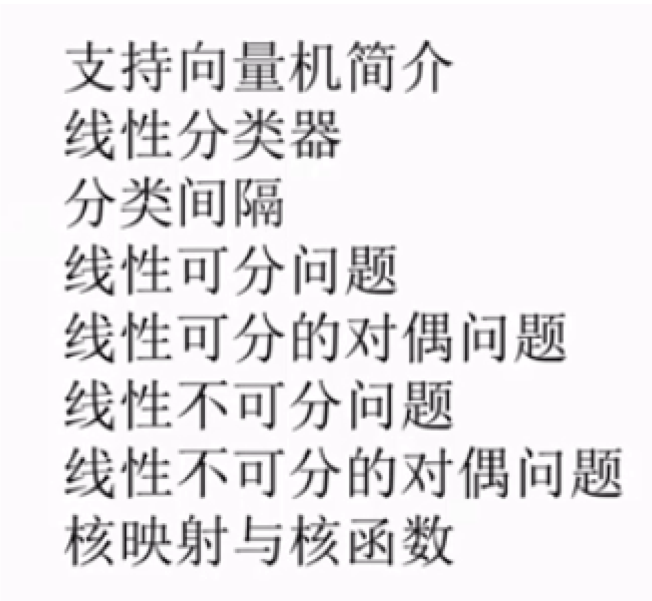
14 支持向量机

14.1 本集内容简介



14.2 支持向量机简介

支持向量机简介
由Vapnik等人提出，在出现后的20多年里它是最有影响力的机器学习算法之一
在深度学习技术出现之前，使用高斯核（RBF）的支持向量机在很多分类问题上一度取得了最好的结果
不仅可以用于分类问题，还可以用于回归问题
具有泛化性能好，适合小样本和高维特征等优点

支持向量机在机器学习里面，对数学要求比较高的。

- 对偶
- KKT

14.3 线性分类器

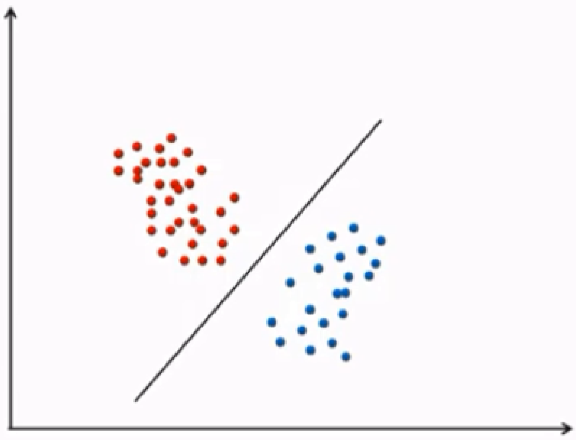
支持向量机是从线性分类器里衍生出来的。

线性分类器

二分类问题，样本标签值为+1或-1

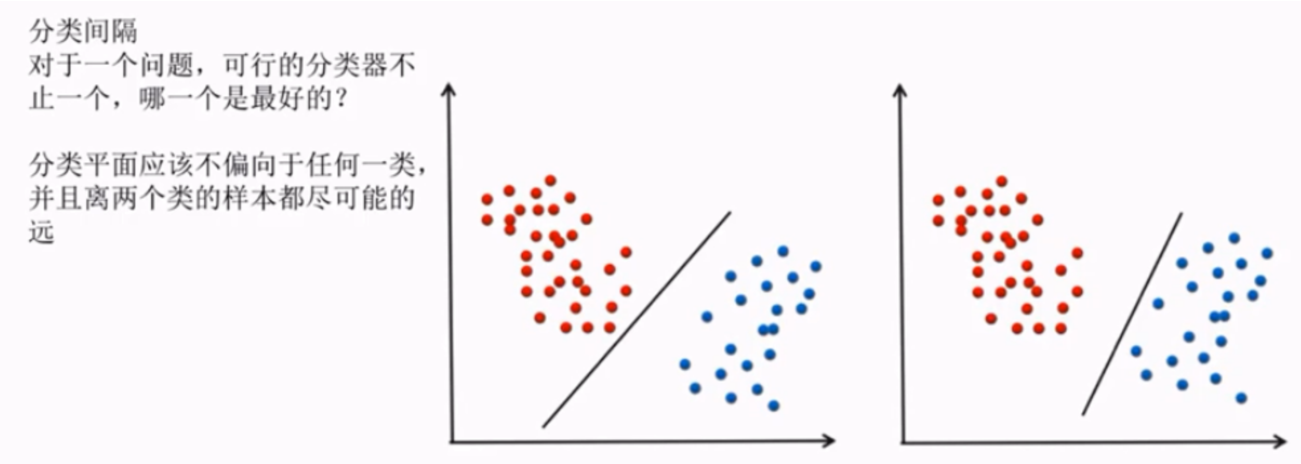
$$w^{T}x + b = 0$$
$$\operatorname{sgn}\left(w^{T}x + b\right)$$

