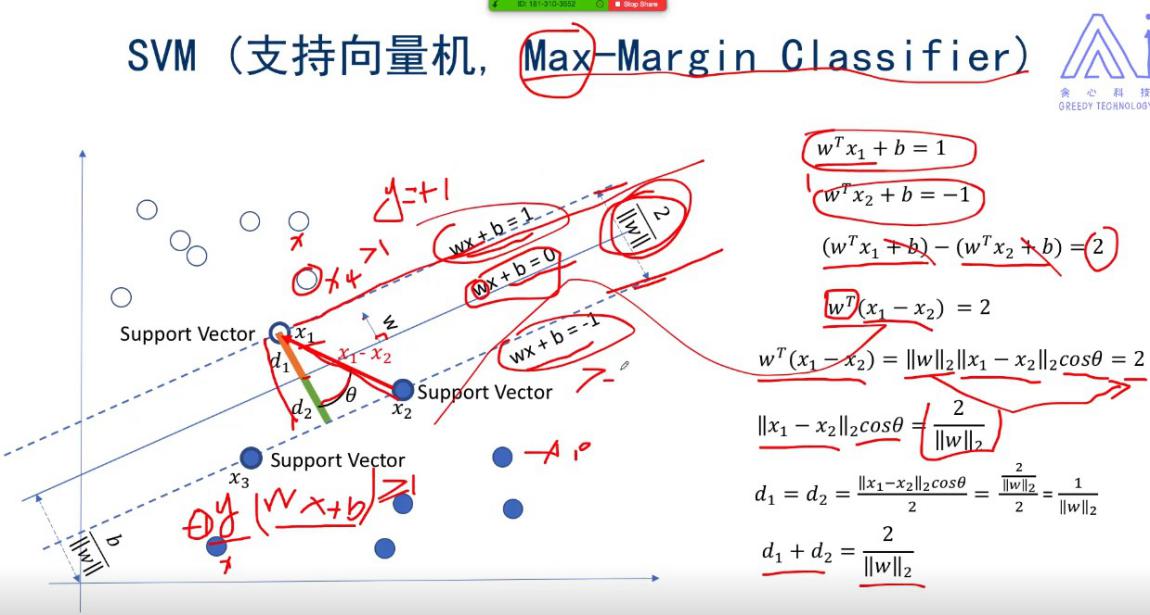
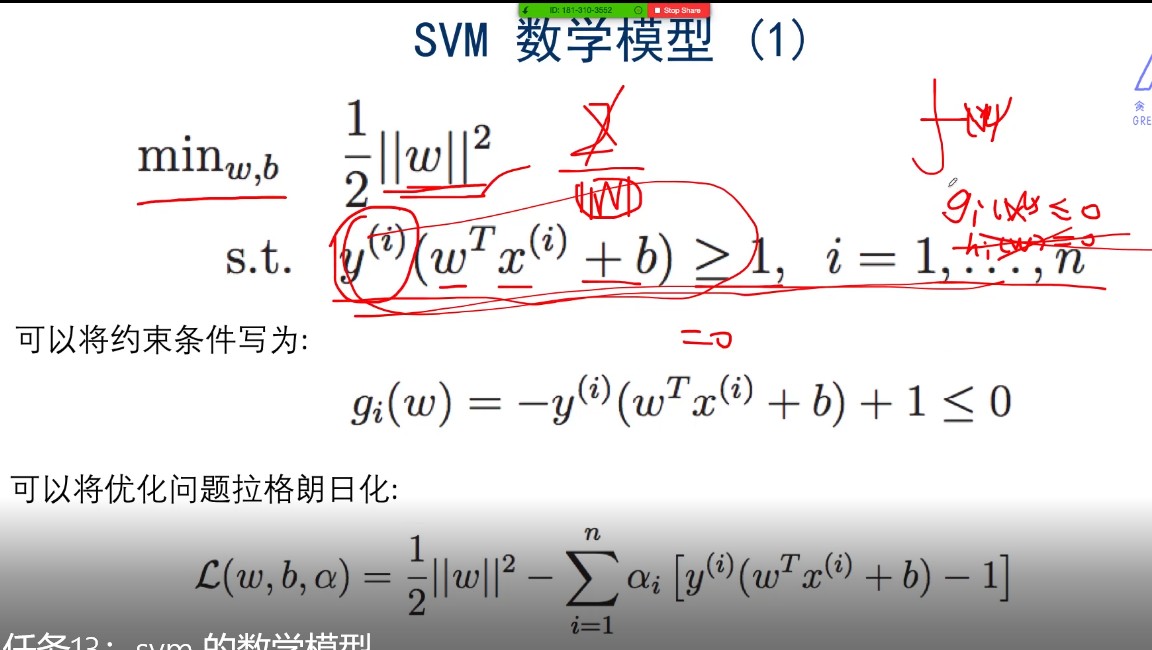
# SVM

