

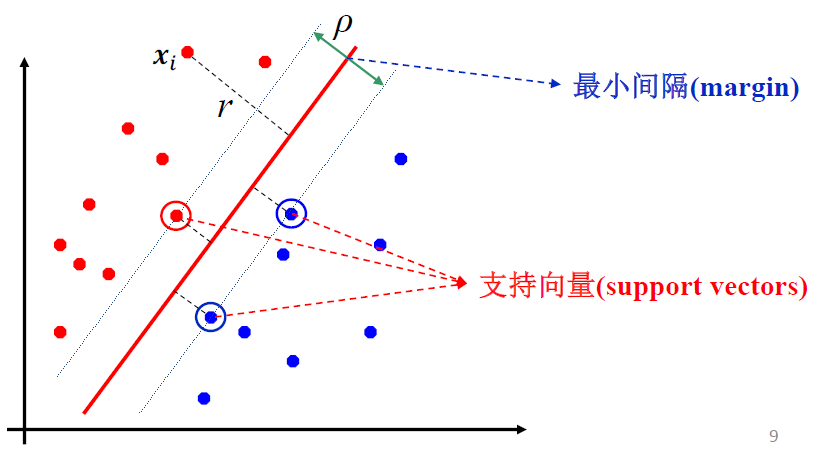
后验概率是机器学习的目标

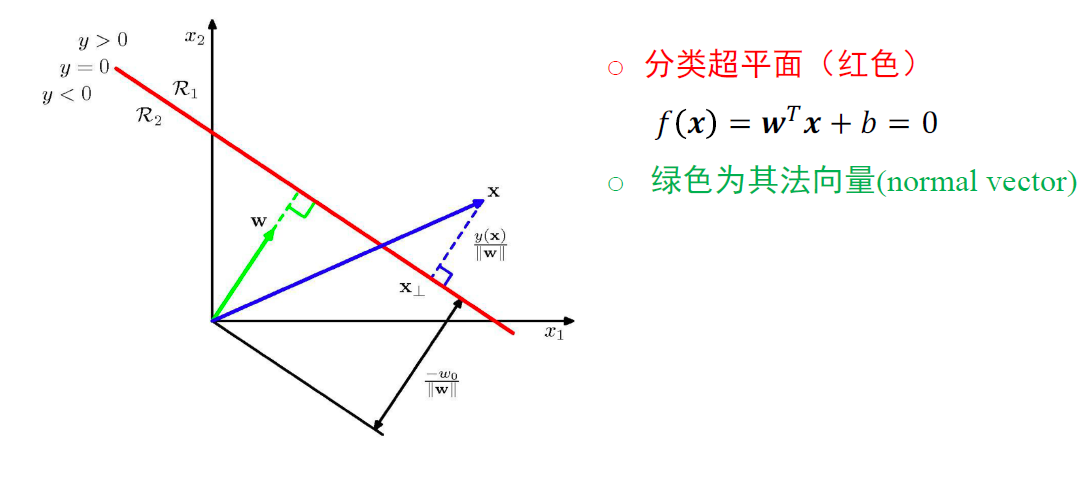


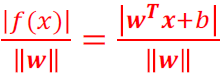
一个点的**边际margin**是**其到分界超平面的垂直距离**

SVM**最大化（所有训练样本的）最小边际margin**

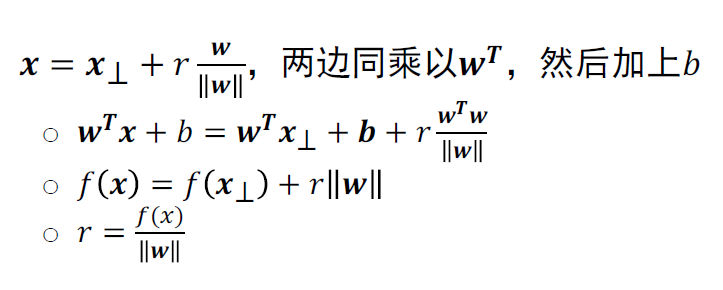
有最小边际的点称为**支持向量**

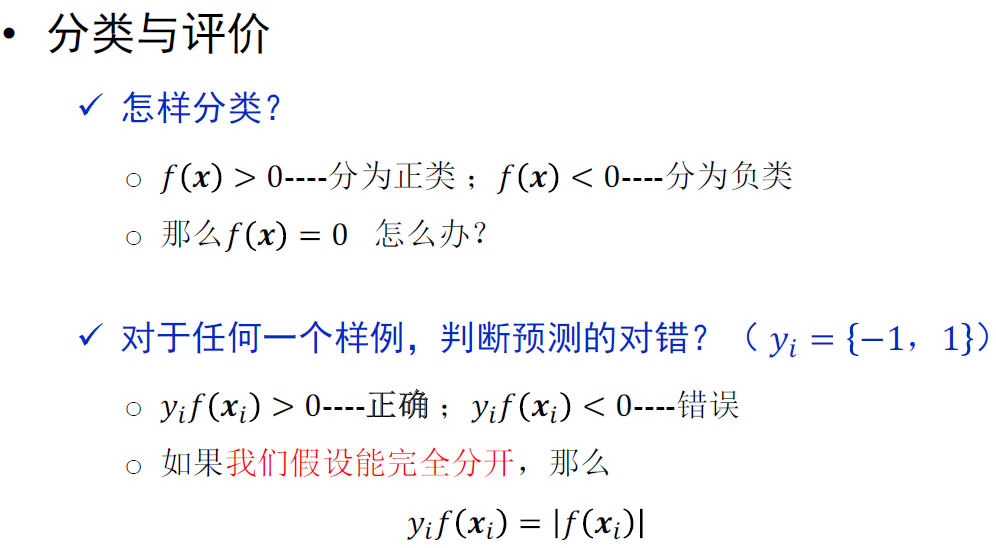


（w为列向量）

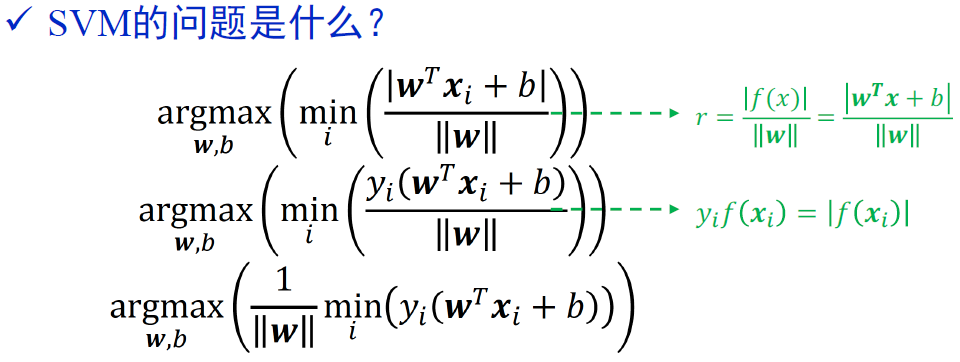
