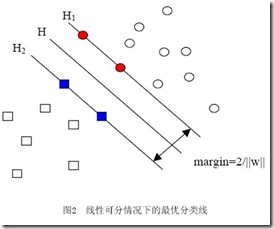
**SVM**

一、线性可分与硬间隔最大化

[](http://images.cnitblog.com/blog/458371/201212/31223651-cb3b6f9f79e34304a30c1042681d46f0.jpg)

  这个超平面可以用分类函数http://img.blog.csdn.net/20131107201211968表示，当f(x) 等于0的时候，x便是位于超平面上的点，而f(x)大于0的点对应 y=1 的数据点，f(x)小于0的点对应y=-1的点。

因此，最优超平面到点的距离为：http://img.blog.csdn.net/20131107201759093

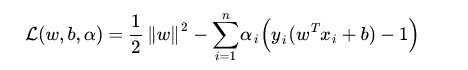
带颜色的支持向量刚好在边界上，所以它们满足http://img.blog.csdn.net/20131111155244218。

因此，H1和H2之间的距离为:2/||W||.

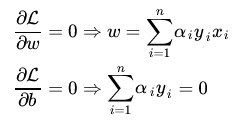
最大化此距离等价于最小化公式1：

http://img.my.csdn.net/uploads/201210/25/1351141994_1802.jpg

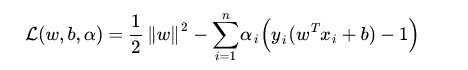
利用拉格朗日乘子法进行求解：

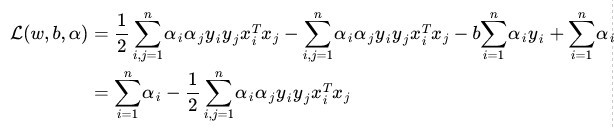


首先固定*http://img.blog.csdn.net/20131111195836468，*要让 L 关于 w 和 b 最小化，我们分别对w，b求偏导数，即令 ∂L/∂w 和 ∂L/∂b 等于零（对w求导结果的解释请看本文评论下第45楼回复）：

