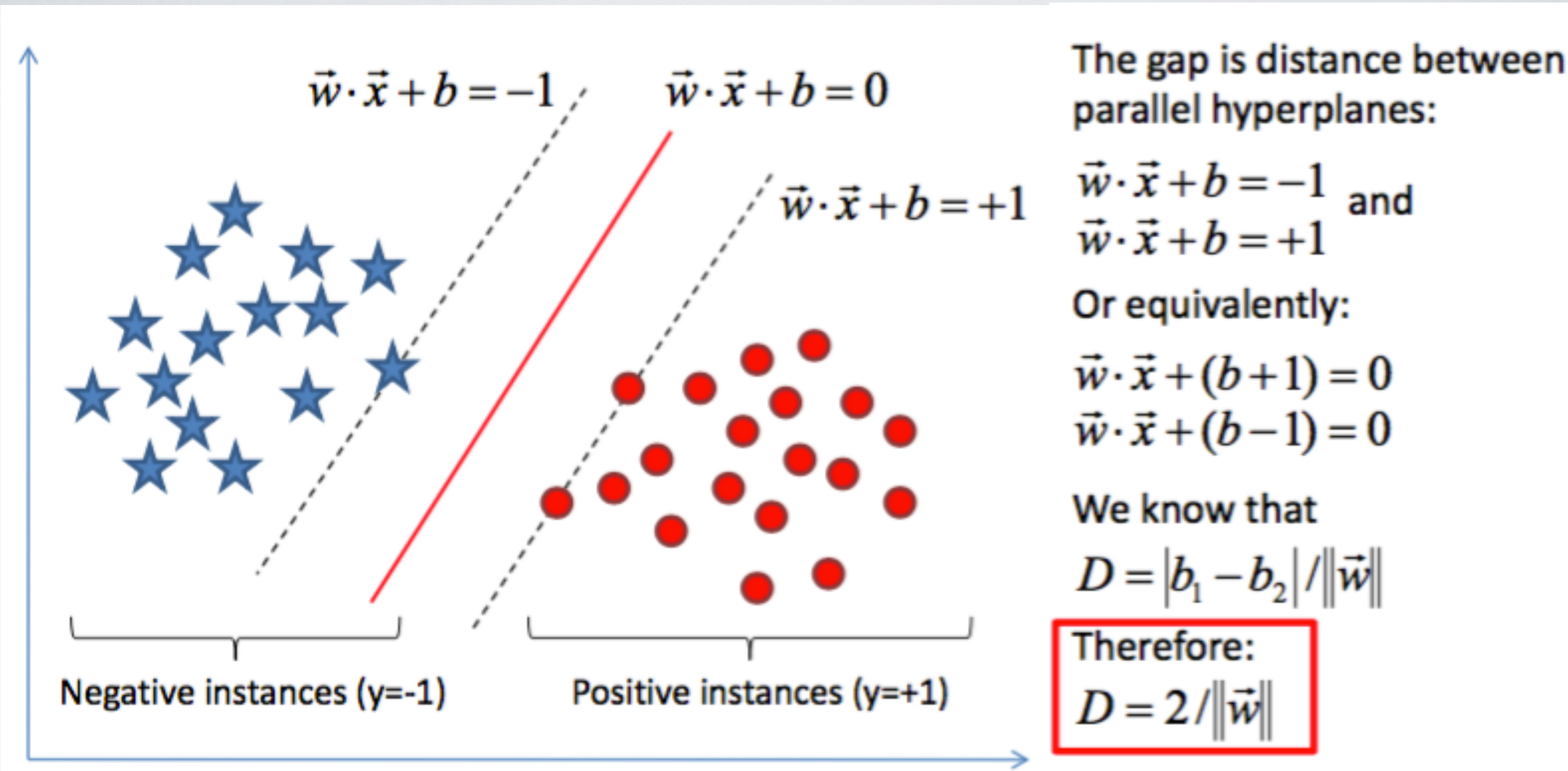


计算间隔



- 固定w的方向，对正样本来说，可以找到一个平面 $wx+b_{+1}=0$ ，并且满足所有样本都在线的一侧
- 同样对负样本也可以在相同方向，找到另外一个平面 $wx+b_{-1}=0$
- 调整w的大小，我们一定可以到得到 $b_{+1}-b_{-1}=2$ ，因此重新取 $b=b_{+1}-b_{-1}-1$
- 因此得到图中的三个平面，中间的平面可以用来预测新数据