**任一点 x到超平面的距离𝑟/margin为：**

推导：

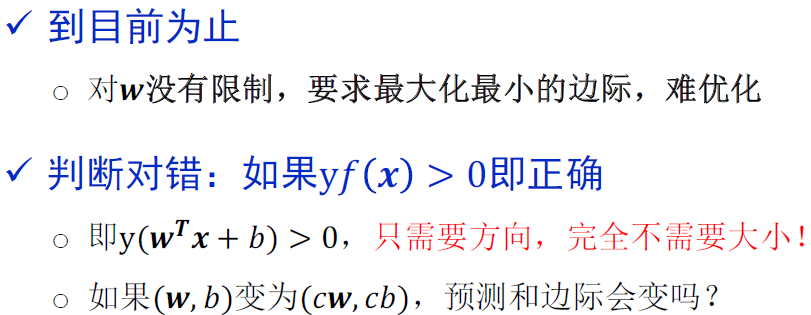




（如果不能完全分开，等号前的才是对的）

（i表示第i个训练样本）

最大化最小margin

**不变**。这里c>0。

理解：

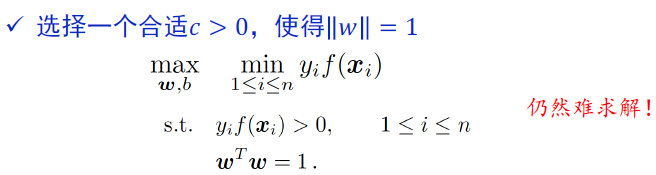
如果c>0，则

所以**(w\*,b\*)下和(cw\*,cb\*)下的最小margin一样（其实不加“最小”二字也成立）**

所以如果(w\*,b\*)可以使得最小margin最大，(cw\*,cb\*)也能使得最小margin最大

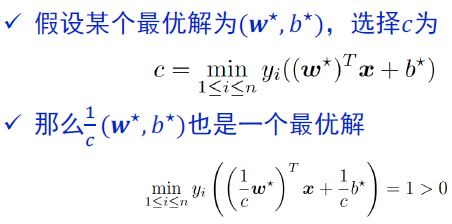
**结论：如果(w\*,b\*)是最优解，那么(cw\*,cb\*)也是最优解**