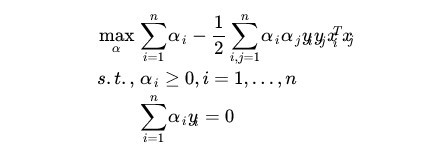


    将以上结果代入之前的L，得到：



从而有：  


由此，得到公式1的对偶问题：

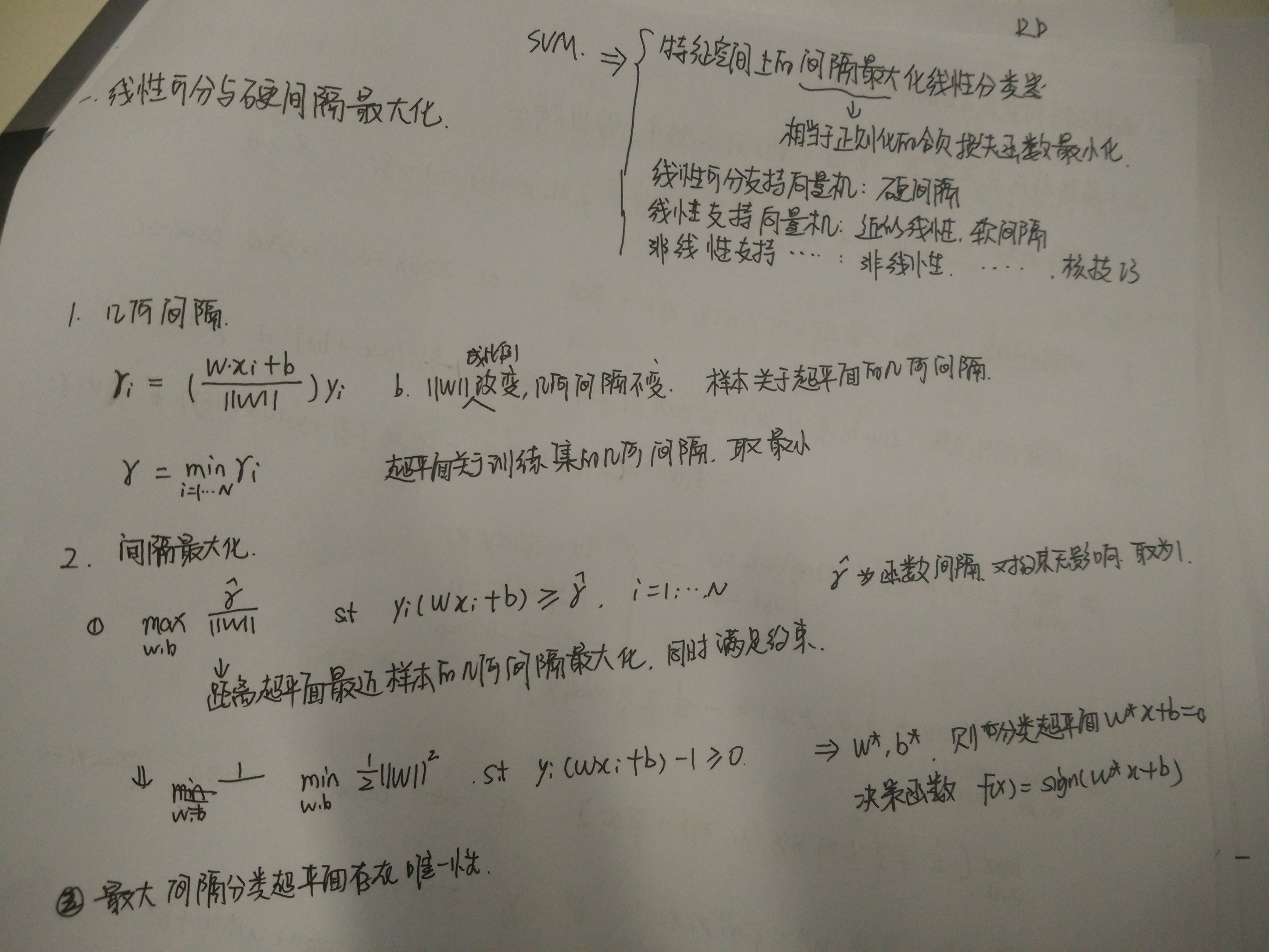


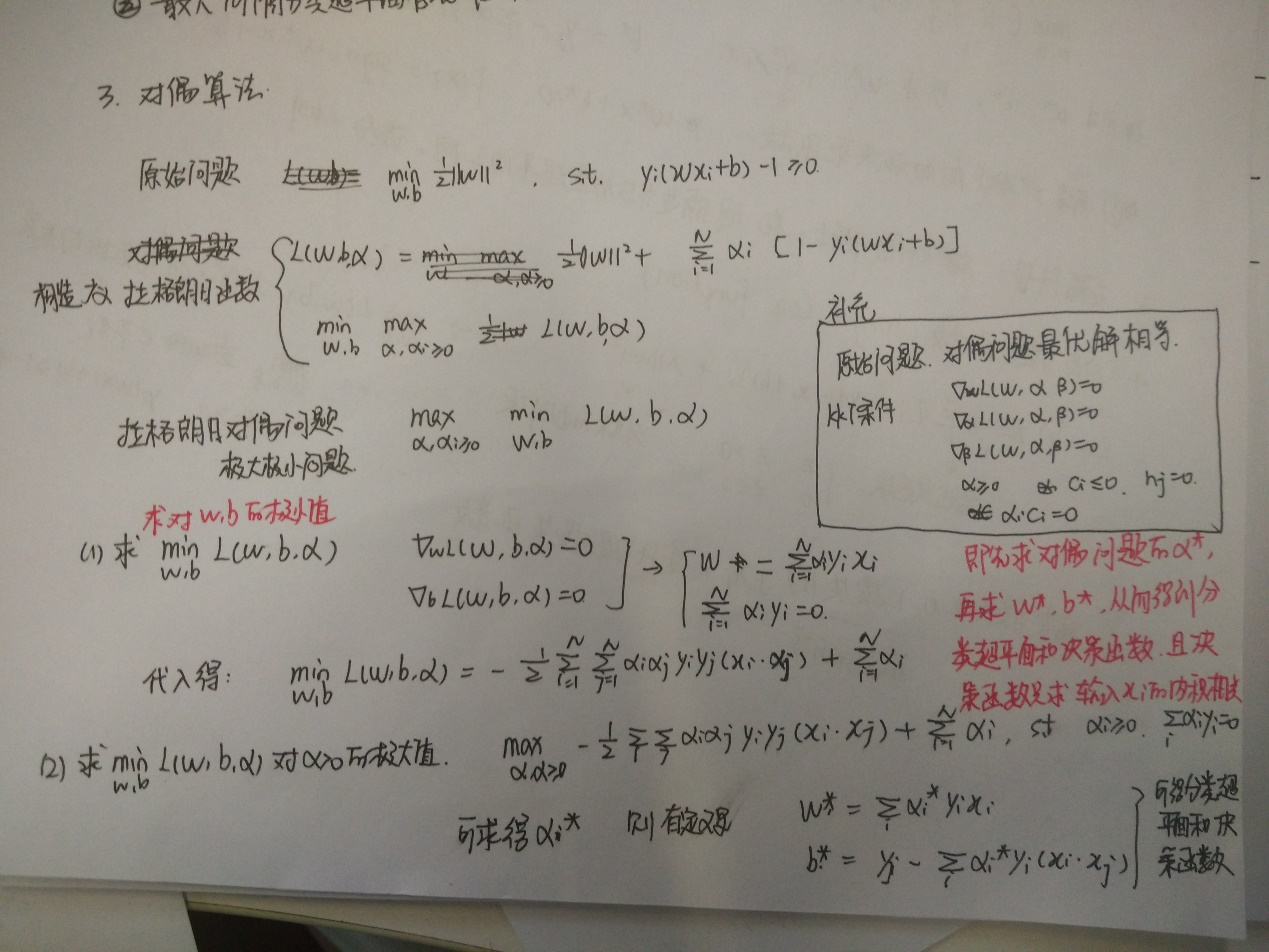
求出http://img.blog.csdn.net/20131111195824031之后，根据：

