机器学习数学：拉格朗日对偶问题

岭大王
啊，好热！

对偶问题是利用拉格朗日对偶性将原始问题转换为对偶问题，通过解对偶问题得到原始问题的解。优点是：

- 对偶问题往往更易于求解
- 自然引入核函数，推广到非线性分类问题的求解

原始问题

我们以硬间隔svm的原始问题为例，求解一个约束最优化问题：

$$\begin{aligned} &\min_{w,b} \quad \frac{1}{2}|w|^2 \\ s.t. \quad & y_i(wx_i+b)-1\geq 0 \\ & i=1,2..N \end{aligned}$$

（注：这个问题用二次规划求解时复杂度与x的维度有关）

我们将这个问题一般化：

$$\begin{aligned} &\min_{x\in R^n} \quad f(x) \\ s.t. \quad & c_i(x)\leq 0 \quad i=1,2..k \\ & h_j(x)=0 \quad j=1,2..l \end{aligned}$$

极小极大问题

引入拉格朗日函数，由于约束条件有k+l个，所以我们有：

$$L(x,\alpha,\beta)=f(x)+\sum_{i=1}^{k}\alpha_i c_i(x)+\sum_{j=1}^{l}\beta_j h_j(x)$$

(求对x有限制条件的f(x)的最优化问题转为对 x,α,β 没有限制条件的 $L(x,\alpha,\beta)$ 极值问题)

其中 α_i,β_j 是拉格朗日乘子， $\alpha_i\geq 0$ ，我们再定义一个函数：

$$\theta_p(x)=\max_{\alpha,\beta,\alpha_i\geq 0} L(x,\alpha,\beta)$$

（注：在 α， β， $\alpha_i\geq 0$ 条件下的最大值）

我们来研究这个函数：

$$\theta_p(x)=\max_{\alpha,\beta,\alpha_i\geq 0}(f(x)+\sum_{i=1}^{k}\alpha_i c_i(x)+\sum_{j=1}^{l}\beta_j h_j(x))$$

为无穷大，因为可以取 $\alpha_i = \infty, \beta_j h_j(x) = \infty$

- 相反，假设x满足原始问题的约束条件，则上式最大值一定是 $\alpha = 0$ ，同时h(x)=0，所以上式的最大值的条件不存在了，即最大值就是一个定值f(x)
- 所以在x满足原始问题约束的条件下， $\theta_p(x) = f(x)$

所以，我们可以把原始问题中的s.t.条件去掉，得到原始问题的等价问题：
$\min_x f(x) = \min_x \theta_p(x) = \min_x \max_{\alpha,\beta,\alpha\geq 0} L(x,\alpha,\beta)\,\mathrm{r}$

如果没看懂，没关系，现在来证明这个等式，即：

$$\min_{x\in R^n} f(x) \quad s.t. \quad c_i(x) \leq 0(i=1,2..k), h_j(x) = 0(j=1,2..l)$$

$$\Updownarrow$$

$$\min_x \max_{\alpha,\beta,\alpha_i\geq 0} (f(x) + \sum_{i=1}^{k} \alpha_i c_i(x) + \sum_{j=1}^{l} \beta_j h_j(x)) = \min_x \theta_p(x)$$

- $\max_{\alpha,\beta,\alpha_i\geq 0}(\dots.)$ ：在x是常数的前提下，在取值范围是max下标的条件下的，某个 α， β 使得后面括号中取最大值。
- $\min_x \max_{\alpha,\beta,\alpha_i\geq 0}(\dots.)$ ：在上一步对x取值全部遍历之后，得到的所有的最大值中的一个最小值。
- 将 $\max_{\alpha,\beta,\alpha_i\geq 0} (f(x) + \sum_{i=1}^{k} \alpha_i c_i(x) + \sum_{j=1}^{l} \beta_j h_j(x))$ 记做 $P(x)$
- 如果x满足 $c_i(x) > 0\text{或者} h_j(x) \neq 0$ ，则 $P(x)$ 值无限大，则第一步就无解
- 如果x满足 $c_i(x) \leq 0\text{或者} h_j(x) = 0$ ，又有max下标的条件限制，则一定是 $\sum_{i=1}^{k} \alpha_i c_i(x) = 0$ ， $\sum_{j=1}^{l} \beta_j h_j(x) = 0$ ，则 有：
  $\max_{\alpha,\beta,\alpha_i\geq 0} (f(x) + \sum_{i=1}^{k} \alpha_i c_i(x) + \sum_{j=1}^{l} \beta_j h_j(x)) = \max_{\alpha,\beta,\alpha_i\geq 0}(f(x)) = f(x)$
- 上式两边加 $\min$ ，即证明得等价

