

# 支持向量机: Duality

by pluskid, on 2010-11-02, in Machine Learning 69 comments

本文是“支持向量机系列”的番外篇(1), 参见本系列的[其他文章](#)。

在之前关于 support vector 的推导中, 我们提到了 dual, 这里再来补充一点相关的知识。这套理论不仅适用于 SVM 的优化问题, 而是对于所有带约束的优化问题都适用的, 是优化理论中的一个重要部分。简单来说, 对于任意一个带约束的优化都可以写成这样的形式:

$$\begin{aligned} \min f_0(x) \\ s.t. f_i(x) \leq 0, \quad i = 1, \dots, m \\ h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

形式统一能够简化推导过程中不必要的复杂性。其他的形式都可以归约到这样的标准形式, 例如一个  $\max f(x)$  可以转化为  $\min -f(x)$  等。假如  $f_0, f_1, \dots, f_m$  全都是凸函数, 并且  $h_1, \dots, h_p$  全都是仿射函数 (就是形如  $Ax + b$  的形式), 那么这个问题就叫做凸优化 (Convex Optimization) 问题。凸优化问题有许多优良的性质, 例如它的极值是唯一的。不过, 这里我们并没有假定需要处理的优化问题是一个凸优化问题。

虽然约束条件能够帮助我们减小搜索空间, 但是如果约束条件本身就是比较复杂的形式的话, 其实是一件很让人头痛的问题, 为此我们希望把带约束的优化问题转化为无约束的优化问题。为此, 我们定义 Lagrangian 如下:

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

它通过一些系数把约束条件和目标函数结合在了一起。当然 Lagrangian 本身并不好玩, 现在让我们来让他针对  $\lambda$  和  $\nu$  最大化, 令:

$$z(x) = \max_{\lambda \geq 0, \nu} \mathcal{L}(x, \lambda, \nu)$$

这里  $\lambda \geq 0$  理解为向量  $\lambda$  的每一个元素都非负即可。这个函数  $z(x)$  对于满足原始问题约束条件的那些  $x$  来说, 其值等于  $f_0(x)$ , 这很容易验证, 因为满足约束条件的  $x$  会使得  $h_i(x) = 0$ , 因此最后一项消掉了, 而  $f_i(x) \leq 0$ , 并且我们要求了  $\lambda \geq 0$ , 因此  $\lambda_i f_i(x) \leq 0$ , 所以最大值只能在它们都取零的时候得到, 这个时候就只剩下  $f_0(x)$  了。因此, 对于满足约束条件的那些  $x$  来说,  $f_0(x) = z(x)$ 。这样一来, 原始的带约束的优化问题其实等价于如下的无约束优化问题:

$$\min_x z(x)$$

因为如果原始问题有最优值, 那么肯定是在满足约束条件的某个  $x^*$  取得, 而对于所有满足约束条件的  $x$ ,  $z(x)$  和  $f_0(x)$  都是相等的。至于那些不满足约束条件的  $x$ , 原始问题是无法取到的, 否则极值问题无解。很容易验证对于这些不满足约束条件的  $x$  有  $z(x) = \infty$ , 这也和原始问题是一致的, 因为求最小值得到无穷大可以和“无解”看作是相容的。

到这里, 我们成功把带约束问题转化为了无约束问题, 不过这其实只是一个形式上的重写, 并没有什么本质上的改变。我们只是把原来的问题通过 Lagrangian 写作了如下形式:

$$\min_x \max_{\lambda \geq 0, \nu} \mathcal{L}(x, \lambda, \nu)$$

这个问题 (或者说原始的带约束的形式) 称作 primal problem。如果你看过之前关于 SVM 的推导, 那么肯定就知道了, 相对应的还有一个 dual problem, 其形式非常类似, 只是把 min 和 max 交换了一下:

$$\max_{\lambda \geq 0, \nu} \min_x \mathcal{L}(x, \lambda, \nu)$$

交换之后的 dual problem 和原来的 primal problem 并不相等, 直观地, 我们可以这样来理解: 胖子中最瘦的那个都比瘦骨精中最胖的那个要胖。当然这是很不严格的说法, 而且扣字眼的话可以纠缠不休, 所以我们还是来看严格数学描述。和刚才的  $z(x)$  类似, 我们也用一个记号来表示内层的这个函数, 记:

$$g(\lambda, \nu) = \min_x \mathcal{L}(x, \lambda, \nu)$$

并称  $g(\lambda, \nu)$  为 Lagrange dual function (不要和  $\mathcal{L}$  的 Lagrangian 混淆了)。 $g$  有一个很好的性质就是它是 primal problem 的一个下界。换句话说, 如果 primal problem 的最小值记为  $p^*$ , 那么对于所有的  $\lambda \geq 0$  和  $\nu$ , 我们有:

$$g(\lambda, \nu) \leq p^*$$

因为对于极值点 (实际上包括所有满足约束条件的点)  $x^*$ , 注意到  $\lambda \geq 0$ , 我们总是有

$$\sum_{i=1}^m \lambda_i f_i(x^*) + \sum_{i=1}^p \nu_i h_i(x^*) \leq 0$$

因此

$$\mathcal{L}(x^*, \lambda, \nu) = f_0(x^*) + \sum_{i=1}^m \lambda_i f_i(x^*) + \sum_{i=1}^p \nu_i h_i(x^*) \leq f_0(x^*)$$

于是

$$g(\lambda, \nu) = \min_x \mathcal{L}(x, \lambda, \nu) \leq \mathcal{L}(x^*, \lambda, \nu) \leq f_0(x^*) = p^*$$

这样一来就确定了  $g$  的下界性质, 于是

$$\max_{\lambda \geq 0, \nu} g(\lambda, \nu)$$

实际上就是最大的下界。这是很自然的, 因为得到下界之后, 我们自然地就希望得到最好的下界, 也就是最大的那一个——因为它离我们要逼近的值最近呀。记 dual problem 的最优值为  $d^*$  的话, 根据上面的推导, 我们就得到了如下性质:

$$d^* \leq p^*$$

这个性质叫做 weak duality，对于所有的优化问题都成立。其中  $p^* - d^*$  被称作 duality gap。需要注意的是，无论 primal problem 是什么形式，dual problem 总是一个 convex optimization 的问题——它的极值是唯一的（如果存在的话），并且有现成的软件包可以对凸优化问题进行求解（虽然求解 general 的 convex optimization 实际上是很慢并且只能求解规模较小的问题的）。这样一来，对于那些难以求解的 primal problem（比如，甚至可以是 NP 问题），我们可以通过找出它的 dual problem，通过优化这个 dual problem 来得到原始问题的一个下界估计。或者说我们甚至都不去优化这个 dual problem，而是（通过某些方法，例如随机）选取一些  $\lambda \geq 0$  和  $\nu$ ，带到  $g(\lambda, \nu)$  中，这样也会得到一些下界（只不过不一定是最大的那个下界而已）。当然选  $\lambda$  和  $\nu$  也并不是总是“随机选”那么容易，根据具体问题，有时候选出来的  $\lambda$  和  $\nu$  带入  $g$  会得到  $-\infty$ ，这虽然是一个完全合法的下界，然而却并没有给我们带来任何有用的信息。

故事到这里还没有结束，既然有 weak duality，显然就会有 strong duality。所谓 strong duality，就是

$$d^* = p^*$$

这是一个很好的性质，strong duality 成立的情况下，我们可以通过求解 dual problem 来优化 primal problem，在 SVM 中我们就是这样做的。当然并不是所有的问题都能满足 strong duality，在讲 SVM 的时候我们直接假定了 strong duality 的成立，这里我们就来提一下 strong duality 成立的条件。不过，这个问题如果要讲清楚，估计写一本书都不够，应该也有不少专门做优化方面的人在研究这相关的问题吧，我没有兴趣（当然也没有精力和能力）来做一个完整的介绍，相信大家也没有兴趣来看这样的东西——否则你肯定是专门研究优化方面的问题的了，此时你肯定比我懂得更多，也就不需要看我写的介绍啦。😄

所以，这里我们就简要地介绍一下 Slater 条件和 KKT 条件。Slater 条件是指存在严格满足约束条件的点  $x$ ，这里的“严格”是指  $f_i(x) < 0$  中的“小于或等于号”要严格取到“小于号”，亦即，存在  $x$  满足

$$\begin{aligned} f_i(x) &< 0 & i = 1, \dots, m \\ h_i(x) &= 0 & i = 1, \dots, p \end{aligned}$$

我们有：如果原始问题是 Convex 的并且满足 Slater 条件的话，那么 strong duality 成立。需要注意的是，这里只是指出了 strong duality 成立的一种情况，而并不是唯一情况。例如，对于某些非 convex optimization 的问题，strong duality 也成立。这里我们不妨回顾一下 SVM 的 primal problem，那是一个 convex optimization 问题（QP 是凸优化问题的一种特殊情况），而 Slater 条件实际上在这里就等价于是存在这样的一个超平面将数据分隔开来，亦即是“数据是可分的”。当数据不可分是，strong duality 不能成立，不过，这个时候我们寻找分隔平面这个问题本身也就是没有意义的了，至于我们如何通过把数据映射到特征空间中来解决不可分的问题，这个当时已经介绍过了，这里就不多说了。