超平面的方程只是给出了分界面，哪边为正，哪边为负，是可以灵活控制的



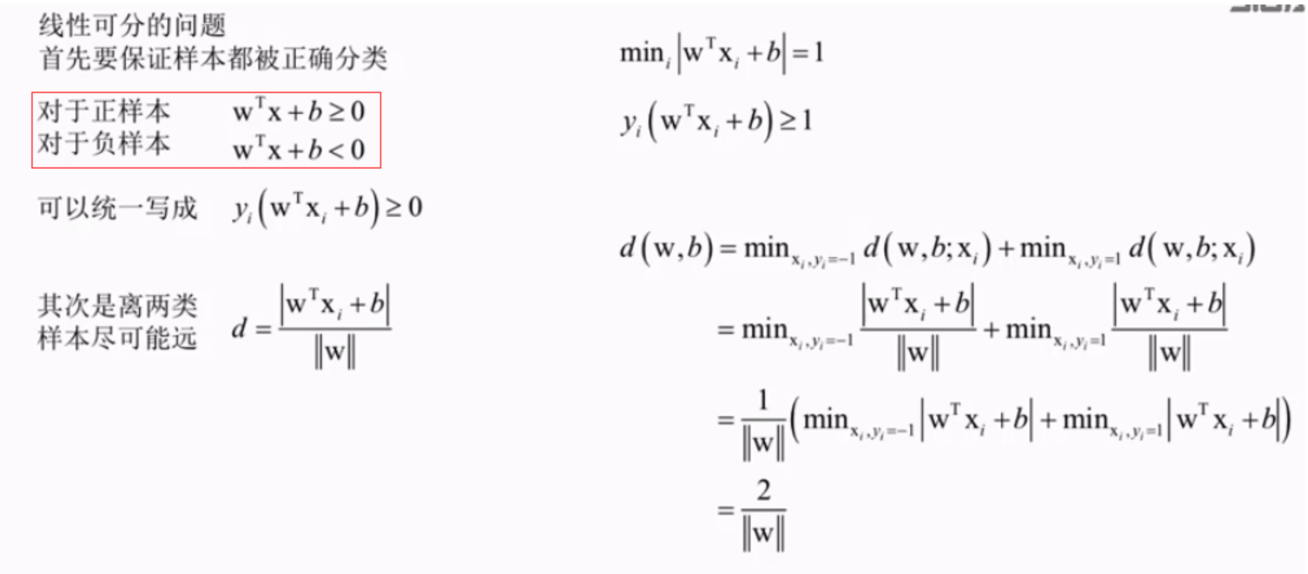
14.4 分类间隔

前面都是对二分类问题，如果对于多分类问题。



尽可能使得分类间隔最大化的那个最好。

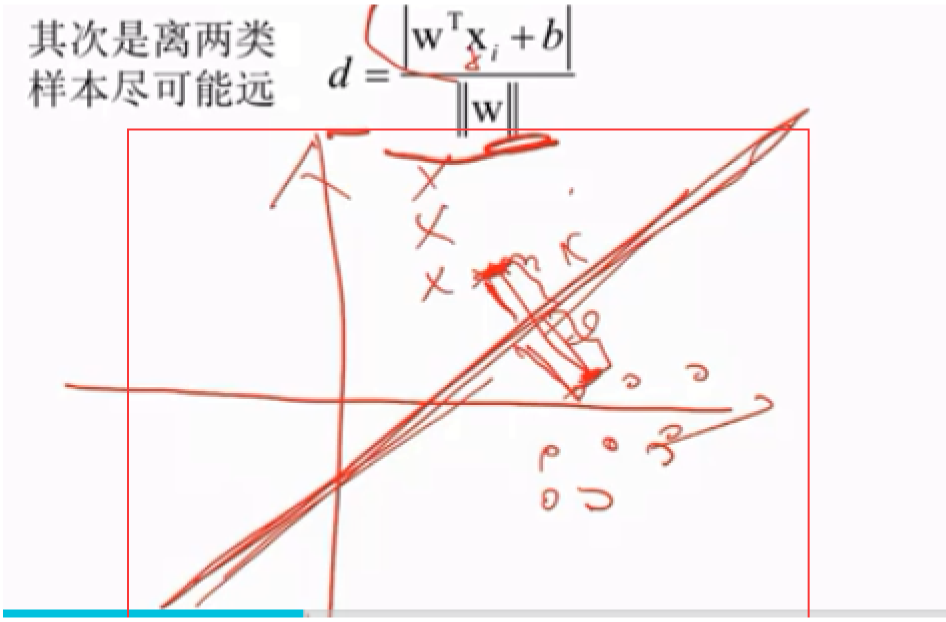
14.5 线性可分的原问题



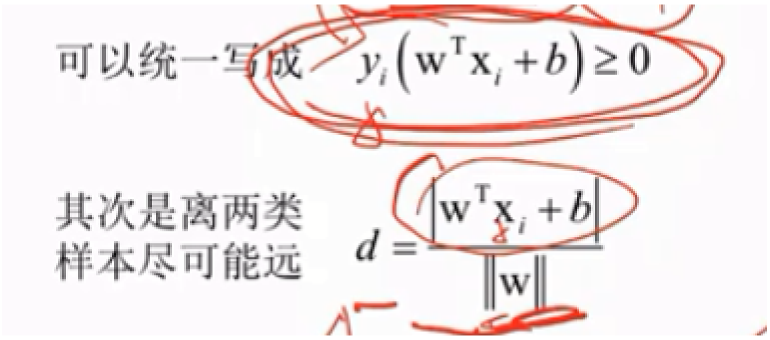
离两类样本都远直线如下

两类样本距离尽可能的最大化

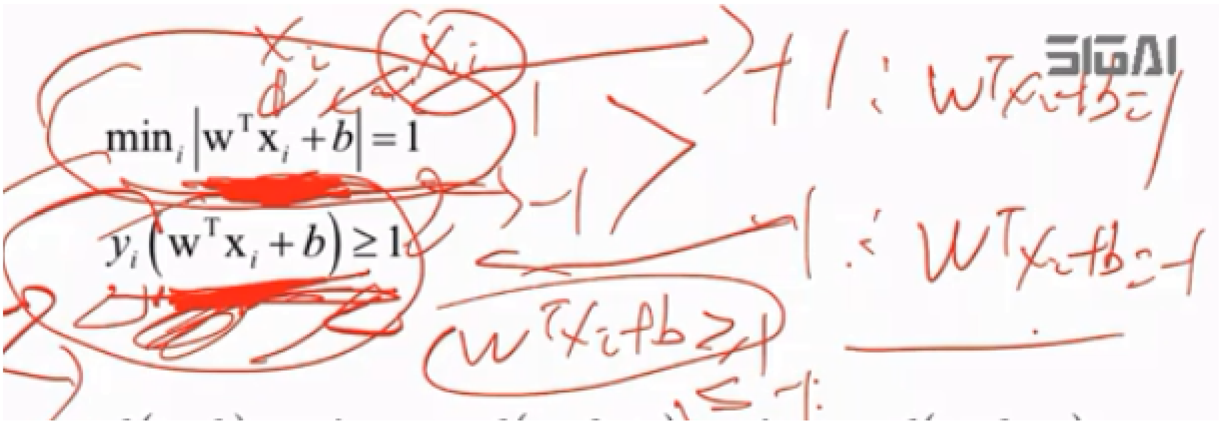
计算距离



回忆一下解析几何里面，点到超平面的距离。



张样本的距离和负样本



计算距离

$$
\begin{aligned}
d(\mathrm{w},b) &= \boxed{\min_{\mathrm{x}_i,y_i=-1} d(\mathrm{w},b;\mathrm{x}_i)} + \boxed{\min_{\mathrm{x}_i,y_i=1} d(\mathrm{w},b;\mathrm{x}_i)} \\
