

支持向量机 (SVM) 的详细推导过程及注解

更新时间: 2018-03-03 21:51:01

[点击这里!!! \[厉害了\]牛人都关注了他们, 伴你前行的技术良友](#)[点击这里!!! 改变命运: 免费人工智能机器学习深度学习Python视频2018最新版](#)[冲刺80万年薪, 扫码公众号回复: python, 获得视频教程||★☆☆☆](#)

[马开东搜索--为百万程序员, 站长服务, 关注Python, Java及Web框架, 大数据Hadoop MR Hive Hbase Spark Storm, 自然语言NLP, 机器学习 K-means 朴素贝叶斯 SVM 随机森林等, 推荐系统协同过滤, 神经网络CNN RNN, 深度学习TensorFlow, IT互联网, 科技资讯, 想你所想, 知你不知, 共享信息, 改变世界! 做个用代码改变世界的程序员, 我喂自己袋盐。](#)



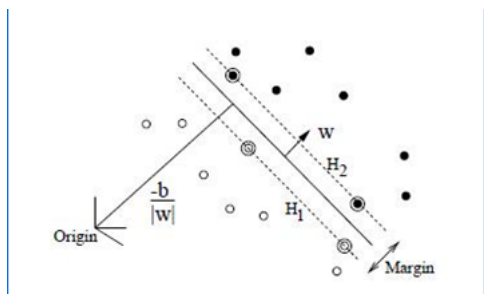
数据分析师

CPDA 数据分析行业权威认证

百度广告

[我是搬运工: http://my.oschina.net/wangguolongnk/blog/111353](http://my.oschina.net/wangguolongnk/blog/111353)

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。



如上图所示, 有一堆训练数据的正负样本, 标记为: $\{x_i, y_i\}, i = 1, \dots, l, y_i \in \{-1, 1\}, x_i \in R^d$, 假设有一个超平面H: $w \cdot x + b = 0$, 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:

$$\begin{aligned} w \cdot x + b &= 1 \\ w \cdot x + b &= -1 \end{aligned}$$

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:

$$\begin{cases} w \cdot x_i + b \geq 1 & \text{for } y_i = 1 \\ w \cdot x_i + b \leq -1 & \text{for } y_i = -1 \end{cases}$$

写成统一的式子就是：

$$y_i (w \cdot x_i + b) - 1 \geq 0 \quad (1)$$

而超平面H1和H2的距离可知为：

$$\text{Margin} = 2 / \|w\|$$

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 $\|w\|^2$ 。于是可以构造如下的条件极值问题：

$$\begin{cases} \min \|w\|^2 / 2 \\ \text{s.t. } y_i (w \cdot x_i + b) - 1 \geq 0 \end{cases} \quad (2)$$

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：

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y_i (w \cdot x_i + b) - 1) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i y_i (w \cdot x_i + b) + \sum_{i=1}^l \alpha_i \quad (3)$$

其中：

$$\alpha_i \geq 0 \quad (4)$$

那么我们要处理的规划问题就变为：

$$\min_{w, b} \max_{\alpha_i \geq 0} L(w, b, \alpha_i) \quad (5)$$

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：

$$\min_{w, b} \max_{\alpha_i \geq 0} L(w, b, \alpha_i) = \max_{\alpha_i \geq 0} \min_{w, b} L(w, b, \alpha_i)$$

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：

$$\max_{\alpha_i \geq 0} \min_{w, b} L(w, b, \alpha_i) \quad (6)$$

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：

$$\begin{cases} \frac{\partial L(w, b, \alpha_i)}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i \\ \frac{\partial L(w, b, \alpha_i)}{\partial b} = -\sum_{i=1}^l \alpha_i y_i \end{cases} \quad (7)$$

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：

$$\begin{cases} w = \sum_{i=1}^l \alpha_i y_i x_i \\ \sum_{i=1}^l \alpha_i y_i = 0 \end{cases} \quad (8)$$

将（8）代回（3）式，可得：

$$\begin{aligned} \min_{w, b} L(w, b, \alpha_i) &= \frac{1}{2} \|w\|^2 - w \cdot \sum_{i=1}^l \alpha_i y_i x_i - b \sum_{i=1}^l \alpha_i y_i + \sum_{i=1}^l \alpha_i \\ &= \frac{1}{2} \|w\|^2 - w \cdot w - b \cdot 0 + \sum_{i=1}^l \alpha_i \\ &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \|w\|^2 \\ &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \end{aligned} \quad (9)$$