**继续优化一：限定w为单位向量**



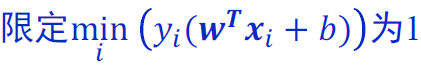
选择：

**继续优化二：限定最小margin为1（不限定w为单位向量）**

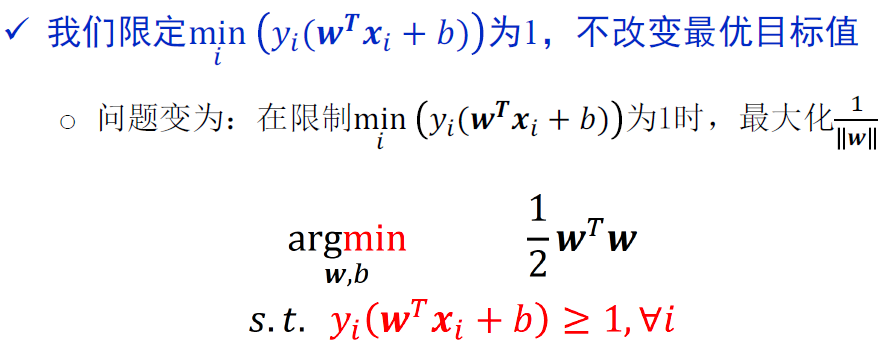


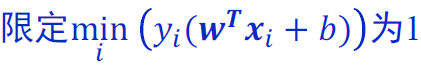
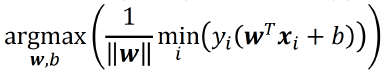
理解：这里貌似错了，因为不限定w为单位向量，所以最小margin的计算（上下两个式子）还要除w的模长。

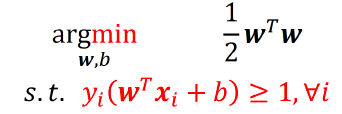
下式把1/c提出来，等于1/c × c = 1

结论：抛弃限定w为单位向量，而是

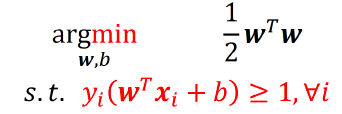
**继续优化二：**

（s.t.指“约束条件”）

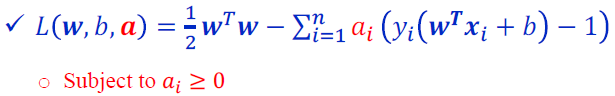
理解：当时，为了最大化，所以要，所以要最小化。

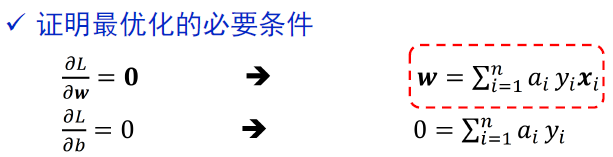
结论：问题转化为（约束的意思是：除了支持向量margin等于1，其它点的margin都大于1）