&= \min_{\mathrm{x}_i,y_i=-1} \frac{\left|\mathrm{w}^{\mathrm{T}}\mathrm{x}_i + b\right|}{\|\mathrm{w}\|} + \min_{\mathrm{x}_i,y_i=1} \frac{\left|\mathrm{w}^{\mathrm{T}}\mathrm{x}_i + b\right|}{\|\mathrm{w}\|} \\
&= \frac{1}{\|\mathrm{w}\|}\left(\boxed{\min_{\mathrm{x}_i,y_i=-1}\left|\mathrm{w}^{\mathrm{T}}\mathrm{x}_i} + b\right| + \boxed{\min_{\mathrm{x}_i,y_i=1}\left|\mathrm{w}^{\mathrm{T}}\mathrm{x}_i + b\right|}\right) \\
&= \frac{2}{\|\mathrm{w}\|}
\end{aligned}
$$

所以有，

通过范数优化

$$\max \frac{2}{\|w\|}$$

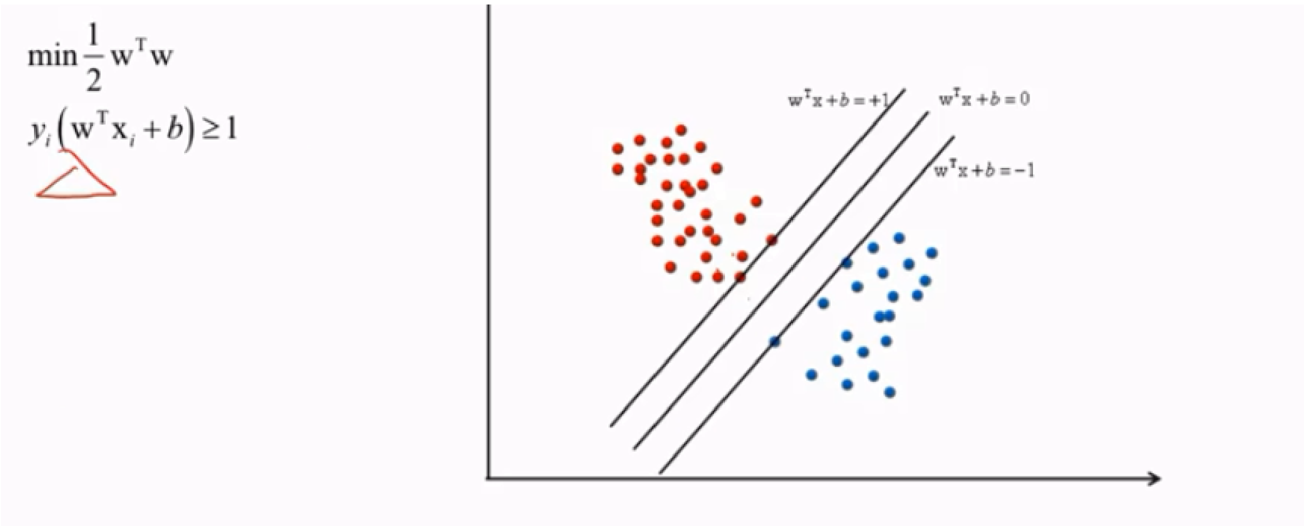
$$\min \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w}$$