再把（9）代入（6）式有：

$$\max_{\alpha \geq 0} \min_{w, b} L(w, b, \alpha_i) = \max_{\alpha \geq 0} \left\{ \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \right\} \quad (10)$$

考虑到（8）式，我们的对偶问题就变为：

$$\begin{cases} \max_{\alpha} \left\{ \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \right\} \\ s.t. \quad \sum_{i=1}^l \alpha_i y_i = 0 \\ \alpha_i \geq 0 \end{cases} \quad (11)$$

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—个约束，即：

$$\alpha_i (y_i (w \cdot x_i + b) - 1) = 0 \quad (12)$$

这个约束是这样得来的，如果（2）和（5）等效，必有：

$$\max_{\alpha \geq 0} L(w, b, \alpha_i) = \frac{1}{2} \|w\|^2$$

把（3）式代入上式中，得到：

$$\begin{aligned} \frac{1}{2} \|w\|^2 &= \max_{\alpha \geq 0} \left\{ \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y_i (w \cdot x_i + b) - 1) \right\} \\ &= \frac{1}{2} \|w\|^2 - \min_{\alpha \geq 0} \left\{ \sum_{i=1}^l \alpha_i (y_i (w \cdot x_i + b) - 1) \right\} \end{aligned}$$

化简得到：



$$\min_{\alpha} \left\{ \sum_{i=1}^I \alpha_i (y_i (w \cdot x_i + b) - 1) \right\} = 0 \quad (13)$$

又因为约束 (1) 式和 (4) 式, 有:

$$\alpha_i (y_i (w \cdot x_i + b) - 1) \geq 0$$

所以要使 (13) 式成立, 只有令: $\alpha_i (y_i (w \cdot x_i + b) - 1) = 0$, 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

$$w = \sum_{i=1}^I \alpha_i y_i x_i$$

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。

线性 SVM 和线性可分数据集:

➤ 原问题: $\min \frac{1}{2} (w \cdot w)$
 $s.t. y_i (w \cdot x_i + b) \geq 1 \quad i = 1, \dots, N$

➤ 对偶问题: $\max W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$
 $s.t. \sum_{i=1}^N \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad i = 1, \dots, N$

线性 SVM 和线性不可分数据集:

➤ 原问题: $\min \frac{1}{2} (w \cdot w) + C \sum_{i=1}^N \xi_i$
 $s.t. y_i (w \cdot x_i + b) \geq 1 - \xi_i \quad i = 1, \dots, N$

➤ 对偶问题: $\max W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$
 $s.t. \sum_{i=1}^N \alpha_i y_i = 0, \quad C \geq \alpha_i \geq 0, \quad i = 1, \dots, N$

C-SVM

采用软间隔和核函数两种方法所做出的分类器有着更好的适应性。这时的原问题和对偶问题为:

➤ 原问题: $\min \frac{1}{2} (w \cdot w) + C \sum_{i=1}^N \xi_i$
 $s.t. y_i (w \cdot x_i + b) \geq 1 - \xi_i \quad i = 1, \dots, N$

➤ 对偶问题: $\max W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j)$
 $s.t. \sum_{i=1}^N \alpha_i y_i = 0, \quad C \geq \alpha_i \geq 0, \quad i = 1, \dots, N$

这样的分类器称为非线性软间隔支持向量分类。



本地美女正在直播热舞

身边同事们每天晚上必看的直播, 到底有什么不一样呢?

百度广告

此文链接: http://makaidong.com/JustForCS/0/636469_3428618.html

转载请注明出处: 支持向量机 (SVM) 的详细推导过程及注解

来源: 马开东云搜索 (电话: 15110131480 微信: makaidongzi QQ: 1130122167 微信公众号: makaidong-com)

欢迎分享本文, 转载请保留出处!

加入QQ群 粉丝交流QQ群:

免责声明: 本站仅提供平台, 所有内容均来自互联网收集或网友原创、转发而来, 版权归原创作者所有, 本站不承担任何由于内容的侵权, 合法性及所引起的争议和法律责任