让我们回到 duality 的话题。来看看 strong duality 成立的时候的一些性质。假设  $x^*$  和  $(\lambda^*, \nu^*)$  分别是 primal problem 和 dual problem 的极值点，相应的极值为  $p^*$  和  $d^*$ ，首先  $p^* = d^*$ ，此时我们可以得到

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \min_x (f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x)) \\ &\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

由于两头是相等的，所以这一系列的式子里的不等号全部都可以换成等号。根据第一个不等号我们可以得到  $x^*$  是  $L(x, \lambda^*, \nu^*)$  的一个极值点，由此可以知道  $L(x, \lambda^*, \nu^*)$  在  $x^*$  处的梯度应该等于 0，亦即：

$$\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) = 0$$

此外，由第二个不等式，又显然  $\lambda_i^* f_i(x^*)$  都是非正的，因此我们可以得到

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \dots, m$$

这个条件叫做 complementary slackness。显然，如果  $\lambda_i^* > 0$ ，那么必定有  $f_i(x^*) = 0$ ；反过来，如果  $f_i(x^*) < 0$  那么可以得到  $\lambda_i^* = 0$ 。这个条件正是我们在介绍支持向量的文章末尾时用来证明那些非支持向量（对应于  $f_i(x^*) < 0$ ）所对应的系数  $\alpha_i$ （在本文里对应  $\lambda_i$ ）是为零的。:) 再将其其他一些显而易见的条件写到一起，就是传说中的 KKT (Karush-Kuhn-Tucker) 条件：

$$\begin{aligned} f_i(x^*) &\leq 0, & i = 1, \dots, m \\ h_i(x^*) &= 0, & i = 1, \dots, p \\ \lambda_i^* &\geq 0, & i = 1, \dots, m \\ \lambda_i^* f_i(x^*) &= 0, & i = 1, \dots, m \\ \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) &= 0 \end{aligned}$$

任何满足 strong duality（不一定要要求是通过 Slater 条件得到，也不一定要求是凸优化问题）的问题都满足 KKT 条件，换句话说，这是 strong duality 的一个必要条件。不过，当原始问题是凸优化问题的时候（当然还要求一函数是可微的，否则 KKT 条件的最后一个式子就没有意义了），KKT 就可以升级为充要条件。换句话说，如果 primal problem 是一个凸优化问题，且存在  $\tilde{x}$  和  $(\tilde{\lambda}, \tilde{\nu})$  满足 KKT 条件，那么它们分别是 primal problem 和 dual problem 的极值点并且 strong duality 成立。其证明也比较简单，首先 primal problem 是凸优化问题的话， $g(\lambda, \nu) = \min_x L(x, \lambda, \nu)$  的求解对每一组固定的  $(\lambda, \nu)$  来说也是一个凸优化问题，由 KKT 条件的最后一个式子，知道  $\tilde{x}$  是  $\min_x L(x, \tilde{\lambda}, \tilde{\nu})$  的极值点（如果不是凸优化问题，则不一定能推出来），亦即：

$$\begin{aligned} g(\tilde{\lambda}, \tilde{\nu}) &= \min_x L(x, \tilde{\lambda}, \tilde{\nu}) \\ &= L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\ &= f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i^* f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i^* h_i(\tilde{x}) \\ &= f_0(\tilde{x}) \end{aligned}$$

最后一个式子是根据 KKT 条件的第二和第四个条件得到。由于  $g$  是  $f_0$  的下界，这样一来，就证明了 duality gap 为零，也就是说，strong duality 成立。到此为止，做一下总结。我们简要地介绍了 duality 的概念，基本上没有给什么具体的例子。不过由于内容比较多，为了避免文章超长，就挑了一些重点讲了一下。总的来说，一个

优化问题, 通过求出它的 dual problem, 在只有 weak duality 成立的情况下, 我们至少可以得到原始问题的一个下界。而如果 strong duality 成立, 则可以直接求解 dual problem 来解决原始问题, 就如同经典的 SVM 的求解过程一样。有可能 dual problem 比 primal problem 更容易求解, 或者 dual problem 有一些优良的结构 (例如 SVM 中通过 dual problem 我们可以将问题表示成数据的内积形式从而使得 kernel trick 的应用成为可能)。此外, 还有一些情况会同时求解 dual 和 primal problem, 比如在迭代求解的过程中, 通过判断 duality gap 的大小, 可以得出一个有效的迭代停止条件。

Tags: [Optimization](#), [Support Vector Machine](#)

## 69 comments to 支持向量机: Duality



颢风

[November 3rd, 2010 at 9:48 am](#) · [Reply](#)



继续学习啊~

这一篇讲的太好了!