计算最大间隔



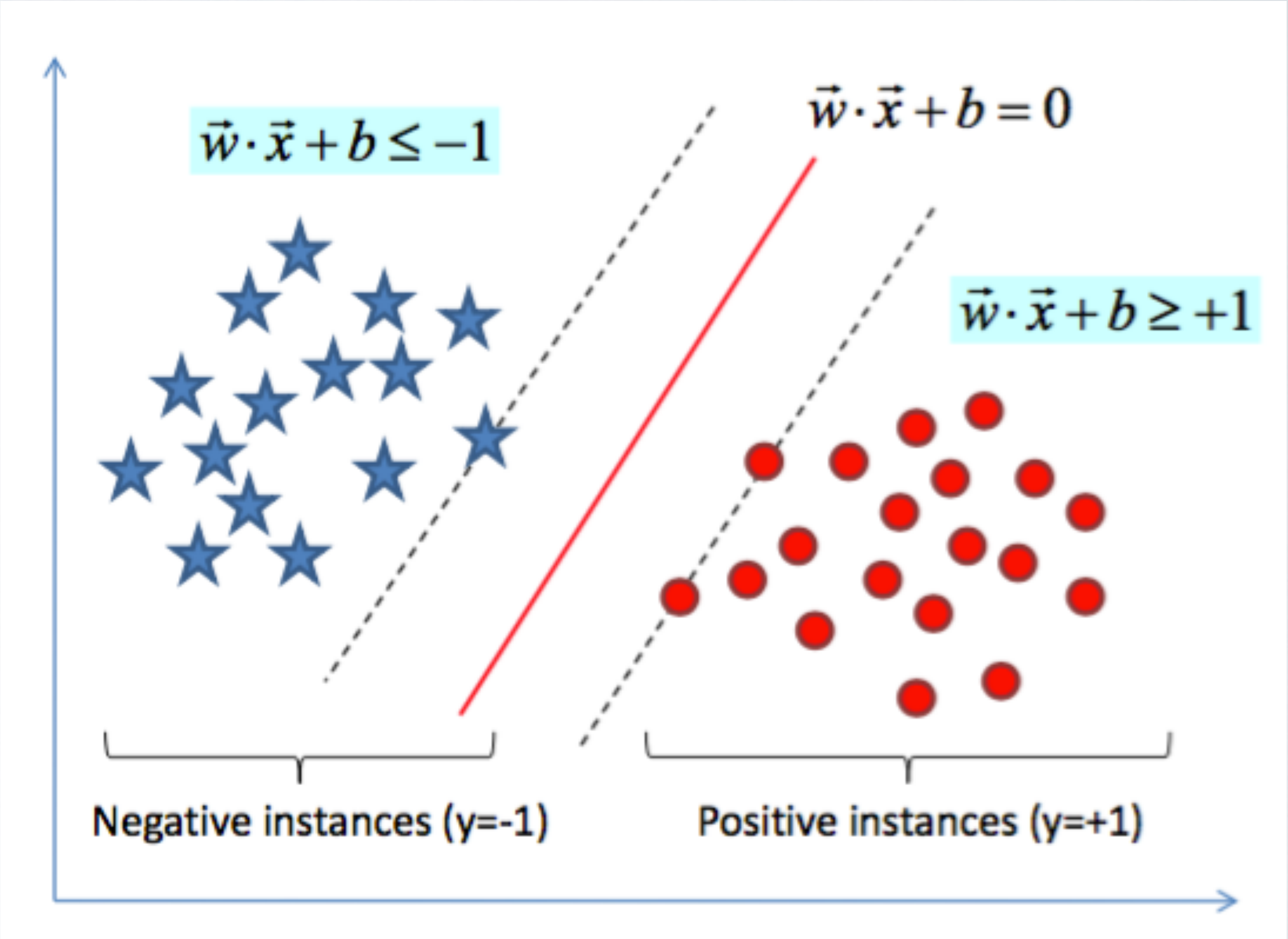
- 最大化 $\frac{2}{||w||}$, 等价于最小化 $\frac{1}{2}||w||^2$

SVM问题1：最优化目标

- 最小化 $\frac{1}{2}\|w\|^2$

- 满足约束条件：

$$y_i(w^T x_i + b) \geq 1$$



$$\vec{w}\cdot\vec{x}_i+b\leq -1 \quad \text{if} \quad y_i=-1$$
$$\vec{w}\cdot\vec{x}_i+b\geq +1 \quad \text{if} \quad y_i=+1$$