$$y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i + b\right) \geq 1$$

$$\geq 0$$

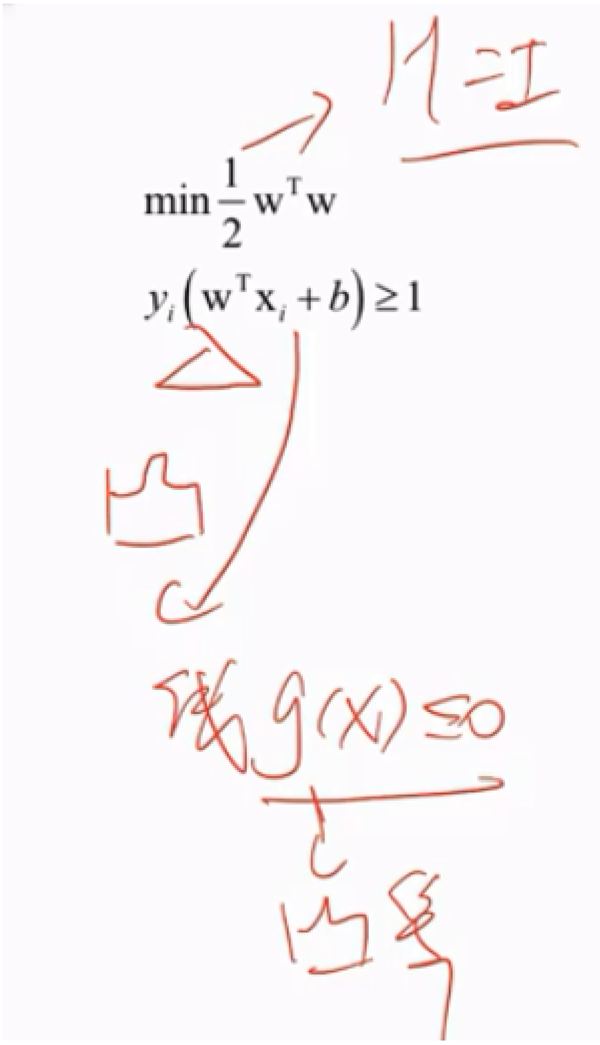
$$\sqrt{w^T w}$$

看看示意图，红色为正样本



证明是凸优化

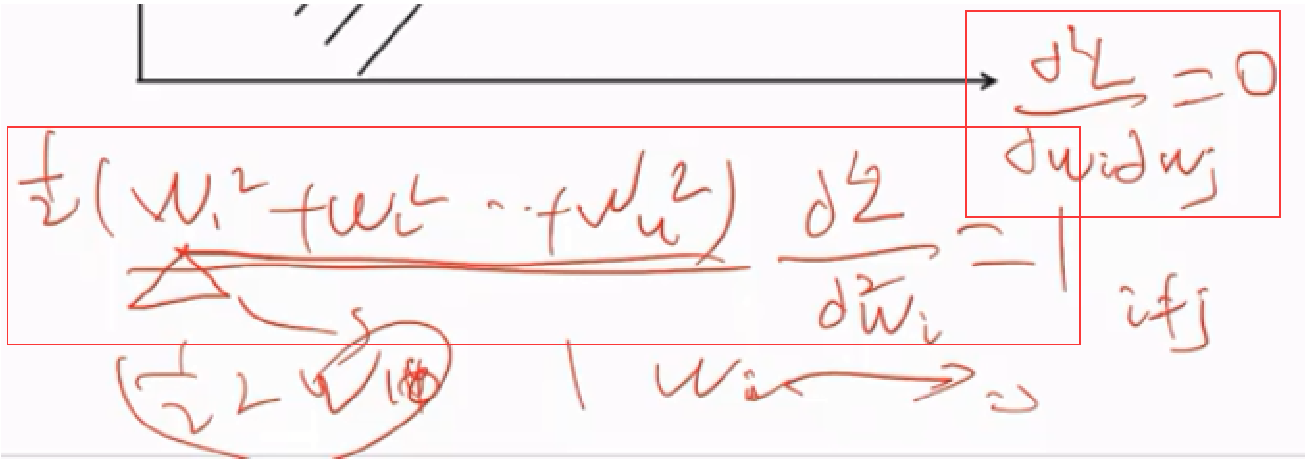
假设有一个优化目标



线性可分围成的都是凸集

我们计算一下优化目标函数

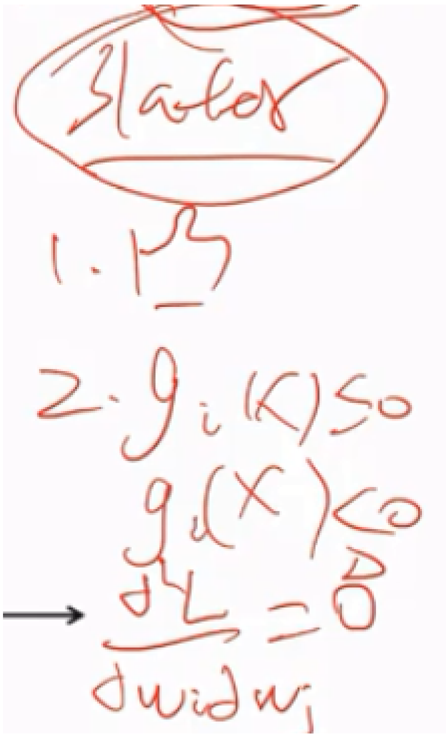
如果原问题难求解，满足凸优化的条件，可以转化为对偶问题来求解。



矩阵的形状如如，可以知道，这是一个严格凸集

来看看slator条件

1. 是凸优化
2. 至少存在图中公式2



如果原问题难求解，满足凸优化的条件，可以转化为对偶问题来求解。

14.6 线性可分的对偶问题

前面的最优化问题带有太多的不等式约束，不易求解
这个优化问题是凸优化，而且满足Slater条件，因此可以强对偶成立，可以用拉格朗日对偶转化成对偶问题