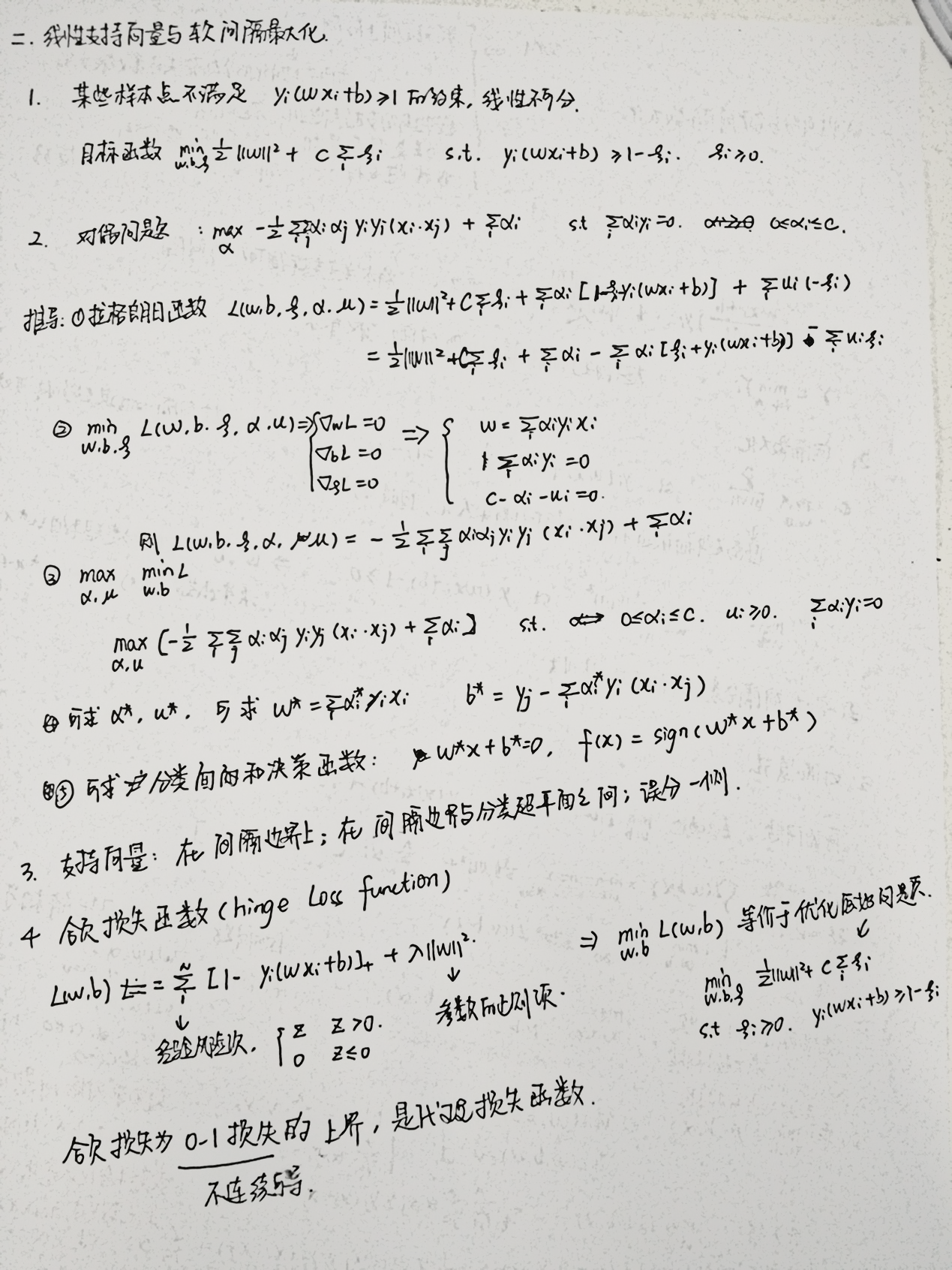
http://img.my.csdn.net/uploads/201301/11/1357838666_9138.jpg