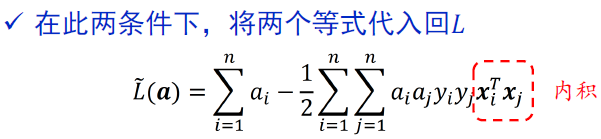
**求解：**

采用**拉格朗日乘子法**，把约束放到目标函数里面。然后**求导**。拉格朗日乘子法的约束不一定是等式，也可以是不等式：。

引入KKT乘子ai，得到拉格朗日函数：





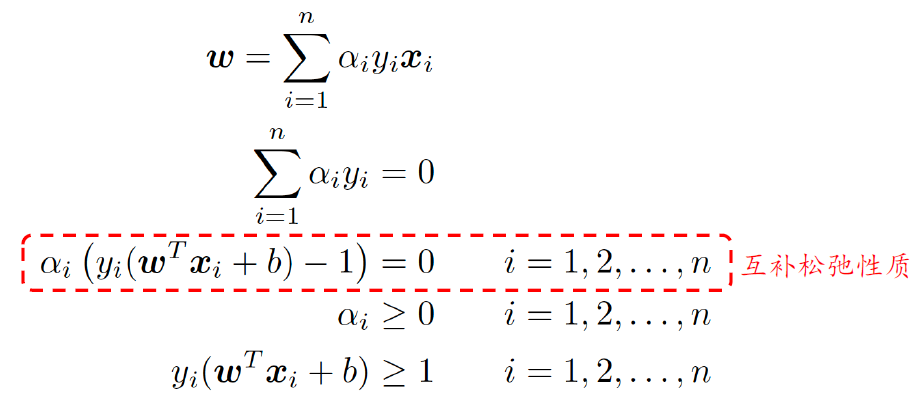


需要满足KKT条件：

KKT条件是**非线性规划最优解的必要条件**。

但在原始**SVM**问题里，KKT条件对于确定最优解是**充分必要**的。

KKT条件：

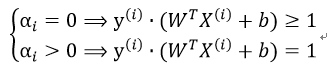


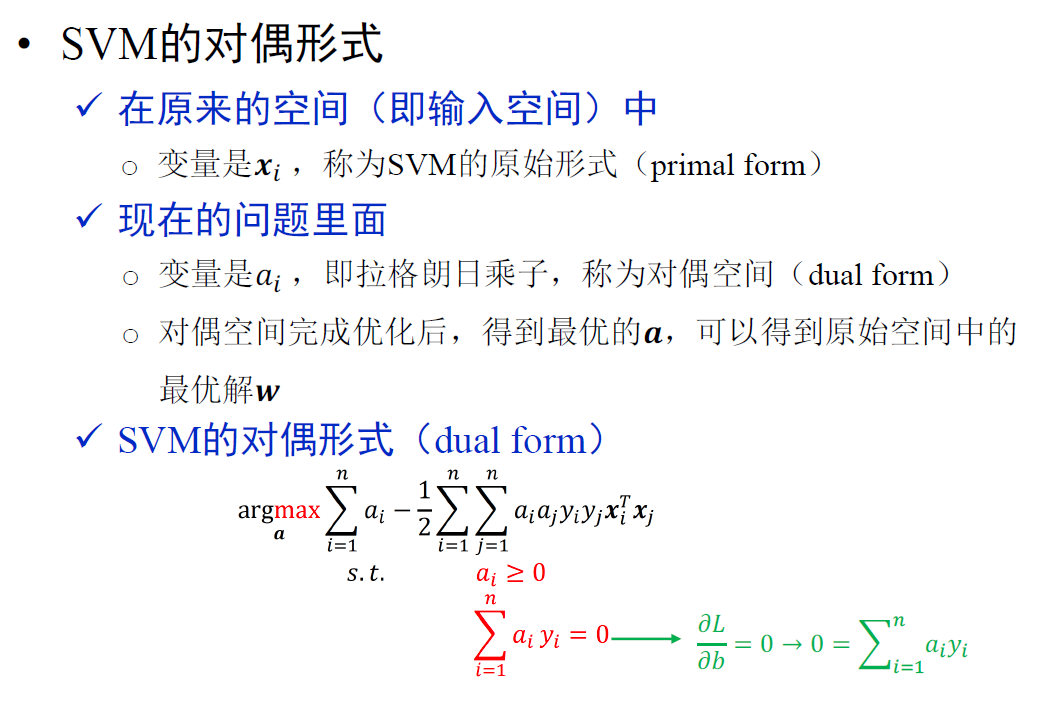
前两个条件是拉格朗日乘子法求导后得到的。

**互补松弛性质**非常重要。

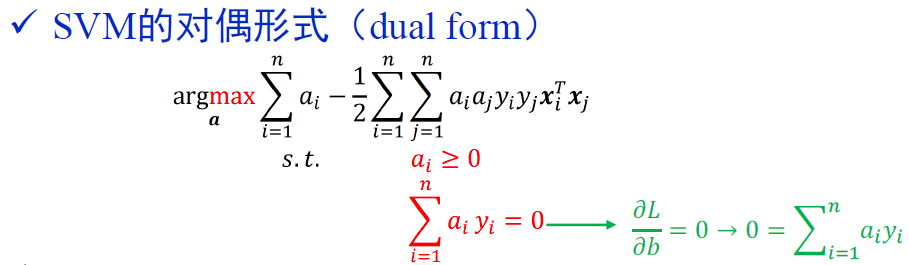
第四个条件是KKT乘子ai需要满足的条件。

由上述条件可得：

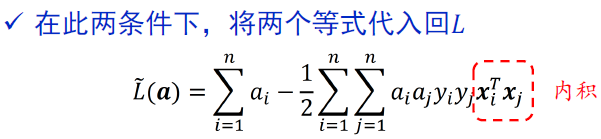
，即**支持向量对应的ai>0；非支持向量对应的的ai=0，对模型的学习没有起到作用**。

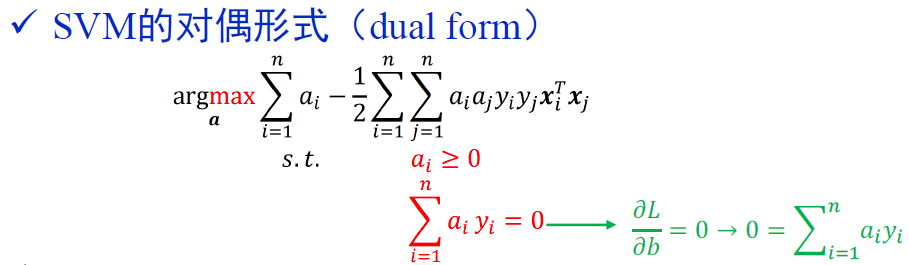
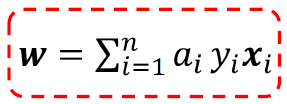