$$L(\mathrm{w},b,\alpha)=\frac{1}{2}\|\mathrm{w}\|^2-\sum_{i=1}^{l}\alpha_i\left(y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}+b\right)-1\right)$$

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

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

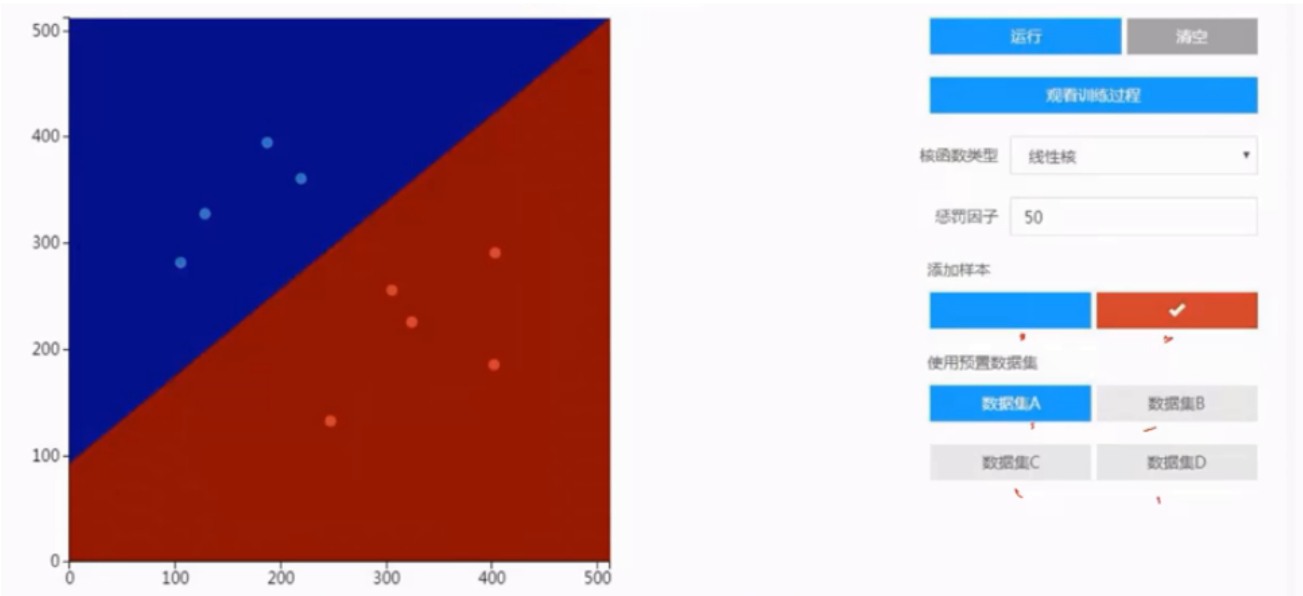
$$\nabla_{\mathrm{w}}L=0 \qquad \mathrm{w}=\sum_{i=1}^{l}\alpha_i y_i \mathrm{x}_i$$

$$\max_{\alpha}-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \mathrm{x}_i^{\mathrm{T}}\mathrm{x}_j+\sum_{k=1}^{l}\alpha_k$$