Equivalently:

$$\boxed{y_i(\vec{w}\cdot\vec{x}_i+b)\geq 1}$$

SVM问题1: 问题求解

- 这是一个二次凸优化问题，可以直接使用QP软件求解w和b
- 另外一种方法求解对偶问题

对偶问题

- 构造拉格朗日乘子

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

- 原问题是极小极大，对偶问题是极大极小

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

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

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

- 对L(w,b,a)分别对w,b求偏导，求极小值，令其＝0

$$\frac{\partial L}{\partial b}=0 \text{ -> } w=\sum_{i=1}^{n}\alpha_i y_i x_i$$

如果得到了a的值，我们可以计算w

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

得到了a一个约束条件

继续化简对偶问题

$$\begin{aligned}
\mathcal{L}(\boldsymbol{w}, b, \boldsymbol{\alpha}) &= \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left[y_i \left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i + b\right) - 1\right] \\
&= \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} - \boldsymbol{w}^{\mathrm{T}} \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i - b \sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \alpha_i \\
&= \frac{1}{2}\boldsymbol{w}^{\mathrm{T}} \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i - \boldsymbol{w}^{\mathrm{T}} \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i - b \cdot 0 + \sum_{i=1}^{n} \alpha_i \\
&= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \left(\sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i\right)^{\mathrm{T}} \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i \\
&= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j
\end{aligned}$$

重新整理对偶问题

- 最大化：

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

- s.t. $\alpha_i \geq 0, i = 1,2 \ldots n$

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

- 最小化：

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

- s.t. $\alpha_i \geq 0, i = 1,2 \ldots n$

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

SVM问题1: 重新定义

- 最小化:

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

- s.t. $\alpha_i \geq 0, i = 1,2\ldots n$

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

- 输出 W', b'

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

$$b' = y_j - w'x_j = y_j - \sum_{i=1}^{n}\alpha_i y_i x_i^T x_j \text{ , } \alpha_j \neq 0$$