## 一、SVM：找最优分界线

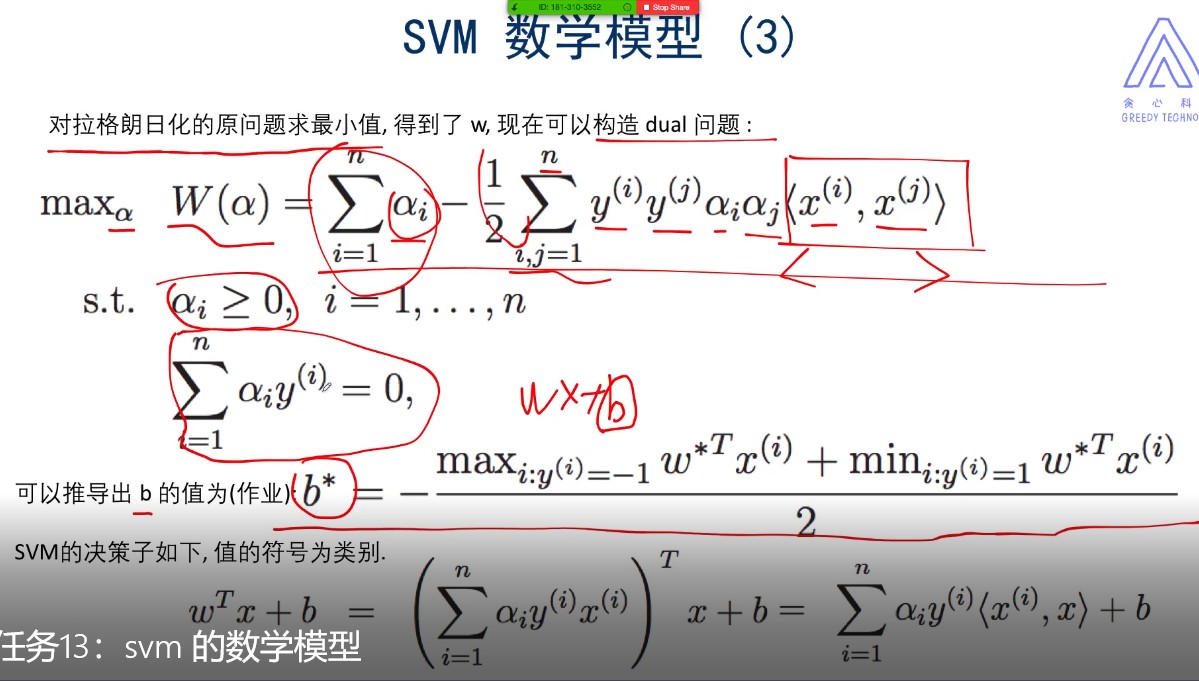
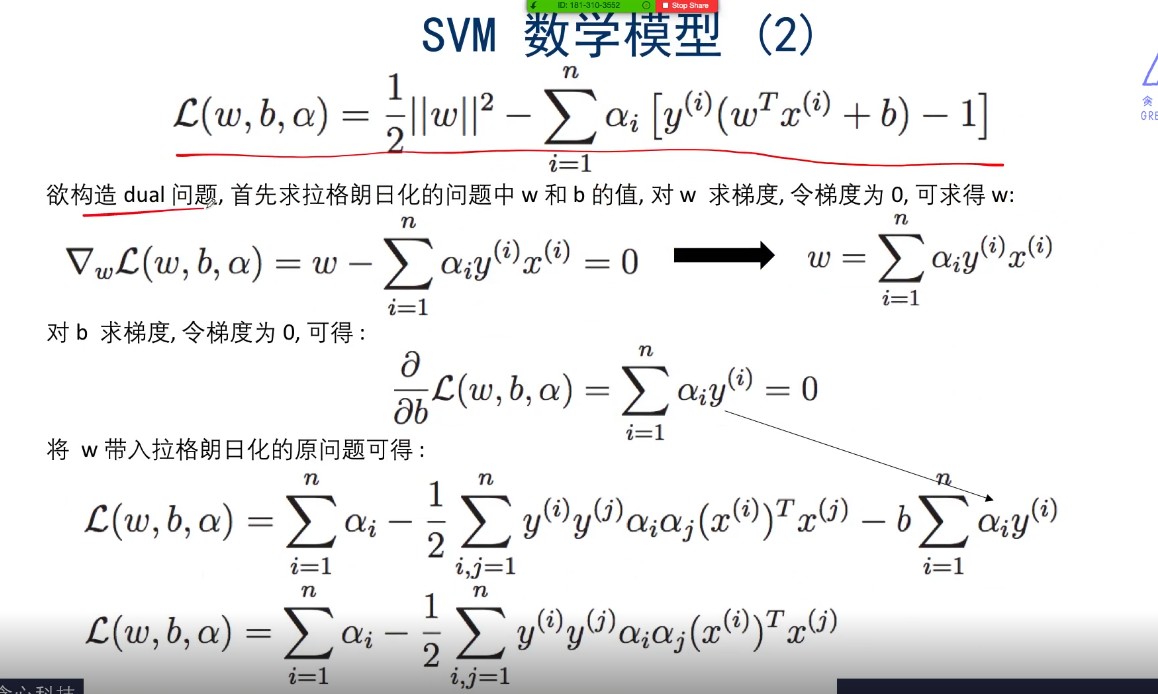
### 1.1 线性理想SVM == max margin classiler（margin：d1+d2 = 2/w越大越好）