回忆拉格朗日对偶，要到前面的章节去看。

14.7 线性可分的实验

直观看出把样本分开的是一条直线



14.9 线性不可分的对偶问题

转换成对偶问题

$$y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i+b\right)\geq 1-\xi_i \Rightarrow -\left(y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i+b\right)-1+\xi_i\right)\leq 0$$

$$\xi_i \geq 0 \Rightarrow -\xi_i \leq 0$$

$$L\left(\mathrm{w},b,\alpha,\xi,\beta\right)=\frac{1}{2}\|\mathrm{w}\|^2+C\sum_{i=1}^{l}\xi_i-\sum_{i=1}^{l}\alpha_i\left(y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i+b\right)-1+\xi_i\right)-\sum_{j=1}^{l}\beta_i\xi_i$$

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

$$\nabla_{\xi}L=0 \qquad \alpha_i+\beta_i=C$$

$$\nabla_{\mathrm{w}}L=0 \qquad \mathrm{w}=\sum_{i=1}^{l}\alpha_i y_i \mathrm{x}_i$$

$$\max_{\alpha}-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \mathrm{x}_i^{\mathrm{T}}\mathrm{x}_j+\sum_{k=1}^{l}\alpha_k$$

$$\alpha_i+\beta_i=C \qquad \beta_i\geq 0 \qquad \alpha_i\leq C$$

$$Q_{ij} = y_i y_j \mathrm{x}_i^{\mathrm{T}} \mathrm{x}_j$$

$$\min_{\alpha} \frac{1}{2} \alpha^{\mathrm{T}} \mathrm{Q} \alpha - \mathrm{e}^{\mathrm{T}} \alpha$$

$$0 \le \alpha_i \le C$$

$$\mathrm{y}^{\mathrm{T}} \alpha = 0$$

这是一个凸优化问题

$$\mathrm{Q} = \mathrm{X}^{\mathrm{T}} \mathrm{X}$$

$$\mathrm{X} = \left[y_1 \mathrm{x}_1, \ldots, y_l \mathrm{x}_l \right]$$

$$\mathrm{x}^{\mathrm{T}} \mathrm{Q} \mathrm{x} = \mathrm{x}^{\mathrm{T}} \left(\mathrm{X}^{\mathrm{T}} \mathrm{X} \right) \mathrm{x} = \left(\mathrm{X} \mathrm{x} \right)^{\mathrm{T}} \left(\mathrm{X} \mathrm{x} \right) \ge 0$$

$$\min_{\alpha} \frac{1}{2} \alpha^{\mathrm{T}} \mathrm{Q} \alpha - \mathrm{e}^{\mathrm{T}} \alpha$$

$$0 \le \alpha_i \le C$$

$$\mathrm{y}^{\mathrm{T}} \alpha = 0$$

这是一个凸函数，

$$\begin{bmatrix} \cancel{y_1 x_1^T} \\ y_l x_l^T \\ x_l x_l^T \end{bmatrix} [y_1 x_1 \cdots y_l x_l]$$