$$y = sign(w'x + b')$$

SVM问题1: 重新理解

- 原问题是求解w, 其维度和x是一致的，现在求解 α ，其维度和样本数目一致的，改变了求解问题的维度

- 观察最优化目标函数，包含样本x的计算只有点积的形式，方便我们引入核函数。核函数也只能在对偶问题形式下才能引入。

SVM问题2: 线性不可分的问题

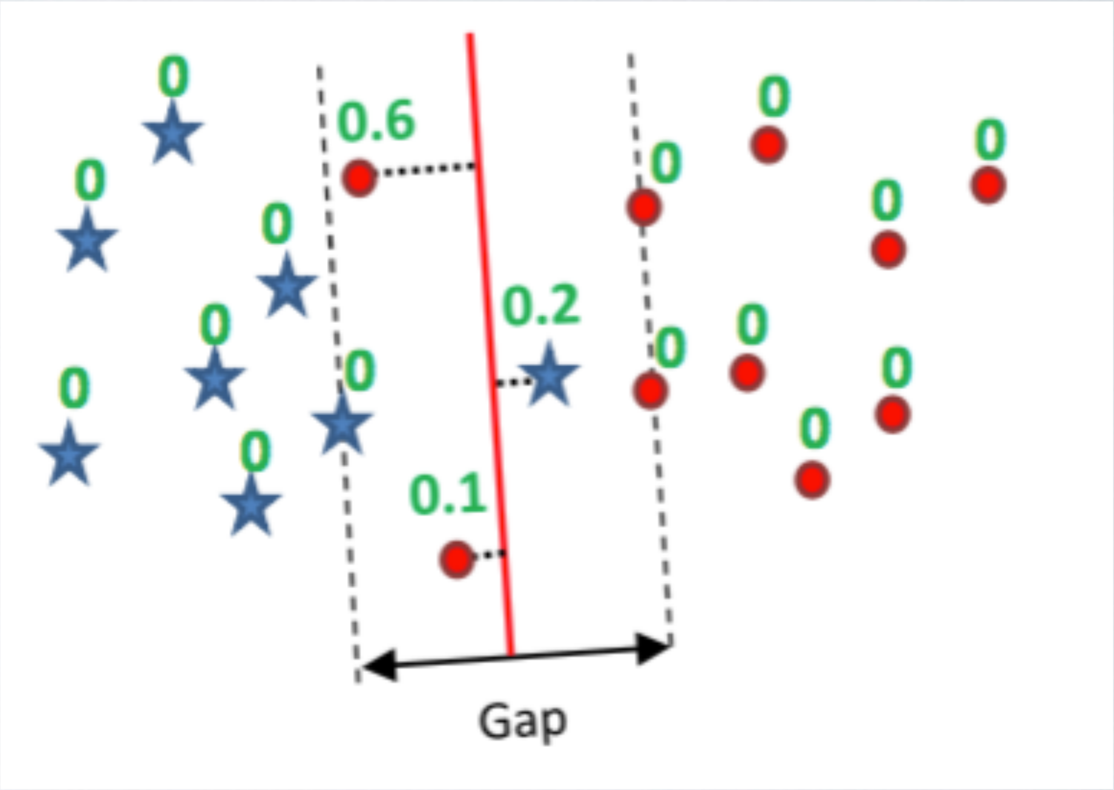
- 引入松弛变量 ζ

- 最小化

$$\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\zeta_i$$

- 满足约束条件:

$$y_i(w^t x+b)\geq 1-\zeta_i,\ \zeta_i\geq 0$$



SVM问题2: 对偶问题形式

$$L(w,b,\zeta,\alpha,\mu)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\zeta_i-\sum_{i=1}^{n}\alpha_i\{y_i(w^Tx+b)-1+\zeta_i\}-\sum_{i=1}^{n}\mu_i\zeta_i$$

求解极大极小问题，首先求偏导 w,b,ζ ,并且代入L

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}}=0 \Rightarrow \boldsymbol{w}=\sum_{i=1}^{n}\alpha_i y_i \boldsymbol{x}_i$$

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

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

$$L(w,b,\zeta,\alpha,\mu)=\frac{1}{2}||w||^2+C\sum_{i=1}^{n}\zeta_i-\sum_{i=1}^{n}\alpha_i\{y_i(w^Tx+b)-1+\zeta_i\}-\sum_{i=1}^{n}\mu_i\zeta_i$$

对偶形式，最小化问题

$$\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu)=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i$$

对上面的问题，求最大化

$$\max_{\alpha}\ -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i$$

$$\begin{aligned} s.t.\ & \sum_{i=1}^{N}\alpha_i y_i=0 \\ & C-\alpha_i-\mu_i=0 \\ & \alpha_i\geq 0 \\ & \mu_i\geq 0,\quad i=1,2,\ldots N \end{aligned}$$

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

SVM问题2: 重新定义

- 最小化：

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

- s.t. $0\le\alpha_i\le C \quad i=1,2,\ldots n$

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

- 输出 W', b'

