支持向量机

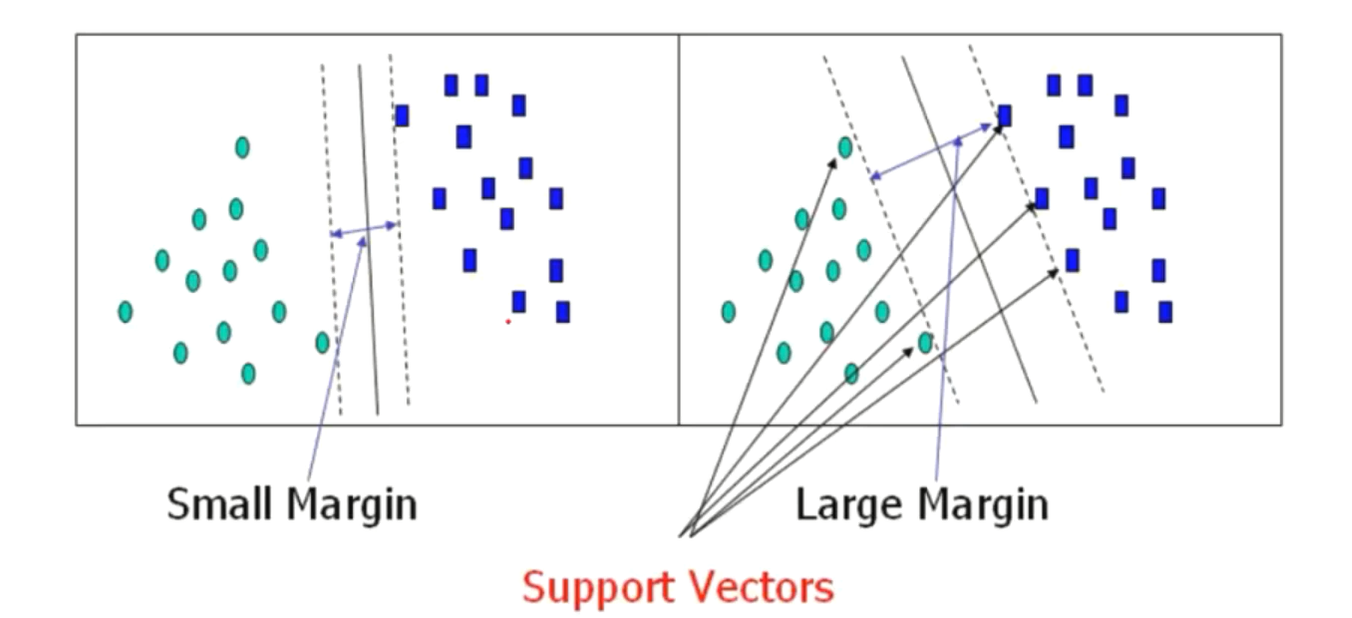
Suupport Vector Machine

1. 支持向量机的作用

目的：做分类，并且分类的边界越宽越好

何谓边界的宽度呢？请看下图：

大家可以看到，左图的边界宽度明显比右图的宽度要小很多，边界越宽当然越好，即分类的泛性越大，即若有更多的点需要分类，那么右图的分类边界表现一般是更好的。



1. 推导边界点到分类平面的距离公式

要得到最宽的分类边界，跟最靠近分类边界的点有关，因为分类的边界本就是为了分割开不同种类的点，而最靠近边界的点就在边界的最边上。

那么，既然说最靠近边界的点与边界的宽度有关，如何求最靠近边界的点与边界的距离呢？

（一）利用平面向量与法向量

这里，我们可以把边界想象成一个平面，那么要求分类的点与边界的距离，就是求点到平面的距离。点到平面的距离是如何定义的呢？

这里来推导一下公式。

假设我们要求的点为x, 假设边界为平面，

平面的一般公式大家应该都知道，即y = wTx + b， wT为平面的法向量，b为偏置量。

在平面上假设有两个点 x1, x2, 将两个点代入平面的公式中，即

wTx1 + b = 0

wTx2 +b = 0

同时，两个点组成了平面上的一个向量，而平面的法向量wT是垂直于平面上的向量的，因此有：

(wT, (x2 – x1)) = 0

有了平面上的一个向量x2 – x1，因此，我们要求的点x到平面的距离就可以配合着这个向量x2-x1来求，连接点x与点X1，构造一个向量(x-x1)，通过向量(x-x1)在平面的法向量上的投影即可求得点X到边界平面的距离。