理解：在SVM的原始形式里，变量是xi；通过拉格朗日乘子法求导后，变量只剩下拉格朗日乘子ai，已经没有原来的变量了。

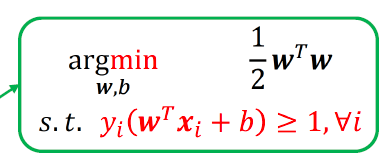


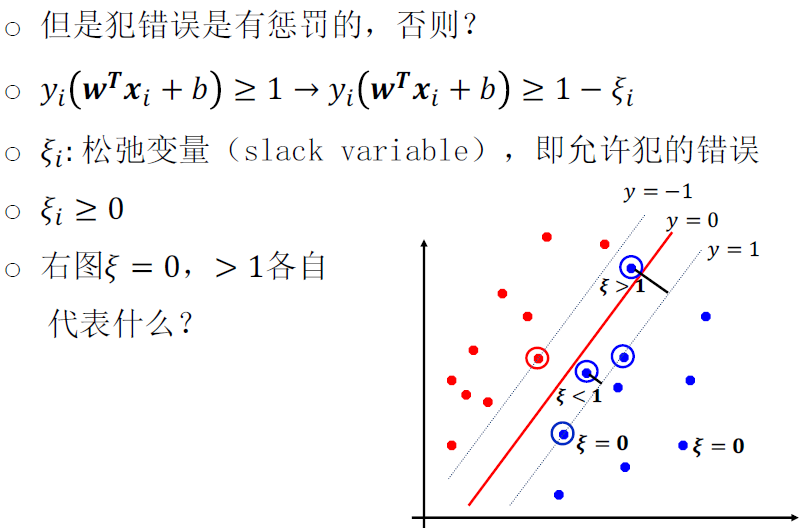
理解：

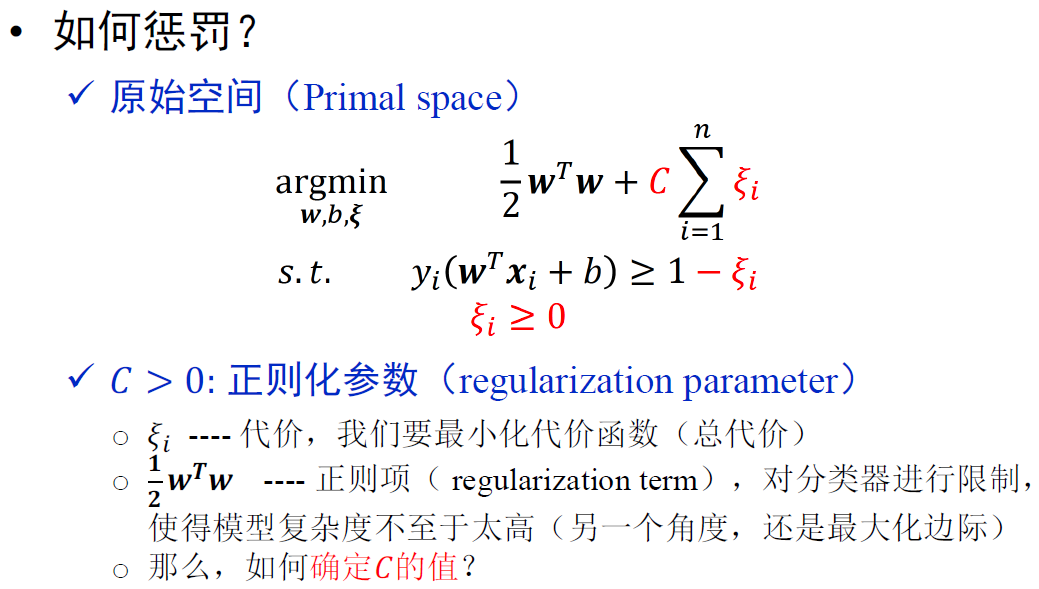
第一行是指“”。

是**凹函数**，可以求**最大值**。求出ai之后带回，得到w。