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

$$b' = y_j - \sum_{i=1}^{n}\alpha_i y_i x_i^T x_j \ , 0<\alpha_j<C$$

$$y = sign(w'x + b')$$

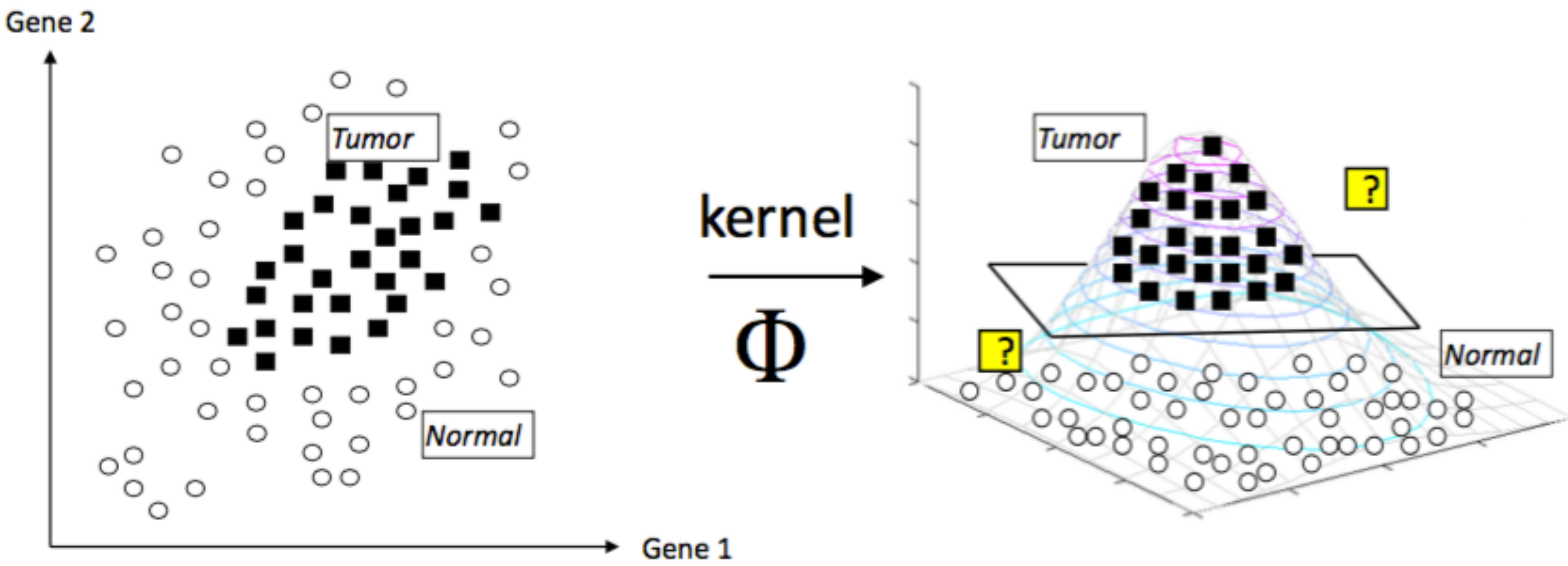
KERNEL TRICK:使用核函数

- 所有的样本计算只使用到了内积，因此我们不需要直接设计一个维度提升函数 ,只需要知道函数内积就可以了，即 $k(x_i,x_j) = R^N \times R^N \text{->} R$

- 核函数必须满足Mercer条件，否则无法用QP来求解

- 将 $x_i^T x_j$ 计算替换为 $k(x_i, x_j)$，点积也被称为线性核

通过核函数提升特征维度



Data is not linearly separable in the input space

Data is linearly separable in the feature space obtained by a kernel

$$\Phi : \mathbf{R}^N \rightarrow \mathbf{H}$$

常用的核函数

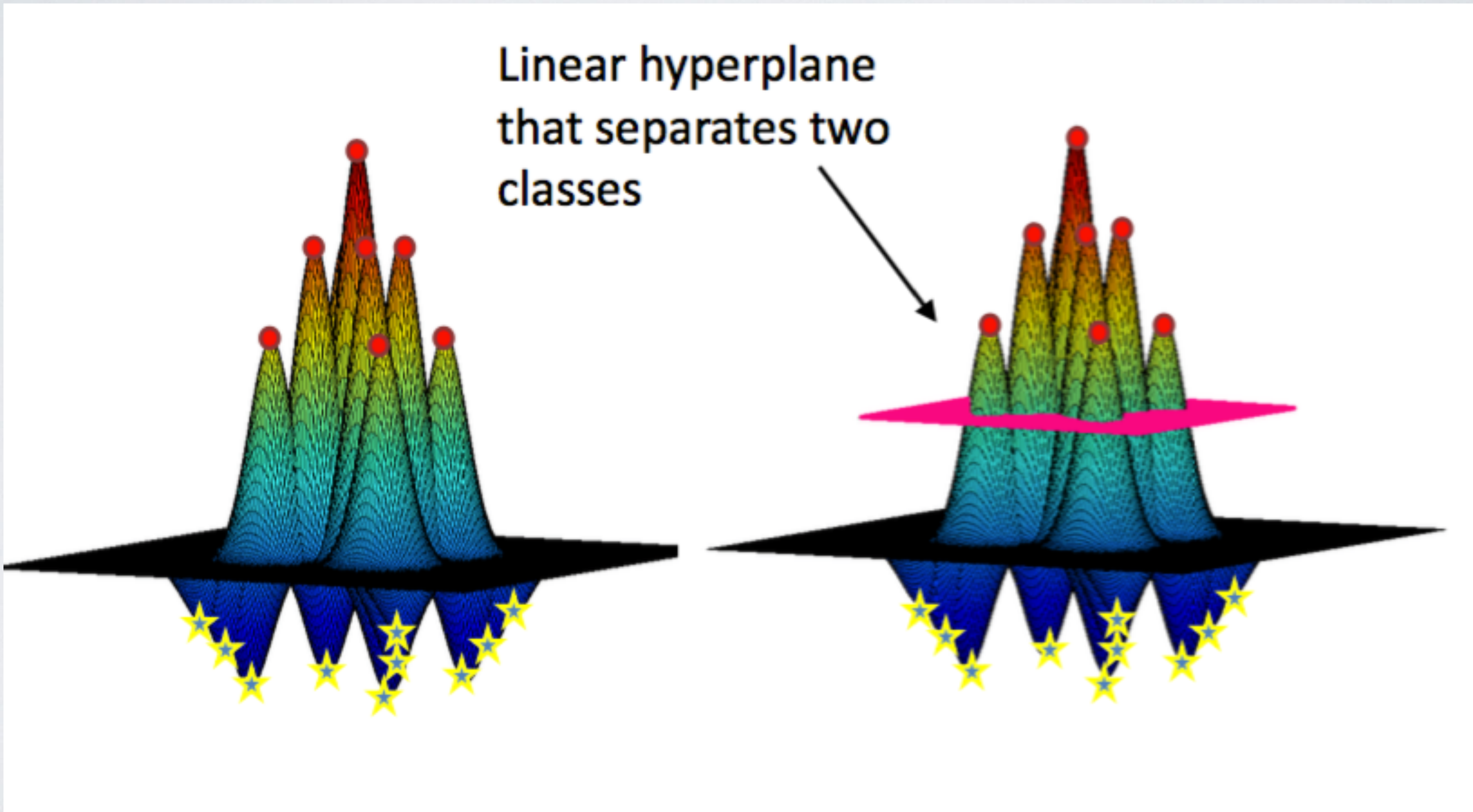
A kernel is a dot product in *some* feature space:

$$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j)$$

<u>Examples:</u>

$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i \cdot \vec{x}_j$	Linear kernel
$K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma \lVert \vec{x}_i - \vec{x}_j \rVert^2)$	Gaussian kernel
$K(\vec{x}_i, \vec{x}_j) = \exp(-\gamma \lVert \vec{x}_i - \vec{x}_j \rVert)$	Exponential kernel
$K(\vec{x}_i, \vec{x}_j) = (p + \vec{x}_i \cdot \vec{x}_j)^q$	Polynomial kernel
$K(\vec{x}_i, \vec{x}_j) = (p + \vec{x}_i \cdot \vec{x}_j)^q \exp(-\gamma \lVert \vec{x}_i - \vec{x}_j \rVert^2)$	Hybrid kernel
$K(\vec{x}_i, \vec{x}_j) = \tanh(k\vec{x}_i \cdot \vec{x}_j - \delta)$	Sigmoidal

特别的核函数：高斯核



SMO算法

- 采用标准的QP软件，直接求解
- 采用SMO算法求解，在实验中说明SMO算法

如何看待线性SVM

SVMs build the following classifiers: $f(\vec{x}) = sign(\vec{w}\cdot\vec{x}+b)$

Consider soft-margin linear SVM formulation:

Find $\vec{w}$ and b that

Minimize $\frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{N}\xi_i$ subject to $y_i(\vec{w}\cdot\vec{x}_i+b)\geq 1-\xi_i$ for $i = 1,\ldots,N$

This can also be stated as:

Find $\vec{w}$ and b that

Minimize $\underbrace{\sum_{i=1}^{N}[1-y_i f(\vec{x}_i)]_+}_{\substack{\textit{Loss} \\ \textit{("hinge loss")}}} + \underbrace{\lambda\|\vec{w}\|_2^2}_{\textit{Penalty}}$

(in fact, one can show that $\lambda = 1/(2C)$).

Flexibility of "loss + penalty" framework

Minimize (*Loss* + λ *Penalty*)

Loss function	Penalty function	Resulting algorithm
Hinge loss: $\sum_{i=1}^{N}[1-y_i f(\vec{x}_i)]_+$	$\lambda\|\vec{w}\|_2^2$	SVMs
Mean squared error: $\sum_{i=1}^{N}(y_i - f(\vec{x}_i))^2$	$\lambda\|\vec{w}\|_2^2$	Ridge regression
Mean squared error: $\sum_{i=1}^{N}(y_i - f(\vec{x}_i))^2$	$\lambda\|\vec{w}\|_1$	Lasso
Mean squared error: $\sum_{i=1}^{N}(y_i - f(\vec{x}_i))^2$	$\lambda_1\|\vec{w}\|_1 + \lambda_2\|\vec{w}\|_2^2$	Elastic net
Hinge loss: $\sum_{i=1}^{N}[1-y_i f(\vec{x}_i)]_+$	$\lambda\|\vec{w}\|_1$	1-norm SVM

SVM扩展

- 多分类的SVM算法
- SVR：支持向量回归问题
- 基于SVM的Novelty detection
- 支持向量聚类算法
- SVM分类的概率输出