[Tweets that mention 支持向量机: Duality « Free Mind -- Topsy.com](#)

[November 3rd, 2010 at 10:31 am](#) · [Reply](#)

[...] This post was mentioned on Twitter by yongsun, 识别男. 识别男 said: 关于优化问题中“对偶理论”最浅显易懂的讲解~ <http://zi.mu/10t8> 还是来自浙大小牛人 [...]



邹炳蔚

[November 19th, 2010 at 11:51 am](#) · [Reply](#)

您好, 我是电子工业出版社下属博文视点的编辑邹炳蔚。

我们出版社现在这里有一本刘未鹏老师的书准备出版, 刘未鹏老师您已经很熟悉了不用我太多介绍, 刘老师是从心理学、认知科学、神经科学、经济学和数学等方面综合来分析和指导我们应该如何去更好的学习和进步。

我在看刘老师的博客时候发现了您在刘老师很关注, 而且又是很有心很理性又对这方面有一定研究的人。因为这本书现在在筹备阶段, 所以想请一些像您这样的读书多又比较有品位的读者来给咱们多做交流和反馈, 让我们了解读者想要的是一本什么样的书。

同时如果您能从他的文字里有更多的收获, 那我想我们的交流就是有意义的。

我的QQ: 4587549, E-mail: [zoubingwei@gmail.com](mailto:zoubingwei@gmail.com), 期待您的回复, 问好!



Cerror

[November 25th, 2010 at 8:13 pm](#) · [Reply](#)

关于Duality这个问题, 我个人觉得讲的最好的还是Arkadi Nemirovski的, 讲的真是行云流水啊。引入的很棒, 后面的扩展更是我见过最清晰的。



pluskid

[November 25th, 2010 at 9:50 pm](#) · [Reply](#)

光哥, 求链接.....



cerror

[November 25th, 2010 at 11:08 pm](#) · [Reply](#)

A. Ben-Tal, A. Nemirovski, Lectures on Modern Convex Optimization, MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2001  
万能的gigapedia上应该有的



成都SEO

[December 9th, 2010 at 1:27 pm](#) · [Reply](#)

文章很有意思 谢谢分享



tmzhishang

[January 7th, 2011 at 11:20 am](#) · [Reply](#)

最近刚开始看SVM和SVR, 找到这个系列, 太有帮助了! 真的写得非常好, 入门的人看再合适不过了。不知道剩下的几个番外啥时候能出来呢, 热烈期待呀!

[支持向量机系列\(转载\) « Masonzms's Blog](#)

[April 10th, 2011 at 9:02 am](#) · [Reply](#)

[...] 支持向量机: Duality —— 关于 dual 问题推导的一些补充理论。 [...]



Liuxian

[April 22nd, 2011 at 4:10 am](#) · [Reply](#)

写得很好, 请问下, svm的二次规划问题 满足Slater条件, 是不是可以说存在一个超平面, 使数据可分, 并且全部数据都远离超平面, 没有支持向量? 而最优的解却不是这个超平面? 此时lambda\_i全部为0.

pluskid

[April 22nd, 2011 at 9:41 am](#) · [Reply](#)



你好, 不会没有支持向量啊。支持向量所在的超平面并不是 separating hyperplane 的, 而是和 separating hyperplane 平行的两个超平面。



Liuxian

[April 22nd, 2011 at 5:07 pm · Reply](#)

感谢回复, 我的问题好像没表达清楚, Slater条件不是说存在 $w$ 使得 $f_i(w) > 0$ , 是不是可以这么说存在一个超平面, 使数据可分, 但没有支持向量。



pluskid

[April 23rd, 2011 at 12:19 pm · Reply](#)

你好, Slater 条件满足就表示数据是可分的, 和支持向量没有关系。实际上, 如果数据可分的话, 支持向量是显然存在的啊。



Liuxian

[April 23rd, 2011 at 9:59 pm · Reply](#)

那麻烦再问下, slater条件是不是就是在svm中, 存在 $w$ 使得所有 $y_i(w \cdot x_i + b) > 1$ 呢?那数据可分, 不就满足 $y_i(w \cdot x_i + b) > 0$ 就可以了么?



pluskid

[April 24th, 2011 at 11:08 am](#)

$y_i(w \cdot x_i + b) > 0, i = 1, \dots, n$ , 在有限个点的情况下是不是可以取到最小值? 即存在  $m > 0$  使得  $y_i(w \cdot x_i + b) \geq m > 0, i = 1, \dots, n$ 。这个时候再相应地缩放  $w$  和  $b$  (比如都除以  $m/2$ ) 不就可以满足  $y_i(w \cdot x_i + b) > 1, i = 1, \dots, n$  了啊?