### 1.2 数学模型

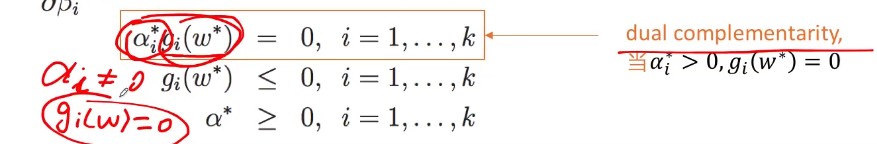


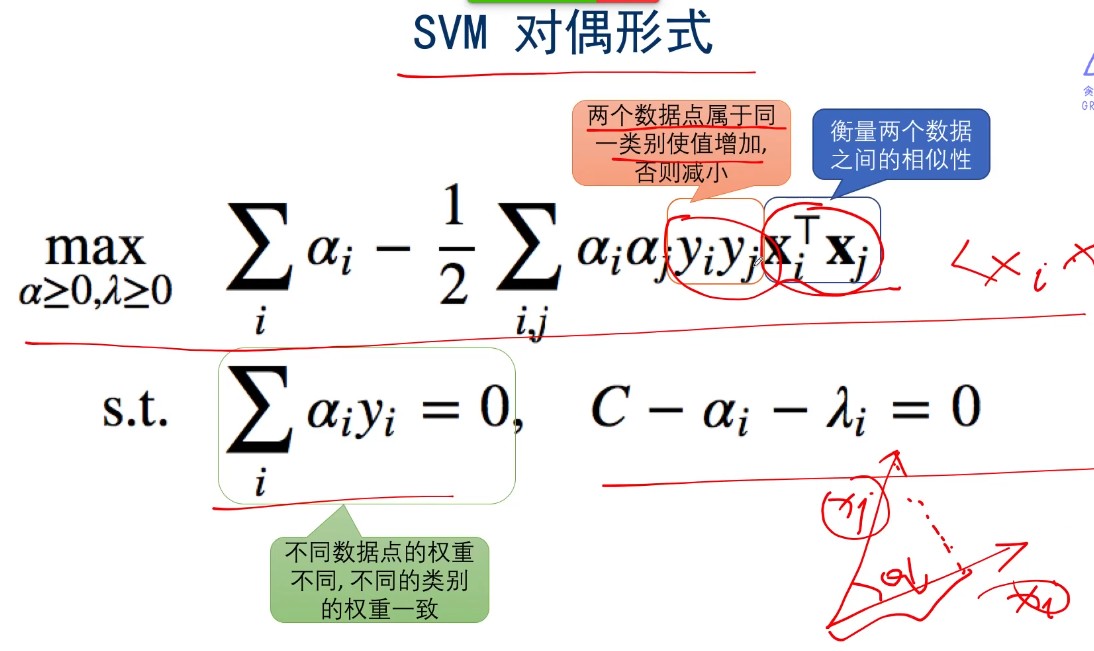
||w||是w的转置和w相乘，他们的求和实质是不一样的，求和的下标i，j不同