根据向量的乘积公式可以得到向量(x-x1)与wT的乘积为

wT(x-x1) = ||wT||||x-x1||costheta

theta 为两个向量的夹角

因为我们要求的就是向量(x-x1)在法向量上的投影距离，即||x-x1||costheta，

那么将右式中的||wT||除过去即可得到。同时，考虑到向量乘积可能有负值，为数值添加绝对值符号，即可以求得点X到边界平面的距离为：  
distance = |wT(x-x1)/||w|||

将x1, x2两点的平面式子wTx1 = -b, wTx2 = -b代入，可得

Distance = |wTx+b| / ||w||

（二）去掉绝对值

支持向量机要优化的目标是：  
每种边界都有一个离边界最近的点，我们要做的优化就是找出哪个边界的最近点离边界的距离是最大的，那么该边界就是我们要求的边界。

边界的方程为 y = wTx + b

那么，有了优化的目标，我们就要朝着这个目标来构造数学公式了。

上面已经求出

Distance = |wTx+b| / ||w||

绝对值比较影响我们的求解，要想办法将绝对值去掉。去掉绝对值，即将负数变为正数，正数依然是正数，那么我给负数乘个-1，给正数乘个+1，不就把绝对值去掉了么！

而-1和+1正好和负例与正例相对应，那么我就将数据集的样本标签yi正例设为+1， 负例设为-1。同时，设置决策分类的结果，

设置决策结果y(xi)为>0时为正例，正好对应了数据集的正例标签为+1；

设置决策结果y(xi)为<0时为负例，正好对应了数据集的负例标签为-1；

因此，给distance的分子|wTxi+b|（即y(xi)）乘以yi，即 yi y(xi) > 0，这样就将绝对值给去掉了。

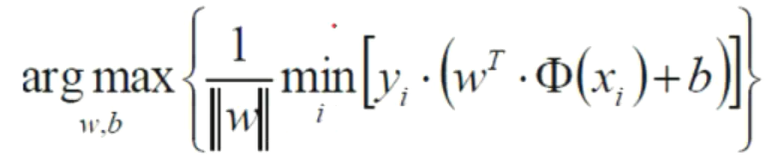
那么现在化简的distance式子为

Distance = y(wTxi+b) / ||w||

1. 目标函数

（一）设置目标函数

去掉了绝对值，就要来设置我们的目标函数了，根据我们之前提出的优化目标，可以写出相应的数学式子：



理解：大括号里面求的是每个分类边界的最近点到边界的距离，大括号外面是求什么样的w, b能够使得某个分类边界的最近点到边界的距离最大。

（二）简化目标函数

这个目标式子过于复杂，我们要想办法简化，因为[y (wTxi+b)] > 0，我们可以通过放缩，使得

[y(wTxi+b)] >=1，

这样，优化目标式子中大括号里面的 min[yi(wTxi+b)]就简化成了1

目标函数简化为：

Argmax 1 / ||w||

目标函数为求解极大值问题，但是，我们一般求解的都是极小值问题，因此，将求解1/||w||的极大值问题转换为求解||w||极小值问题。即

Min w,b ||w||， 且前面假设了 yi(wTxi+b) >= 1这个条件

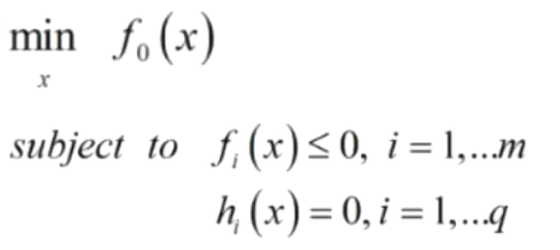
考虑到之后要求导，且求解 ||w||的极小值问题与求解||w2||的极小值问题是一样的。则把目标函数变化为：  
min w,b 1/2 w2， 条件为 yi(wTxi+b) >= 1

（三）求解目标函数

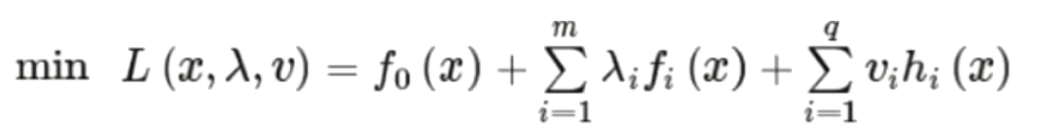
现在就要来求解这个方程了，一看这个问题，是不是很熟悉，考研的同学应该比较熟，用拉格朗日乘子法来求解。