$$y_i y_j x_i^T x$$

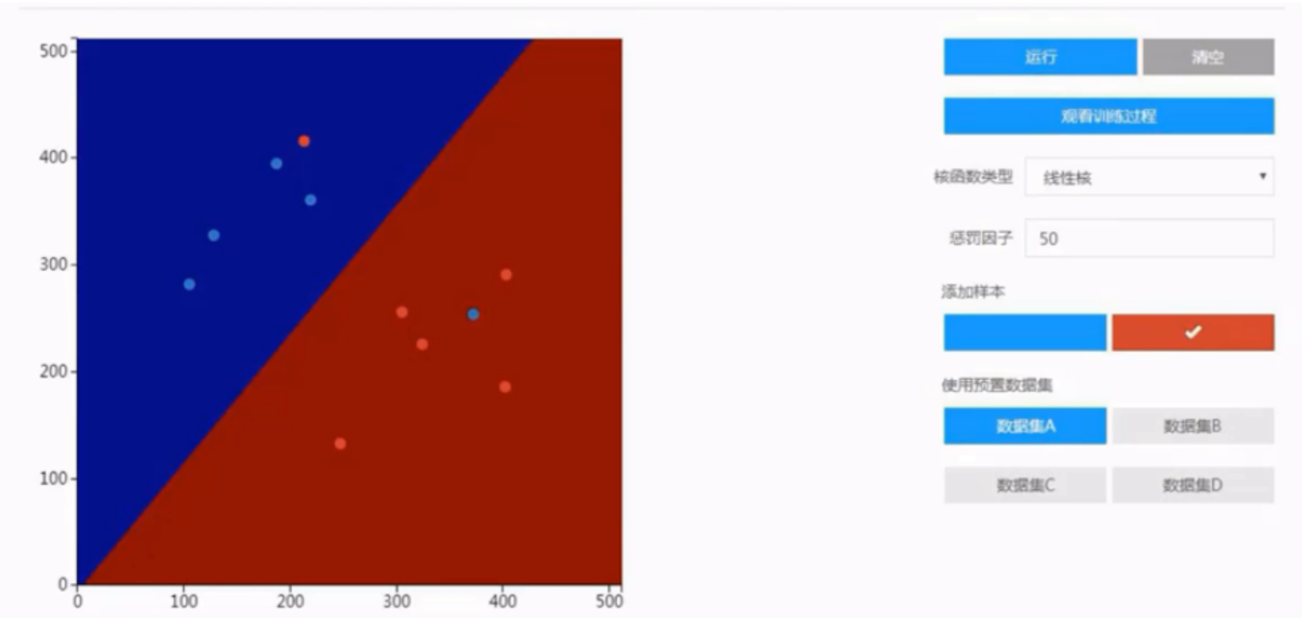
14.10 KKT条件下的使用

什么上界

自由向量

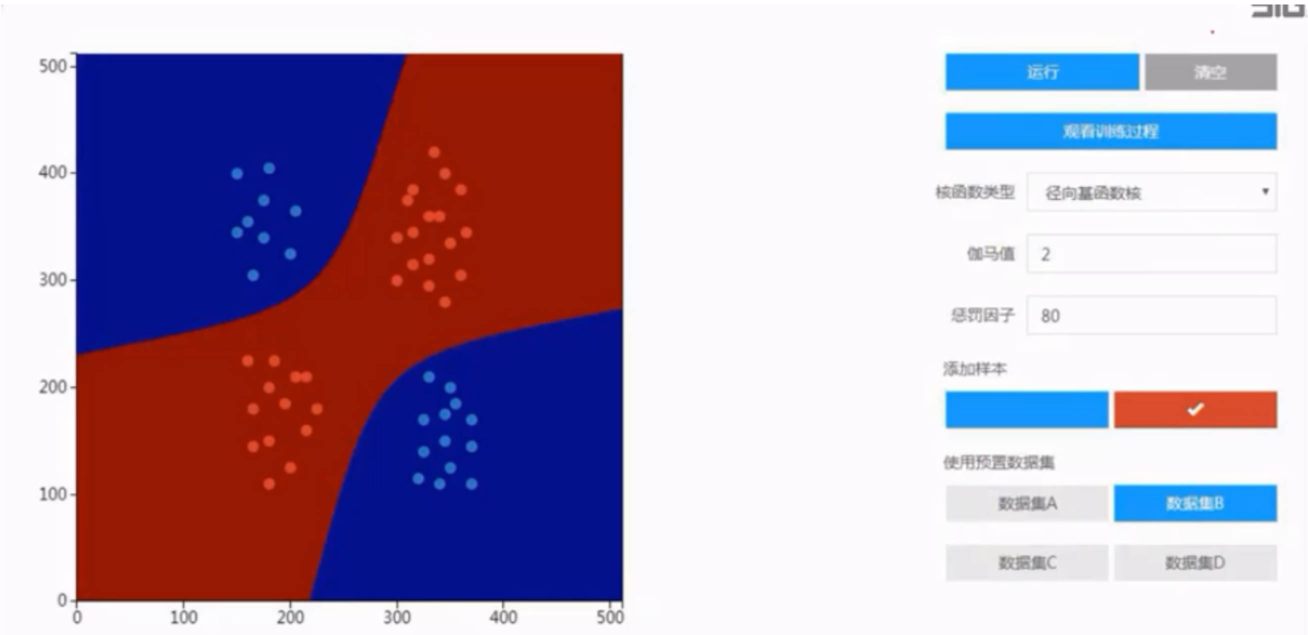
不是说的很抽象，是听得很不懂。不懂每个符号是怎么推导的。

14.11 线性不可分的实验



14.12 核函数实验

高斯核（RBF)



14.13 核映射与核函数

前面通过松弛变量和惩罚因子，将线性

核映射与核函数

虽然引入了松弛变量和惩罚因子，可以处理线性不可分的问题，但SVM还是一个线性模型，只是允许错分样本的存在

核映射 $z=\varphi(x)$

$$z_i^T z_j=\varphi(x_i)^T\varphi(x_j)$$

核函数 $K(x_i,x_j)=K(x_i^T x_j)=\phi(x_i)^T\phi(x_j)$

$$\operatorname{sgn}\left(\sum_{i=1}^{l}\alpha_i y_i K(x_i^T x)+b\right)$$

$$\min_\alpha \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \phi(x_i)^T\phi(x_j)-\sum_{k=1}^{l}\alpha_k$$

