<http://blog.sina.com.cn/s/blog_4298002e010144k8.html>

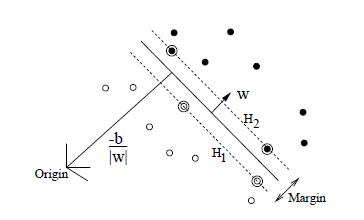
支持向量机（SVM）的详细推导过程及注解

http://simg.sinajs.cn/blog7style/images/common/sg_trans.gif (2012-03-19 15:57:24)

[[http://simg.sinajs.cn/blog7style/images/common/sg_trans.gif](javascript:;)转载▼](javascript:;)

|  |  |
| --- | --- |
| 标签：  [svm](http://search.sina.com.cn/?c=blog&q=svm&by=tag)    [支持向量机](http://search.sina.com.cn/?c=blog&q=%D6%A7%B3%D6%CF%F2%C1%BF%BB%FA&by=tag)    [拉格朗日](http://search.sina.com.cn/?c=blog&q=%C0%AD%B8%F1%C0%CA%C8%D5&by=tag)    [对偶问题](http://search.sina.com.cn/?c=blog&q=%B6%D4%C5%BC%CE%CA%CC%E2&by=tag)    [不等式条件极值](http://search.sina.com.cn/?c=blog&q=%B2%BB%B5%C8%CA%BD%CC%F5%BC%FE%BC%AB%D6%B5&by=tag)    [凸规划](http://search.sina.com.cn/?c=blog&q=%CD%B9%B9%E6%BB%AE&by=tag)  [模式识别](http://search.sina.com.cn/?c=blog&q=%C4%A3%CA%BD%CA%B6%B1%F0&by=tag) | 分类： [无尽之海](http://blog.sina.com.cn/s/articlelist_1117257774_8_1.html) |

    支持向量机的原理很简单，就是VC维理论和最小化结构风险。在阅读相关论文的时候，发现很多文章都语焉不详，就连《A Tutorial on Support Vector Machines for Pattern Recognition》这篇文章对拉格朗日条件极值问题的对偶变换都只是一笔带过，让很多人觉得很困惑。下面我将就SVM对线性可分的情况作详尽的推导。

[](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s2.sinaimg.cn/orignal/4298002etbb8e72012971)

    如上图所示，有一堆训练数据的正负样本，标记为：[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s4.sinaimg.cn/orignal/4298002etbb93f6c484c3)，假设有一个超平面H：[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s2.sinaimg.cn/orignal/4298002etbb93fb859181)，可以把这些样本正确无误地分割开来，同时存在两个平行于H的超平面H1和H2：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s13.sinaimg.cn/orignal/4298002etbb93fefd840c)

使离H最近的正负样本刚好分别落在H1和H2上，这样的样本就是支持向量。那么其他所有的训练样本都将位于H1和H2之外，也就是满足如下约束：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s11.sinaimg.cn/orignal/4298002etbb940783ac1a)

写成统一的式子就是：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s14.sinaimg.cn/orignal/4298002etbb940b643ecd) （1）

而超平面H1和H2的距离可知为：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s4.sinaimg.cn/orignal/4298002etbb9413164c33)

SVM的任务就是寻找这样一个超平面H把样本无误地分割成两部分，并且使H1和H2的距离最大。要找到这样的超平面，只需最大化间隔Margin，也就是最小化[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s13.sinaimg.cn/orignal/4298002etbb941600573c)。于是可以构造如下的条件极值问题：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s12.sinaimg.cn/orignal/4298002etbb941986222b) （2）

    对于不等式约束的条件极值问题，可以用拉格朗日方法求解。而拉格朗日方程的构造规则是：用约束方程乘以非负的拉格朗日系数，然后再从目标函数中减去。于是得到拉格朗日方程如下：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s5.sinaimg.cn/orignal/4298002etbb941ff3c174) （3）

其中：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s14.sinaimg.cn/orignal/4298002etbb942346bc6d) （4）