### 1.3 对偶及其KKT条件

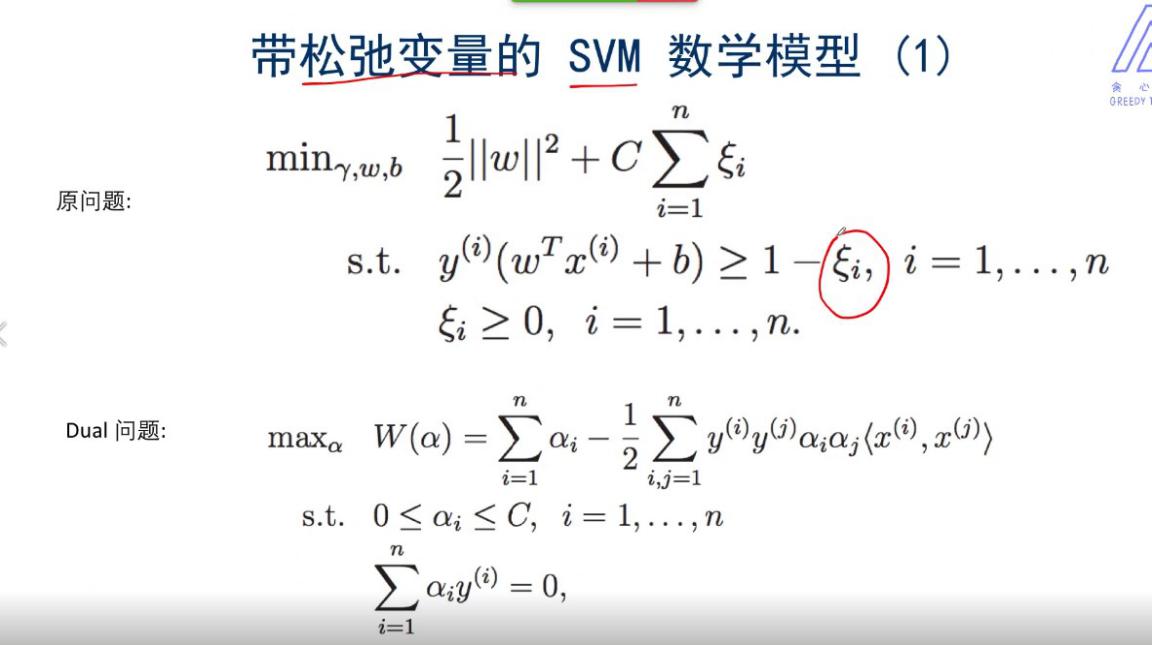
<x^(i),x>点击是全数据集，那计算量是不是很大呢？但是KKT条件，也就是在支持向量上的a才不等于，其余a等于0