#### [机器学习中的算法\(2\)-支持向量机\(SVM\)基础 » NoName](#)

[May 2nd, 2011 at 11:56 pm · Reply](#)

[...] 求解这个式子的过程需要拉格朗日对偶性的相关知识 (另外pluskid也有一篇文章专门讲这个问题), 并且有一定的公式推导, 如果不感兴趣, 可以直接跳到后面用蓝色公式表示的结论, 该部分推导主要参考自plukids的文章。 [...]

#### [LeftNotEasy写的理解SVM的博文 | 丕子](#)

[May 3rd, 2011 at 1:35 pm · Reply](#)

[...] 求解这个式子的过程需要拉格朗日对偶性的相关知识 (另外pluskid也有一篇文章专门讲这个问题), 并且有一定的公式推导, 如果不感兴趣, 可以直接跳到后面用蓝色公式表示的结论, 该部分推导主要参考自plukids的文章。 [...]

#### [Machine learning algorithm \(2\) – the support vector machine \(SVM\) basis](#)

[May 4th, 2011 at 8:34 pm · Reply](#)

[...] management then needs Lagrange duality theorem Related knowledge (in addition pluskid also have an article dedicated to this problem), and a certain formula deduction, if not interested, can directly jump to [...]



Leo

[May 21st, 2011 at 12:07 am · Reply](#)

师兄, 暂且这么叫吧! 这个系列写的好棒, 文字平实易懂, 其中每个公式都仔细编辑过, 看得出你真的是很用心写的!

这已经是第二次从google上随机找资料找到这里, 第一次是language model!

希望有空了继续写些Machine Learning内容!

拜谢!



Hanhan

[November 25th, 2011 at 11:00 pm · Reply](#)

这是最优化里面的东西, 计算机专业的也要搞吗?



pluskid

[November 26th, 2011 at 7:44 pm · Reply](#)

是的, 很重要的东西。。。

#### [Thought this was cool: 机器学习中的算法\(2\)-支持向量机\(SVM\)基础 « CWYAlpha](#)

[December 18th, 2011 at 12:11 am · Reply](#)

[...] 求解这个式子的过程需要拉格朗日对偶性的相关知识 (另外pluskid也有一篇文章专门讲这个问题), 并且有一定的公式推导, 如果不感兴趣, 可以直接跳到后面用蓝色公式表示的结论, 该部分推导主要参考自plukids的文章。 [...]

wjcp

[December 20th, 2011 at 11:57 pm · Reply](#)



应该将 $x$ 定义在可行域中, 否则, 问题描述比较含糊。  
比如 $z(x) = \max_{\{\lambda > 0\}} L(\lambda, \mu, x)$   
如果 $x$ 不在可行域的话,  $z(x)$ 未必是无穷吧?



pluskid  
[December 21st, 2011 at 1:17 pm · Reply](#)

你说的可行域是什么可行域? 如果是说满足  $f(x)$ 、 $h(x)$  的限制的那些  $x$  的话, 必须要在那些限制之外才等于无穷, 之内是等于  $f_0(x)$  的。



iveney  
[February 23rd, 2012 at 1:01 pm · Reply](#)

再次拜讀大作。dual problem 應該總是 concave 才對。原文寫成了 convex。



pluskid  
[February 23rd, 2012 at 1:34 pm · Reply](#)

你好, 只是 dual function 是 concave 的而已, 在这里给定的约束下去 maximize 一个 concave 的函数, 这就是一个 convex 问题没有错。只要在前面加一个负号就很好理解啦。



iveney  
[March 7th, 2012 at 12:14 pm · Reply](#)

oops, 我看錯了。你是對的, dual function 是 concave, dual problem 是 convex



yanglei  
[May 31st, 2012 at 9:04 am · Reply](#)

楼主问下, 就单纯的看  $\max g(\lambda, v)$  这个式子, 当 $x$ 满足条件的时候, 不是应该  $\lambda=0$  的时候 才能得到最大值吗? 为什么还能 通过某些方法, 例如随机 来确定  $\lambda$  和  $v$  的值呢?



pluskid  
[May 31st, 2012 at 9:14 am · Reply](#)

你好, 不知道你说的随机确定  $\lambda$  和  $v$  这个是什么意思.....



yanglei  
[May 31st, 2012 at 11:41 am · Reply](#)

这个是你文章中说的话“而是（通过某些方法, 例如随机）选取一些  $\lambda \geq 0$  和  $v$ , 带到  $g(\lambda, v)$  中.....”我是想问 要想得到  $\max g(\lambda, v)$  的解 不是应该让  $\lambda=0$  么? 谢谢回复~!