那么我们要处理的规划问题就变为：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s8.sinaimg.cn/orignal/4298002etbb942902cca7) （5）

    上式才是严格的不等式约束的拉格朗日条件极值的表达式。对于这一步的变换，很多文章都没有多做表述，或者理解有偏差，从而影响了读者后续的推演。在此我将详细地一步步推导，以解困惑。

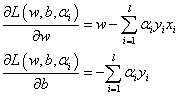
    （5）式是一个凸规划问题，其意义是先对α求偏导，令其等于0消掉α，然后再对w和b求L的最小值。要直接求解（5）式是有难度的，通过消去拉格朗日系数来化简方程，对我们的问题无济于事。所幸这个问题可以通过拉格朗日对偶问题来解决，为此我们把（5）式做一个等价变换：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s8.sinaimg.cn/orignal/4298002etbb94327ddf97)

上式即为对偶变换，这样就把这个凸规划问题转换成了对偶问题：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s16.sinaimg.cn/orignal/4298002etbb94350d0daf) （6）

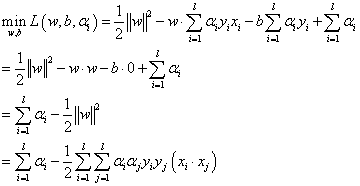
其意义是：原凸规划问题可以转化为先对w和b求偏导，令其等于0消掉w和b，然后再对α求L的最大值。下面我们就来求解（6）式，为此我们先计算w和b的偏导数。由（3）式有：

[](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s14.sinaimg.cn/orignal/4298002etbb9439774fbd) （7）

为了让L在w和b上取到最小值，令（7）式的两个偏导数分别为0，于是得到：

[](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s10.sinaimg.cn/orignal/4298002etbb943cb4d749) （8）

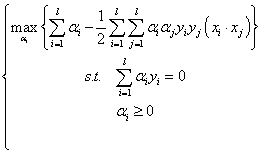
将（8）代回（3）式，可得：

 （9）

再把（9）代入（6）式有：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s5.sinaimg.cn/orignal/4298002etbb94478acc54) （10）

考虑到（8）式，我们的对偶问题就变为：

[](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s1.sinaimg.cn/orignal/4298002etbb944b640330) （11）

    上式这个规划问题可以直接从数值方法计算求解。

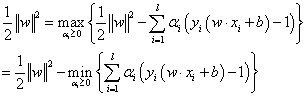
    需要指出的一点是，（2）式的条件极值问题能够转化为（5）式的凸规划问题，其中隐含着一个约束，即：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s3.sinaimg.cn/orignal/4298002etbb9452cdf422) （12）

这个约束是这样得来的，如果（2）和（5）等效，必有：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s2.sinaimg.cn/orignal/4298002etbb945b9dd4b1)

把（3）式代入上式中，得到：

[](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s2.sinaimg.cn/orignal/4298002etbb945ba1a291)

化简得到：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s4.sinaimg.cn/orignal/4298002etbb945bbebd73) （13）

又因为约束（1）式和（4）式，有：

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s5.sinaimg.cn/orignal/4298002etbb945bad05f4)

所以要使（13）式成立，只有令：支持向量机（SVM）的详细推导过程及注解（一），由此得到（12）式的约束。该约束的意义是：如果一个样本是支持向量，则其对应的拉格朗日系数非零；如果一个样本不是支持向量，则其对应的拉格朗日系数一定为0。由此可知大多数拉格朗日系数都是0。

    一旦我们从（11）式求解出所有拉格朗日系数，就可以通过（8）式的

[支持向量机（SVM）的详细推导过程及注解（一）](http://photo.blog.sina.com.cn/showpic.html#blogid=4298002e010144k8&url=http://s2.sinaimg.cn/orignal/4298002etbb945b8b2a01)

计算得到最优分割面H的法向量w。而分割阈值b也可以通过（12）式的约束用支持向量计算出来。这样我们就找到了最优的H1和H2，这就是我们训练出来的SVM。