$$0\le\alpha_i\le C$$

$$\sum_{j=1}^{l}\alpha_j y_j=0$$

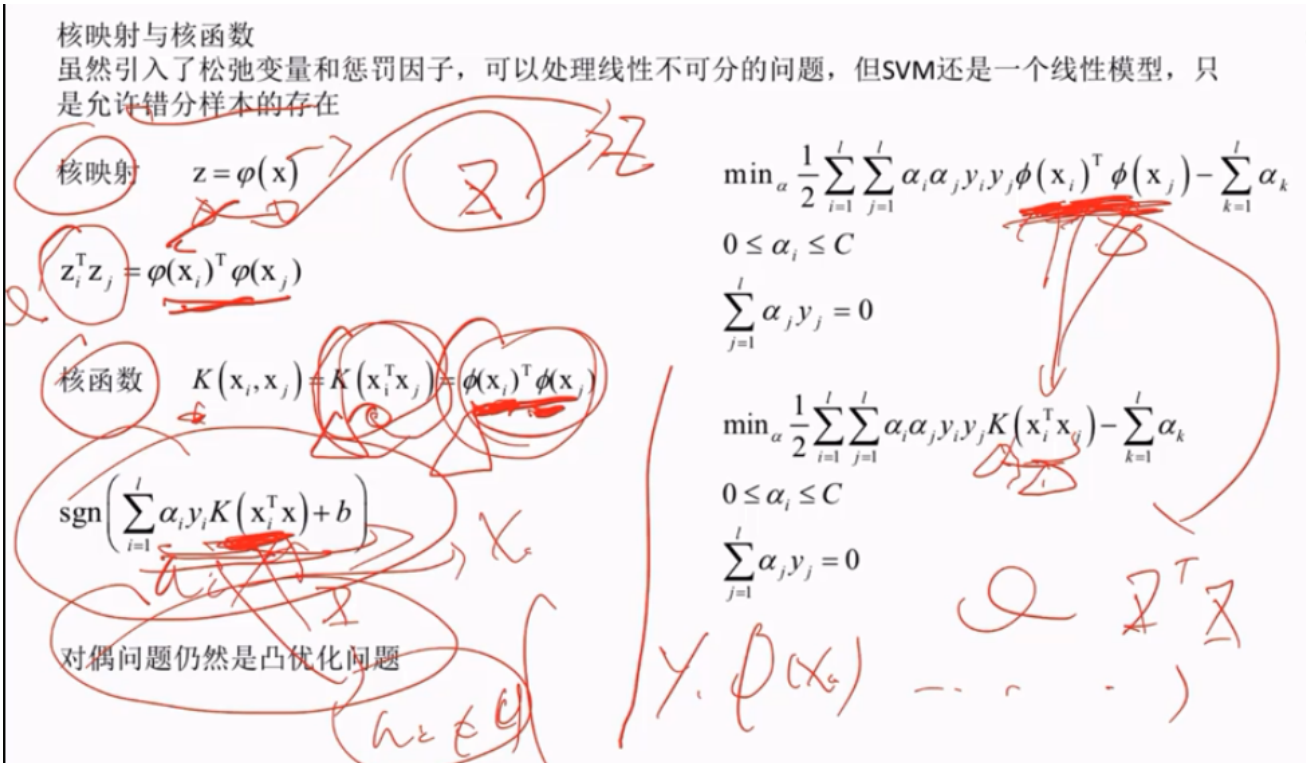
$$\min_\alpha \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i^T x_j)-\sum_{k=1}^{l}\alpha_k$$

$$0\le\alpha_i\le C$$

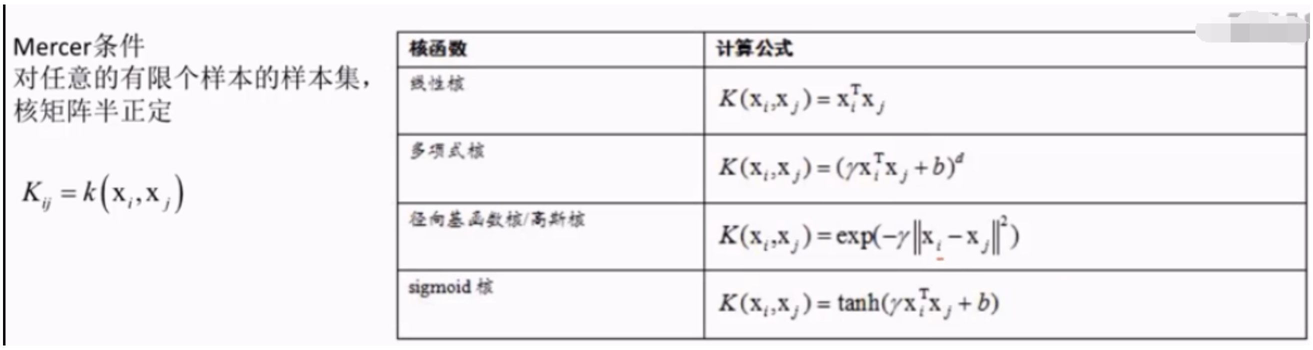
$$\sum_{j=1}^{l}\alpha_j y_j=0$$

对偶问题仍然是凸优化问题

- 核映射
  - 先映射再做内积
- 核函数
  - 先做内积再做映射



Mercer条件



核函数	计算公式
线性核	$K(x_i,x_j)=x_i^T x_j$
多项式核	$K(x_i,x_j)=(\gamma x_i^T x_j+b)^d$
径向基函数核/高斯核	$K(x_i,x_j)=\exp(-\gamma\lVert x_i-x_j\rVert^2)$
sigmoid 核	$K(x_i,x_j)=\tanh(\gamma x_i^T x_j+b)$

支持向量机能够转化为非线性问题，归功于核函数

核函数

14.14 本集总结

1. SVM基本情况
2. 线性分类
3. 分类间隔
4. SVM，引入线性可分的问题，凸优化问题，见了松弛变量，惩罚因子，slater条件。
5. 用和映射和核函数转化成非线性模型。
6. 在原问题上如何用KTT条件来解决。

在openCV上的应用

https://docs.opencv.org/master/d1/d69/tutorial_table_of_content_ml.html