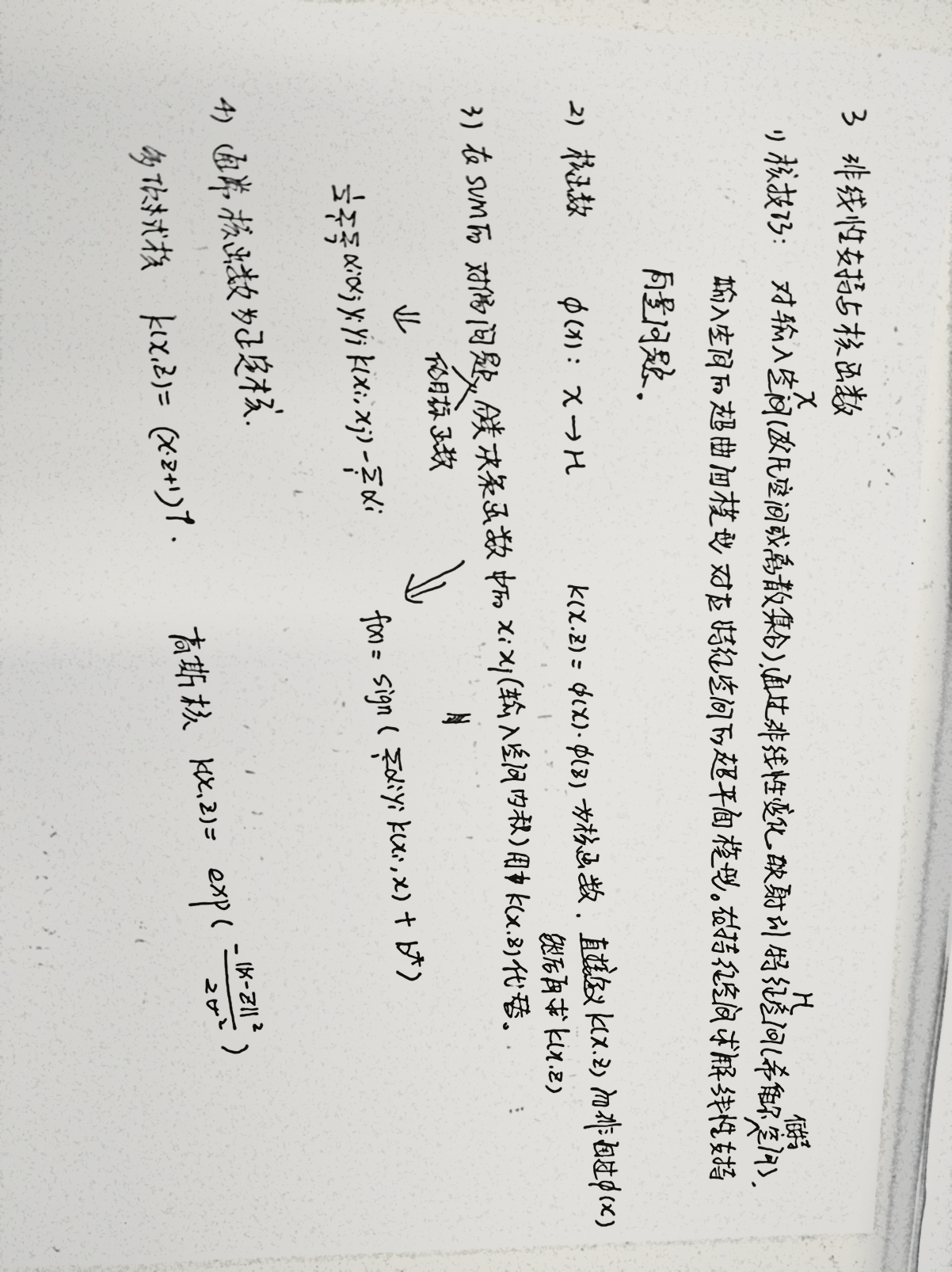
http://img.my.csdn.net/uploads/201301/11/1357838696_3314.png