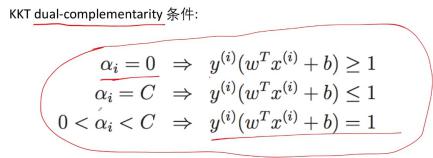


## 二、SVM软间隔异常值处理

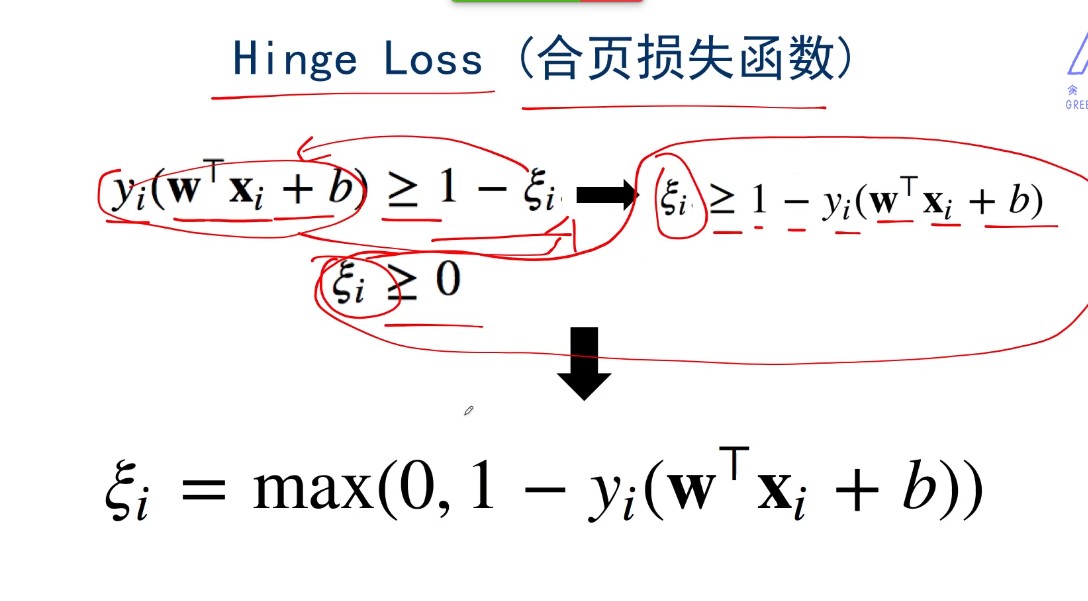
### 2.1 处理1：放松限制（处理线性不可分）

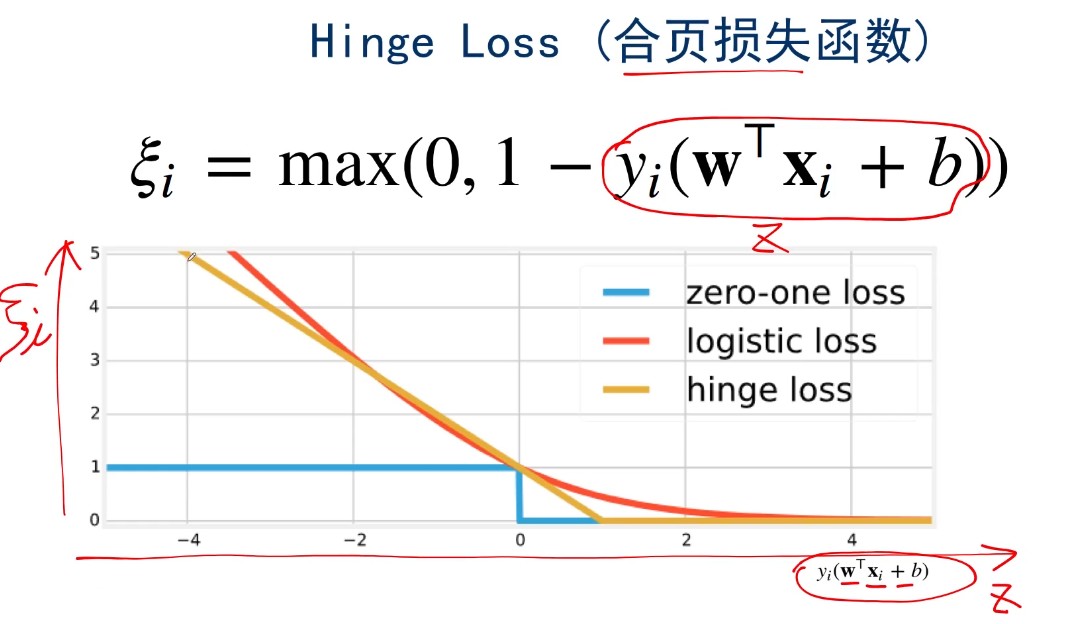


KKT条件：



2.1.1 另一种理解方式：Hinge Loss（合页损失）





### 2.2 处理2：不放松限制（bad）

## 三、带kernel trick的SVM（线性不可分）