1. 拉格朗日乘子法

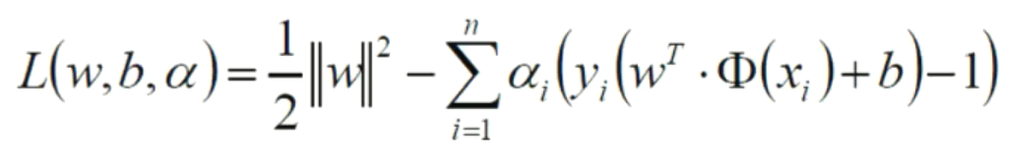
拉格朗日乘子法用于求解极值问题，如下所示，求解f0(x)的极小值，下面是约束条件，可以为不等式或者等式。

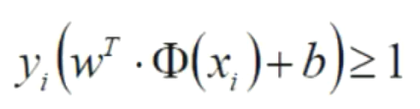


这是拉格朗日乘子法的标准式子：



这是将目标函数和约束条件带入拉格朗日乘子法公式：

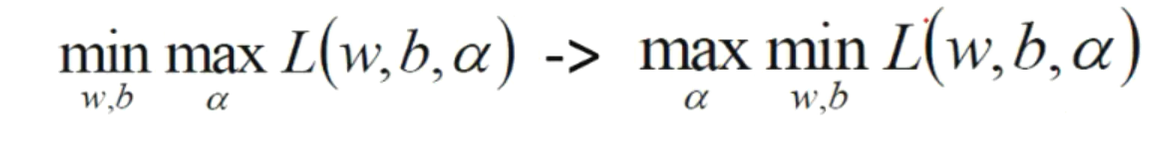




注意，多了参数 二发

由于直接求解目标函数比较困难，通过拉格朗日乘子法求解，将求解w,b与中间参数 二发 联系起来，那么通过求解 二发 的解，然后通过二发 与 w,b的关系，那么w,b姐可以求解了。