就可以求出w和b，由此推出最优超平面。

****

****

****

****

四、问题与答案

1. **介绍SVM原理，核函数等**

通过最大化几何间隔来得到分类超平面和决策函数。（可结合公式）

核函数：输入空间到特征空间的映射函数为 ， ,它俩的内积，直接定义核函数，而非通过定义映射函数在计算核函数。

作用：将输入空间通过非线性映射到特征空间，在特征空间中可以通过线性支持向量的方法求解特征空间的分类超平面，即对应于输入空间的分类超曲面。

1. **谈一谈svm的对偶问题形式**

列出约束表达式

构建拉格朗日函数表达式（原始问题）

根据对偶性转化为拉格朗日极大极小问题（对偶问题）

对w/b求极小值；对 求极大值（推导公式）->可以得出 ->根据原理可以得到w\*,b\*->最终列出分类超平面函数和决策函数

对偶问题概念：一个优化问题可以从两个角度进行考察，一个是primal 问题，一个是dual 问题，就是对偶问题，一般情况下对偶问题给出主问题最优值的下界，在强对偶性成立的情况下由对偶问题可以得到主问题的最优下界，对偶问题是凸优化问题，可以进行较好的求解，SVM中就是将primal问题转换为dual问题进行求解，从而进一步引入核函数的思想。

1. **SVM在哪个地方引入核函数，SVM的Gaussian Kernel 的 dimension**

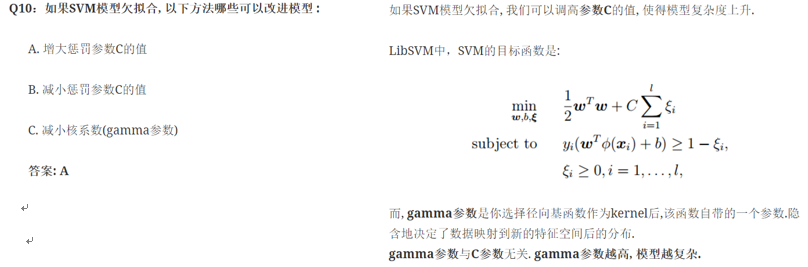
在对偶问题的目标函数和分类决策函数中将xixj核函数。

高斯核将维度映射到无穷多维。（指数根据泰勒展开可以有无穷项）

<https://www.zhihu.com/question/35602879/answer/63963315>

1. **如果SVM模型欠拟合, 可以采用哪些方法改进模型**

使模型复杂度更高。增大惩罚系数，核函数参数等。



其他：

1）添加其他特征项。有时模型出现欠拟合是因为特征项不够导致的。例如，“组合”、“泛化”、“相关性”三类特征是特征添加的重要手段。

2）添加多项式特征，例如将线性模型通过添加二次项或者三次项使模型泛化能力更强。（增加模型复杂度）

3）减少正则化参数，正则化的目的是用来防止过拟合的，但是现在模型出现了欠拟合，则需要减少正则化参数。

[**http://blog.csdn.net/willduan1/article/details/53070777**](http://blog.csdn.net/willduan1/article/details/53070777)