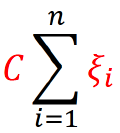
**Soft margin**

限制条件太严格了，可以，即允许少数点margin比1小。

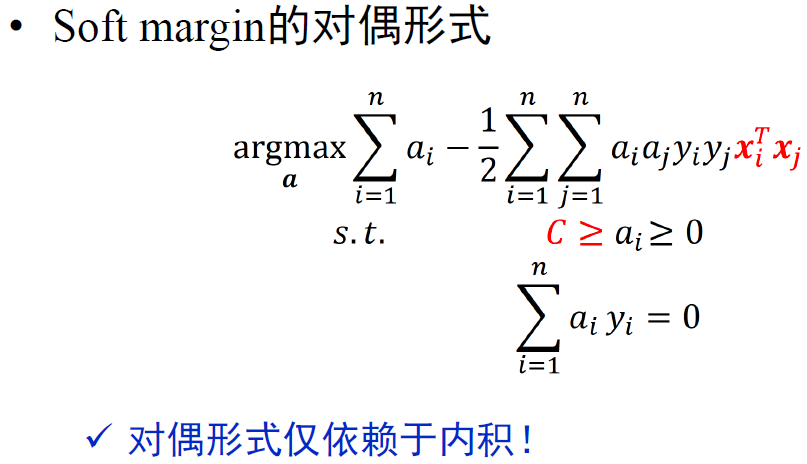


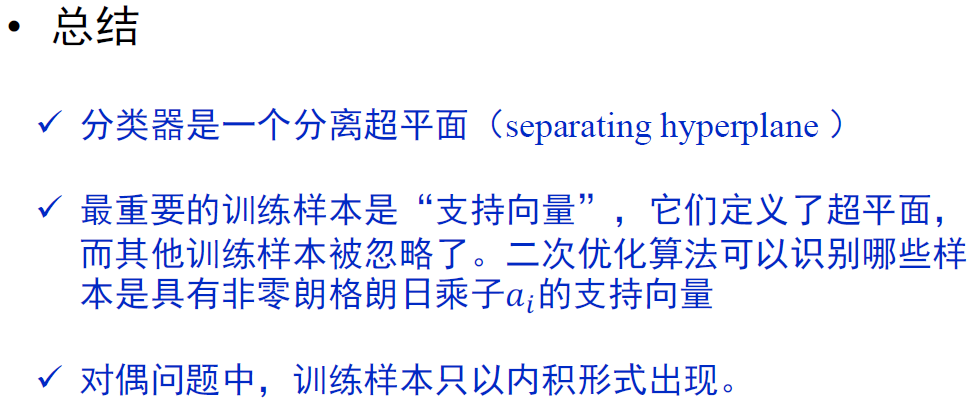


理解：

允许犯错误，但是不能犯错太多，所以要**把犯错****加到目标函数中**

原始空间通过拉格朗日乘子法转化为对偶形式（老师把推导过程忽略了）：





非线性SVM