### 3.1 为什么使用核函数：便于处理线性不可分

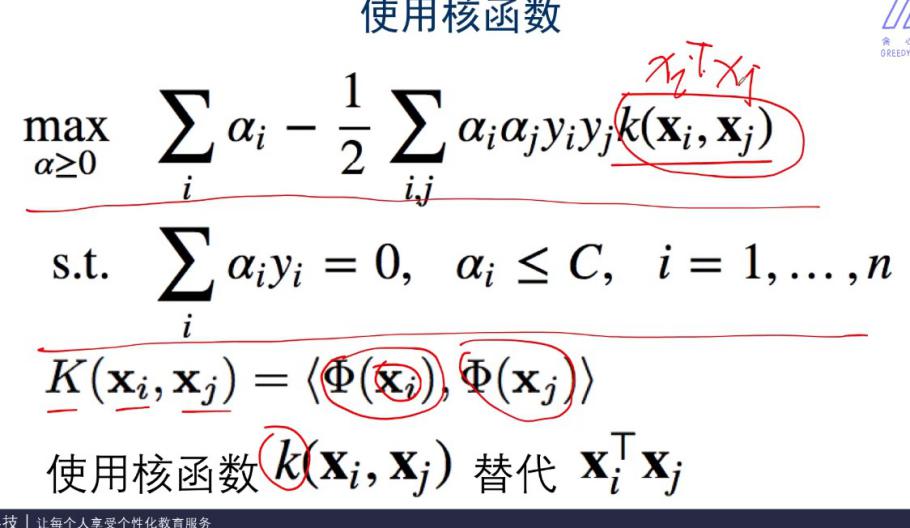
### 3.2 使用核函数代替点积，将特征映射到更高的维度（常用）

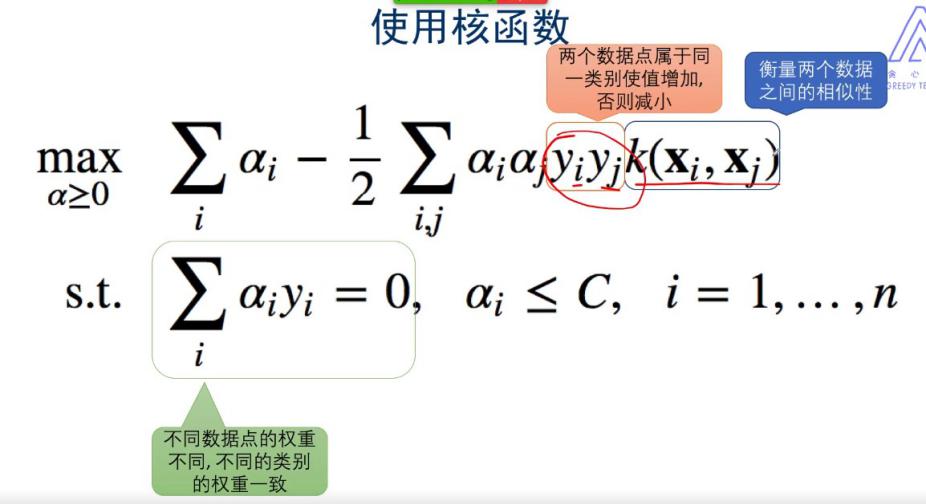
满足kernel的条件：对称矩阵、半正定矩阵

常见的kernel：<https://blog.csdn.net/qq_27231343/article/details/51817866>