pluskid  
[May 31st, 2012 at 12:08 pm · Reply](#)

你好, 随机取的值只是一个下界而已, 不一定是最优解。



yanglei  
[May 31st, 2012 at 3:10 pm · Reply](#)

随机取的是一个下界。 $\max g(\lambda, v)$  的最优解不是应该让  $\lambda=0$  么? 因为不管 $x$ 的值是多少  $f(x) \leq 0$   $h(x) = 0$ , 只有  $\lambda=0$ ,  $g(\lambda, v)$  才能取到最大值, 也就是  $\max g(\lambda, v)$  的最优解。其实我知道我的理解是有错误的, 但是不知道为什么错了, 请指点一下啊!!!



pluskid  
[May 31st, 2012 at 3:36 pm · Reply](#)

并不是对任意  $x$  都有  $f(x) \leq 0$  和  $h(x) = 0$  的, 否则的话我们也不用这么费力地去处理约束条件了。



yanglei  
[May 31st, 2012 at 5:14 pm · Reply](#)

如果 $x$ 是在定义域, 而不是可行域求出了一个 $\lambda$ 和 $v$ 的值, 那就算我得到了一个的 $\lambda$ 和 $v$ 的值, 比如我看很多文献里面用次梯度迭代法 能够求出一个 $\lambda$ 和 $v$ 的值。那么我怎么能保证 用我求得的 $\lambda$ 和 $v$ 的值 最后解出来的 $x$ 是满足约束条件的呢?



pluskid  
[May 31st, 2012 at 7:06 pm · Reply](#)

如果你说的是求解  $\min_x \max_{\{\lambda, \mu\}} L$  的话, 这个是 primal problem, 而不是 dual problem, 他的最优解和原问题的等价性在文章里已经说了的。要么就是你对 min max 和 max min 的顺序没弄对?



[kinslover](#)[November 1st, 2012 at 10:55 am](#) · Reply

想问下, 这里面 $g(\lambda, v)$ 可不可以理解为,  $L(x, \lambda, v)$ 对于每一对 $(\lambda, v)$ 所能取到的最小值么?

谢谢!

[wangjian](#)[January 24th, 2013 at 4:13 pm](#) · Reply

无论 primal problem 是什么形式, dual problem 总是一个 convex optimization 的问题

请问这个出处在哪啊 斯坦福机器学习的关于凸函数的补充讲义中, 他的对偶函数就是concave 的, 这和你的这个结论矛盾了 哪个对呢? ?

[pluskid](#)[January 24th, 2013 at 10:53 pm](#) · Reply

对偶函数是 concave 的, 所以是个 convex optimization problem, 哪里有矛盾吗?

[chen](#)[July 15th, 2013 at 11:28 am](#) · Reply

请问下楼主, 在证strong duality成立时性质中, 你说根据第一个不等式可以得到 $x^*$  是  $L(x, \lambda^*, v^*)$  的一个极值点, 这句话怎么理解? 我觉得 $x^*$ 是满足约束的最优解, 而 $L(x, \lambda^*, v^*)$  的最优解是在全局里面取, 貌似不一样吧. 烦劳楼主扫下盲了

[pluskid](#)[July 19th, 2013 at 4:27 am](#) · Reply

因为 dual problem 总是 convex 的.

[chen](#)[July 19th, 2013 at 9:08 am](#) · Reply

楼主真的好热心, 貌似有问必答, 但是我还是没懂.....  $\min L(x, \lambda^*, v^*)$  是一个数,  $L(x^*, \lambda^*, v^*)$  也是一个数,  $\min L(x, \lambda^*, v^*) < L(x^*, \lambda^*, v^*)$  这个不等式怎么就能说明  $x^*$ 是  $\min L(x, \lambda^*, v^*)$  的一个极值呢? 我只能理解 $L(x, \lambda^*, v^*) < L(x^*, \lambda^*, v^*)$  才说明 $x^*$ 是L的极值

[pluskid](#)[July 19th, 2013 at 9:13 am](#) · Reply

哦, 这个是因为我们这里的情况下所有的不等号都取到等号了.

[chen](#)[July 19th, 2013 at 9:36 am](#) · Reply

对哦, 中间全是等号. 哎, 楼主已经写的很详细了, 都怪自己没仔细看, 瞎琢磨, 一本最优化的书没看完, , 又换另一本, 都没找到好理解的. 还是楼主的博客详细

[celery](#)[August 10th, 2013 at 4:56 pm](#) · Reply

$g$ 小于 $p^*$ , 这个不是叫做 $g$ 有上界么, 怎么成了 $g$ 的下界了?