拉格朗日乘子法， 求解方法：

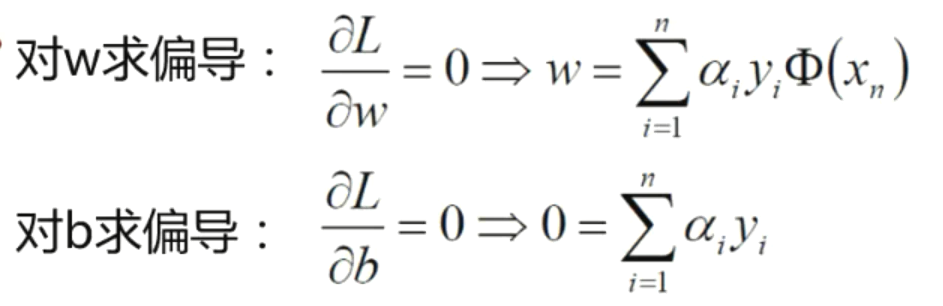


对拉格朗日乘子法的式子先求最大值，再求极小值等价于对它先求极小值，再求极大值。

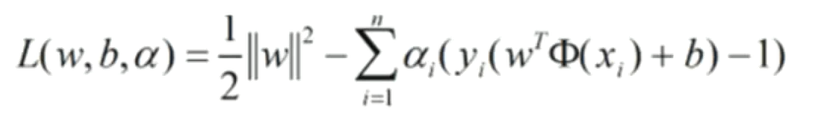
2.拉格朗日乘子法先求极小值

因此，我们就先求L的极小值，即求解什么样的w,b能够使得L最小。

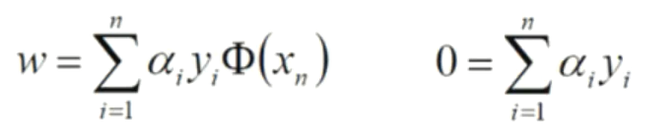
求解极小值，就对L求偏导，有几个参数，就对几个参数分别求偏导，

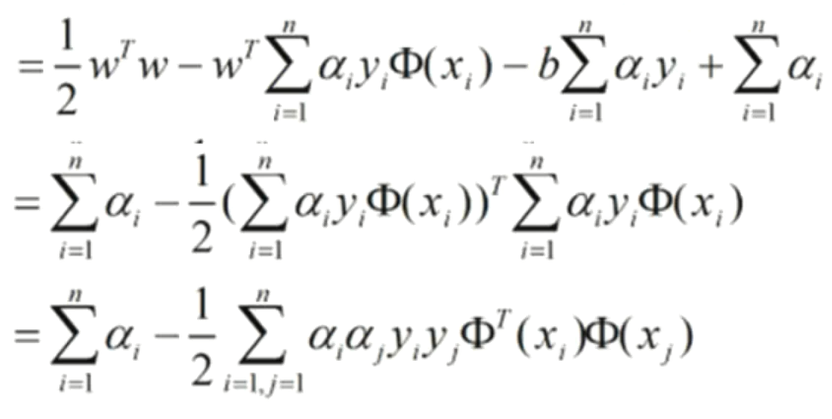


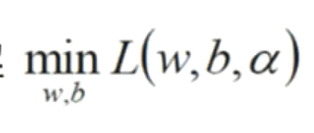
将求出的等式带入原式中



有前面的对w,b求偏导得出的等式带入

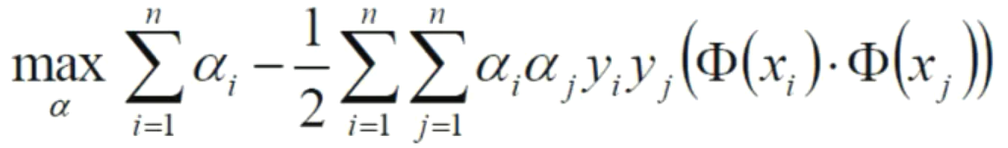


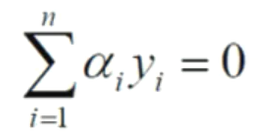


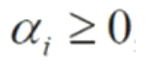
到此，我们完成了拉格朗日乘子法的第一步，求解

3.拉格朗日乘子法再求极大值

接下来求解第二步，求解什么样的 二发 使得 上式结果最大

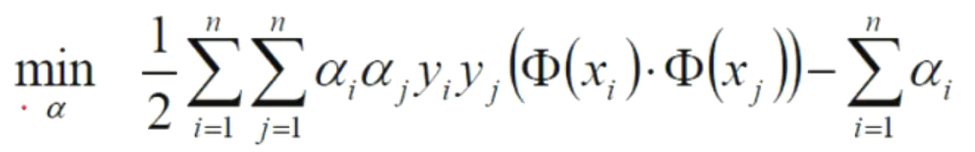


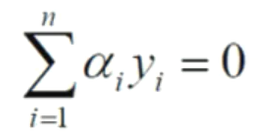
限制条件：  


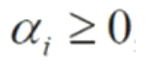


第一个限制条件是之前对w求偏导得出的。

第二个限制条件是拉格朗日乘子法规定的 参数 二发 必须 大于等于 0

又是求解极大值问题，将极大值转变为求解极小值问题，即添个负号即可转变：  


限制条件不变：  


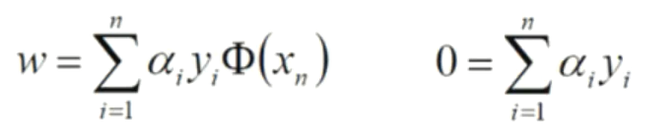


通过求解极小值，求出了 二发 ，就可以求出w

四．支持向量机名字由来

讲到这里，大家对支持向量机这个名称是不是又疑问？明明是一个分类边界的算法为什么叫做支持向量机？

我们可以从w的公式中看出：



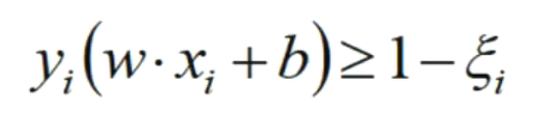
它与 二发， y， x的值有关，可以知道x和y的值不会为0，而 二发的值可能为0.当 二发的值为0时，那么该点就与分类边界的形成毫无关系，就不是支持分类边界生成的支持向量了。那些 二发 不为 0的点都是支持分类边界生成的支持向量，同时，这些点就是边界点，正式边界点对分类边界产生影响。这就是支持向量机这个名称的由来。

五．软间隔

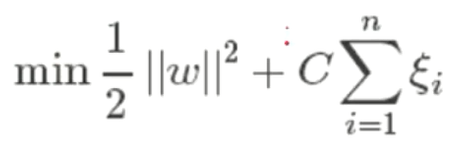
软间隔：数据中有一些噪音点会造成严格的分类，导致更差的分类边界生成。

因此，为了获得最好的分类边界，可以设置分类没有那么严格，可以容忍一些错误分类。

引入 松弛因子，使得



那么，目标函数也会因此有所变化：



引入了控制参数C：（C是我们自己指定的参数）

当C值很大时，为了使目标函数最小，，因此松弛因子就要更小，意味着分类更加严格。

当C值很小时，乘以一个稍微大一点的数也没有影响。因此松弛因此可以适当放大，意味着可以容忍一些错误。

同样，使用拉格朗日乘子法来求解。

六．核变换

低维不可分问题

在低维度的问题中，有些问题用线性分类是无法分类的，那么我们可以将该问题转换到高维度上进行分类。这就叫做核变换。

虽然说是将低维数据映射到高维，但是依旧是在低维度进行计算，免去了高维度计算的烦恼。