### 3.3 SVM使用核函数

SVM之所以可以使用kernel trick,是因为a\_i.a\_j使用了核函数以后不变

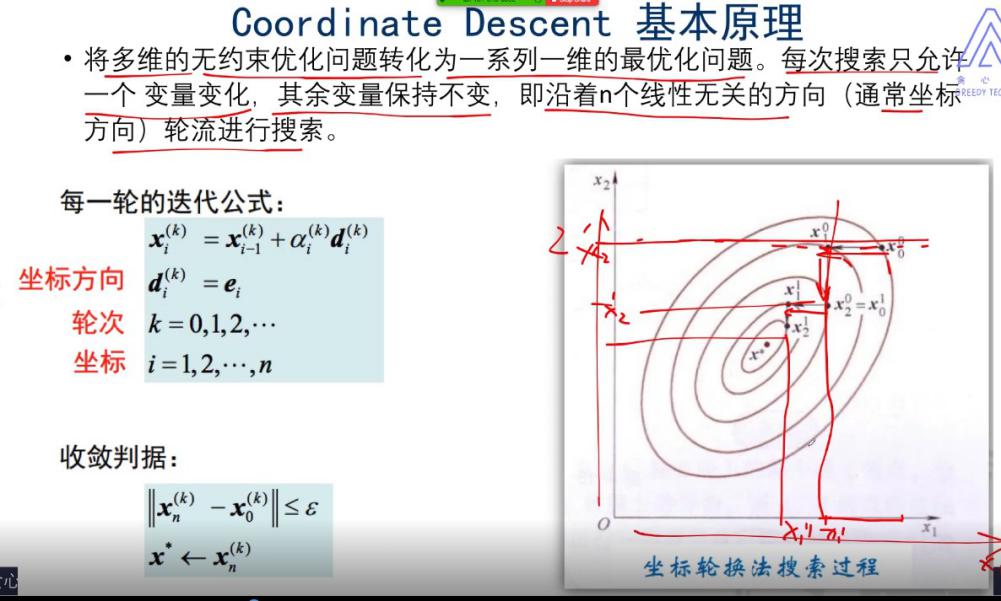


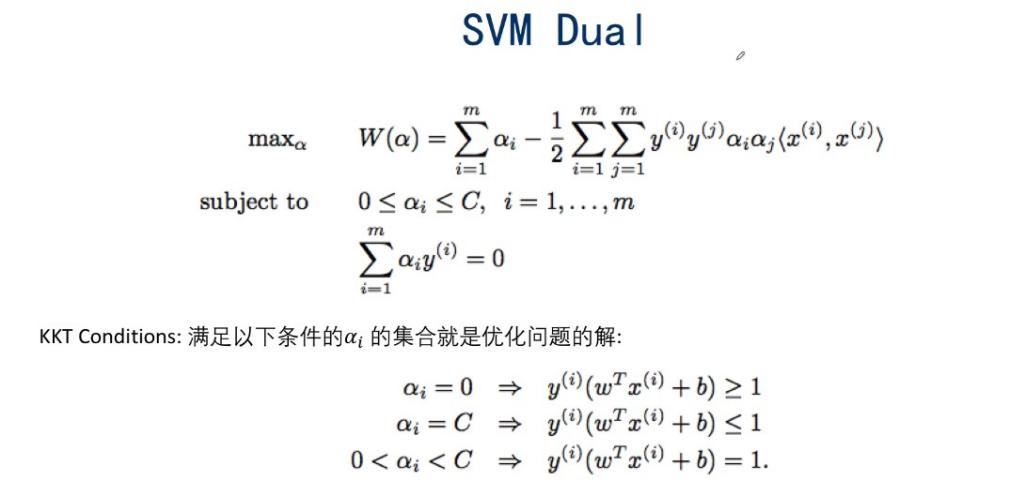


存在问题：增加了计算量，计算量和数据量与每一条数据的维度成正相关；没办法增加到无限维

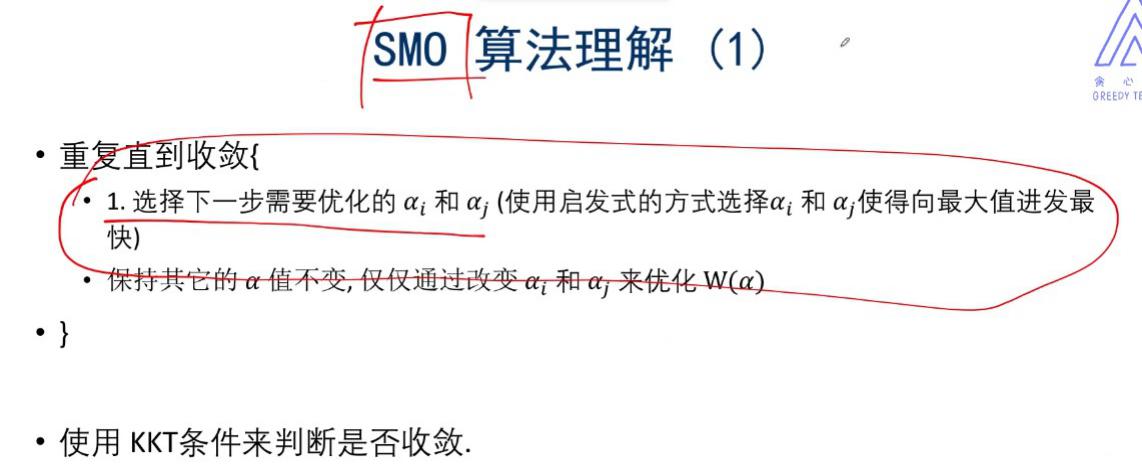
## SMO：求解SVM：二次规划（quadratic programming），在多项式时间内求得最优解