[Michael](#)[September 15th, 2013 at 12:45 pm](#) · Reply

写的非常好, 非常清楚. 偶然发现了这个博客, 会继续支持的.

[zhangjinchao](#)[October 30th, 2013 at 2:00 pm](#) · Reply

当所有的样本都是支持向量的时候, 原始问题是不是就满足Slater条件了? 因为这个时候, 对于不等式约束条件, 小于等于号都会变成等号. 在这种情况下, 不满足Slater条件, SVM算法还怎么往下推导? 很疑惑, 盼楼主快速解答.

[zhangjinchao](#)[October 30th, 2013 at 2:02 pm](#) · Reply

当所有的样本都是支持向量的时候, 原始问题是不是就不满足Slater条件了? 因为这个时候, 对于不等式约束条件, 小于等于号都会变成等号, 不算是严格满足约束条件了. 在这种情况下, 不满足Slater条件, SVM算法还怎么往下推导? 很疑惑, 盼楼主快速解答.

[pluskid](#)[November 2nd, 2013 at 10:11 am](#) · Reply

首先 Slater 是充分但并不一定是必要条件. 其次最优解的时候所有不等号都成为等号了并不代表没有其他可行解使得不等号严格成立啊.



zhangjinchao

[November 3rd, 2013 at 2:24 pm](#) · [Reply](#)那么, 如果所有的样本都是支持向量的话,  $f_i(X)=0$ , 还满足 Slater 条件么?

pluskid

[November 3rd, 2013 at 10:34 pm](#)可能 notation 有点 confusing, 在 SVM 的情况下  $X$  指的是 weight  $w$ , 而不是样本。

zhangjinchao

[November 4th, 2013 at 9:34 am](#) · [Reply](#)

恩, 明白你的意思。如果所有的样本都是支持向量, 那么对于最优解  $w^*$  来说,  $f(w^*)=0$  是成立的。但是可能存在  $w$  是非最优解的时候, 有  $f(w)<0$  成立。我从别的地方看到过弱化的 Slater 条件, 当约束条件  $f(w)$  是线性不等式时, 可以不用遵守严格的约束, 而是放宽到了小于等于。



pluskid

[November 4th, 2013 at 1:29 pm](#) · [Reply](#)

是的。



yyx

[November 15th, 2013 at 5:06 pm](#) · [Reply](#)

请问对于凸优化问题, Slater 条件是强对偶的必要条件吗?



pluskid

[November 15th, 2013 at 11:08 pm](#) · [Reply](#)

Think about where the feasible domain has only one point.



daniel

[March 7th, 2014 at 1:48 pm](#) · [Reply](#)

博主, 你好, 看了你的文章很有收获, 但有些问题想请教下你, 谢谢!

在证明  $g$  是 primal problem 的一个下界。其中, 有  $g(\lambda, v) = \min_x L(x, \lambda, v) \leq L(x^*, \lambda, v) \leq f_0(x^*) = p^*$ , 第一个小于等于是否应该改成等于号更合理, 因为你已经假设了  $x^*$  是所有满足约束条件的极值点。



pluskid

[March 8th, 2014 at 1:17 am](#) · [Reply](#)

是的。



f2m4

[May 7th, 2014 at 5:01 pm](#) · [Reply](#)

博主你好, 看了你的文章很有帮助, 不过这上面你对于  $d^* \leq p^*$  的证明, 感觉有点不妥, 对于  $L(x, \lambda, v)$  来说, 只需如此证明即可,  $d^* = \max_{\lambda, v} (\min_x L(x, \lambda, v)) \leq \max_{\lambda, v} (\min_x (\max_{\lambda, v} L(x, \lambda, v))) = \max_{\lambda, v} L(x^*, \lambda, v) = p^*$ ; 事实上, 对于任意的多元函数均是成立的, 这样证明我感觉会更加好些。



pluskid

[May 9th, 2014 at 5:08 am](#) · [Reply](#)你好, 你这个证明不对吧? 不等式那里怎么会里层和外层同时都有  $\max_{\lambda, v}$  ?

f2m4

[May 9th, 2014 at 4:21 pm](#) · [Reply](#)

对的啊。可以这样理解:

对于任意的  $x, \lambda, v$  来说, 有  $L(x, \lambda, v) \leq \max_{\lambda, v} L(x, \lambda, v)$ ,  
 为了方便起见, 我们设  $\max_{\lambda, v} L(x, \lambda, v) = S$   
 我们就有, 对于任意  $x, \lambda, v$  都有  $L(x, \lambda, v) \leq S$ ,  
 因为是对任意  $x, \lambda, v$  都成立的, 显然有  
 $\max_{\lambda, v} (\min_x L(x, \lambda, v)) \leq \max_{\lambda, v} (\min_x S)$   
 而此不等式的右边, 我们将  $S$  再代回去, 那就有  
 右边  $= \max_{\lambda, v} (\min_x (\max_{\lambda, v} L(x, \lambda, v))) = \max_{\lambda, v} L(x^*, \lambda, v) = p^*$   
 而  $p^*$  是一个数, 当然其关于  $\lambda, v$  的最大值仍然是  $p^*$   
 因此就有右边  $= p^*$ , 所以就  $d^* = p^*$  啦。



f2m4

[May 9th, 2014 at 4:23 pm](#)最后写错了...就有 $d^* \leq p^*$ 啦

f2m4

[May 9th, 2014 at 4:57 pm](#) · Reply

而楼主你对 $d^* \leq p^*$ 证明的问题在于: 我们无法通过 $x^*$ 是 $L(x, \lambda, u)$ 的极值点以及 $\lambda \succeq 0$ 得出 $\sum_{i=1}^m \lambda_i (x^*) + \sum_{i=1}^p u_i (x^*) \leq 0$

因为 $x^*$ 是不一定满足约束条件的。

还有就是博主按这种方法证明的话可能会让人觉得那个不等式是依赖于问题才成立的, 而事实上那是恒成立的, 可能某些情况下会有一些些误导。以上是我对这篇文章的一些看法啊, 当然我是要非常感谢博主的啊, 博主很多文章都对我是有很大帮助的。



pluskid

[May 9th, 2014 at 11:47 pm](#) · Reply

$x^*$  是极值点肯定是满足约束的啊。只是说这种证法在最优解无法达到的时候不成立了。不过你说得没错, 这个不等式确实是不依赖于具体的问题的, 也不论最优解存不存在。



f2m4

[May 10th, 2014 at 1:40 am](#) · Reply

不好意思啊,  $x^*$ 是 $L(x, \lambda, v)$ 的极值还是不明白怎么能够满足约束条件:我是否可以举个例子反驳: 就令 $f_0 = x^2$ ,  $f_1 = x^2 - 1 \leq 0$ ,  $h_1 = x + 1 = 0$ ; 那么 $L = (\lambda + 1)x^2 + v_1x + v_0 - \lambda$  那么 $x^* = -v_1/(2\lambda + 2)$  显然是不一定满足约束条件的, 不知道这个例子可不可以, 有问题不?



pluskid

[May 13th, 2014 at 10:21 pm](#)

你好,  $x^*$  是 primal 问题的极值点, 你这个例子里  $x^* = -1$  啊。

### [SVM算法 – 剑客 | 关注科技互联网](#)

[April 18th, 2015 at 5:48 pm](#) · Reply

[...] 求解这个式子的过程需要拉格朗日对偶性的相关知识 (另外pluskid也有一篇文章专门讲这个问题), 并且有一定的公式推导, 如果不感兴趣, 可以直接跳到后面用蓝色公式表示的结论, 该部分推导主要参考自plukids的文章。 [...]

### [SVM算法 | 一世浮华一场空](#)

[April 19th, 2015 at 8:15 am](#) · Reply

[...] 求解这个式子的过程需要拉格朗日对偶性的相关知识 (另外pluskid也有一篇文章专门讲这个问题), 并且有一定的公式推导, 如果不感兴趣, 可以直接跳到后面用蓝色公式表示的结论, 该部分推导主要参考自plukids的文章。 [...]

### [SVM算法 | 36大数据](#)

[April 20th, 2015 at 2:00 pm](#) · Reply

[...] 求解这个式子的过程需要拉格朗日对偶性的相关知识 (另外pluskid也有一篇文章专门讲这个问题), 并且有一定的公式推导, 如果不感兴趣, 可以直接跳到后面用蓝色公式表示的结论, 该部分推导主要参考自plukids的文章。 [...]

### [SVM算法 » 大数据 » 09大数据](#)

[July 5th, 2015 at 11:26 am](#) · Reply

[...] 求解这个式子的过程需要拉格朗日对偶性的相关知识 (另外pluskid也有一篇文章专门讲这个问题), 并且有一定的公式推导, 如果不感兴趣, 可以直接跳到后面用蓝色公式表示的结论, 该部分推导主要参考自plukids的文章。 [...]



cheng

[January 18th, 2018 at 4:29 pm](#) · Reply

请教一个问题, Slater 条件为什么成立啊, Slater 条件的必要条件是所有的不等式约束要满足严格成立, 而不管 $w$ 取什么值, 在支持向量上的点的不等式约束并不严格成立, 而是取了等号啊。谁能帮我解答一下吗?