1. **训练完SVM模型后, 不是支持向量的那些样本我们可以丢掉, 也可以继续分类么？**

正确。

SVM模型中, 真正影响决策边界的是支持向量

1. **Svm的最大间隔，软间隔，对偶**

软间隔是允许错分类存在，并给与错分类样本惩罚

具体原理见上

1. **常用的核函数有哪些，说说它们的特点及调参方式**

**线性核函数：**简单，速度快，参数少，需要线性可分

**多项式核函数：**比线性核拟合程度更强，知道具体的维度，但是高次容易出现数值不稳定，参数选择比较多。

**高斯核函数：**是应用最广的一个。无论样本大小都有较好的性能，而且相对于多项式核函数参数少（只需要一个参数）。拟合能力强，要注意过拟合问题。大多数不知道用什么核函数的时候，优先使用高斯核函数。

**拉普拉斯核函数：**

**sigmoid核函数：**

选用核函数时，如果对数据有一定的先验知识，就利用先验来选择符合数据分布的核函数；如果不知道，通常使用交叉验证的方法，来试用不同的核函数，误差最小的即为效果最好的核函数，或者也可以将多个核函数结合起来，形成混合核函数。在吴恩达的课上，也曾经给出过一系列的选择核函数的方法：

* 如果特征的数量大到和样本数量差不多，则选用LR或者线性核的SVM；
* 如果特征的数量小，样本的数量正常，则选用SVM+高斯核函数；
* 如果特征的数量小，而样本的数量很大，则需要手工添加一些特征从而变成第一种情况。

[**http://blog.csdn.net/batuwuhanpei/article/details/52354822**](http://blog.csdn.net/batuwuhanpei/article/details/52354822)

1. **SVM 在什么情况下效果会比较差**

大样本

1. **SVM 的推导，特性？多分类怎么处理**

**特性：**

(1)非线性映射：利用内积核函数代替高维空间的非线性映射；  
(2)最大化分类边界是SVM方法的核心；  
(3)支持向量是SVM的训练结果,在SVM分类决策中起决定作用的是支持向量。  
(4)SVM 是一种有坚实理论基础的新颖的小样本学习方法。它基本上不涉及概率测度及大数定律等,因此不同于现有的统计方法。从本质上看,它避开了从归纳到演绎的传统过程,实现了高效的从训练样本到预报样本的“转导推理”,大大简化了通常的分类和回归等问题。  
(5)SVM 的最终决策函数只由少数的支持向量所确定,计算的复杂性取决于支持向量的数目,而不是样本空间的维数,这在某种意义上避免了“维数灾难”。  
(6)少数支持向量决定了最终结果,这不但可以帮助我们抓住关键样本、“剔除”大量冗余样本,而且注定了该方法不但算法简单,而且具有较好的“鲁棒”性。这种“鲁棒”性主要体现在:  
①增、删非支持向量样本对模型没有影响;  
②支持向量样本集具有一定的鲁棒性;  
③有些成功的应用中,SVM 方法对核的选取不敏感

**优点**:适用于线性非线性分类（非线性用核函数），适合小样本，泛化能力好

**缺点**:如何确定核函数,现在还没有合适的方法；确定核函数后,在求解问题分类时,要求解函数的二次规划,需要大量的存储空间，因此大样本难实施。

**多分类问题：**

a.一对多法。某类别样本归为一类,其余样本归为另一类。k个类别的样本就构造出k个SVM。分类时将未知样本分类为具有最大分类函数值的类。

b.一对一法。任意两类样本之间设计一个SVM，k个类别需设计k(k-1)/2个SVM。当对一个未知样本进行分类时，最后得票最多的类别即为该未知样本的类别。Libsvm中的多类分类就是根据这个方法实现的。