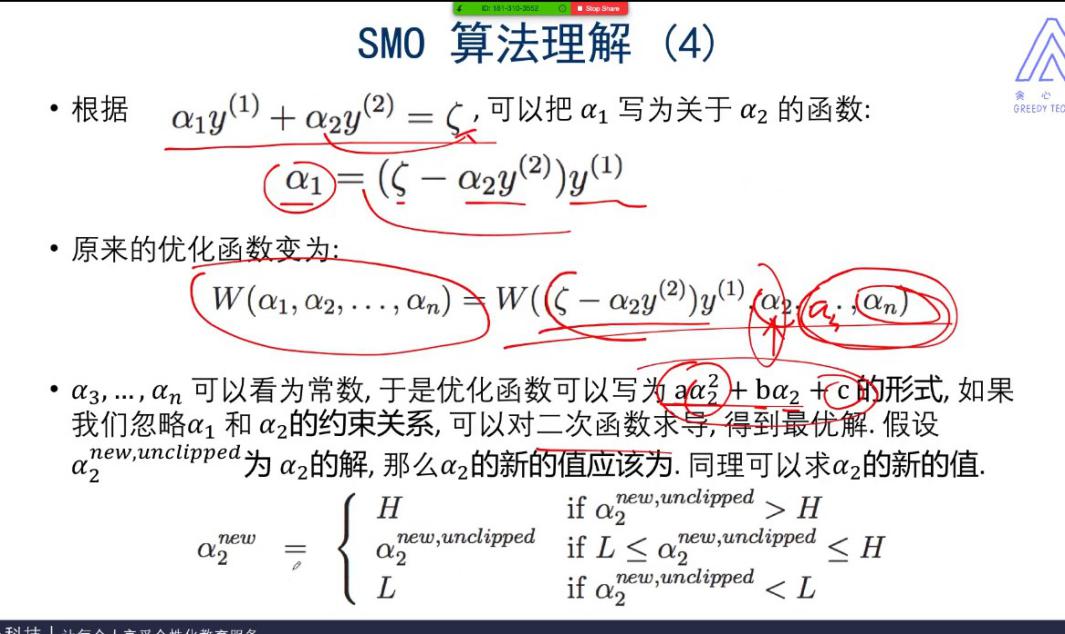
### 4.1 坐标轮换法（coordinate descent）：只允许一个变量变化，其余变量保持不变





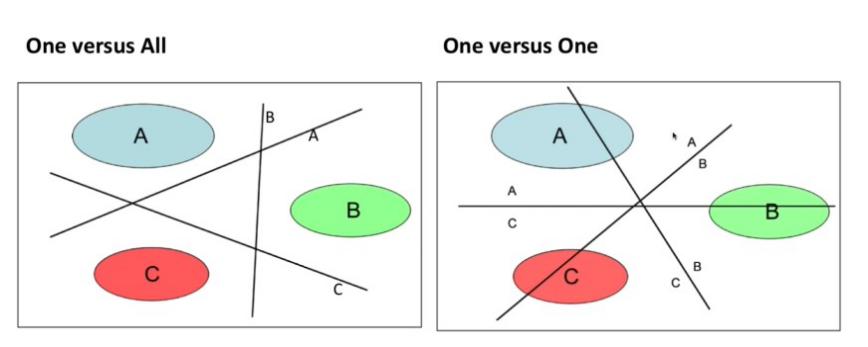
必须选用2个参数，否则回破坏KKT条件



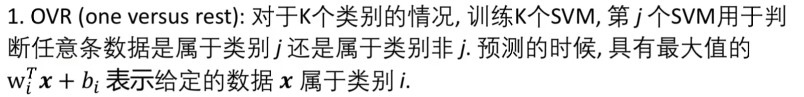


<http://cs229.stanford.edu/materials/smo.pdf>

### 4.2扩展SVM到支持多分类



### OVR（one versus rest）



### OVO（one versus one）