综上，我们得到原始问题等价于 $\min_x \max_{\alpha,\beta,\alpha\geq 0} L(x,\alpha,\beta)$

对偶问题

（我可能是好多年没碰高数了，上面一节要啰啰嗦嗦这么多。。）

我们直接给出以下定理：

若原始问题和对偶问题都有最优值，则对偶问题最优值d* $\leq$ 原始问题最优值p*：

$$d^* = \max_{\alpha,\beta:\alpha_i\geq 0} \min_x L(x,\alpha,\beta) \leq \min_x \max_{\alpha,\beta:\alpha_i\geq 0} L(x,\alpha,\beta) = p^*$$

（通俗点就是宁做凤尾不做鸡头，凤尾（右）是原始问题，鸡头（左）是对偶问题）

定理1易证明，参考统计学习方法或技法视频，此处证明略。

即我们如果把对偶问题解决了，那就能得到原始问题的下限。

对于对偶问题来说，我们求解最小化部分时没有任何限制条件，而没有限制条件的最小化问题的解一定是在求得x的偏导数=0这处，那我们就能得到一些等式，将这些等式代入拉格朗日函数中就可以简化计算（svm中可以最终把min消去，且得到一个只与 α 有关的最大化问题）

KKT条件

推论1：设x⋆和α⋆,β⋆分别是原始问题和对偶问题的可行解，并且d⋆=p⋆，则x⋆和α⋆,β⋆分别是原始问题和对偶问题的最优解。

【即：如果某个解使得最优值相等，则这个解同时是原始和对偶问题的最优解】

定理2：考虑原始问题和对偶问题，假设函数f(x)和ci(x)是凸函数，hj(x)是仿射函数；并且假设不等式约束ci(x)是严格可行的，即存在x，对所有i有c_i(x)<0，则存在x⋆和α⋆,β⋆，使x⋆是原始问题的解，α⋆,β⋆是对偶问题的解，并且有p⋆=d⋆=L(x⋆,α⋆,β⋆).

【即：满足某些条件下，最优解存在且使得最优值相等】

定理3： x^*,α^*,β^* 分别是原始问题和对偶问题的解的充分必要条件是， x^*,α^*,β^* 满足下面的KKT条件：

$$\nabla_x L(x^*,\alpha^*,\beta^*)=0\,\mathrm{diag}\,\nabla_\alpha L(x^*,\alpha^*,\beta^*)=0\,\mathrm{diag}\,\nabla_\beta L(x^*,\alpha^*,\beta^*)=0 \quad (1)\,\mathrm{diag}\,\alpha_i^* c_i(x^*)=0, i=1,2,\cdots,k \quad (2)\,\mathrm{diag}\,c_i(x^*)\le 0, i=1,2,\cdots,k \quad (3)\,\mathrm{diag}\,\alpha_i^*\ge 0, i=1,2,\cdots,k \quad (4)\,\mathrm{diag}\,h_j(x^*)=0, j=1,2,\cdots,l \quad (5)$$

上式条件中1和2是最重要的，3、4、5都是将原始问题表示成拉格朗日函数时的条件。2称为对偶互补条件，根据此条件，当 $\alpha_i^*>0$则$c_i(x^*)=0$ 。

【即：在满足KKT条件下，原始问题的求解可以转换成对偶问题来求解，强对偶成立（等号成立），而强对偶成立，又一定满足kkt条件】

两天了，基本理解并写完。

既然都看到这儿了，少年点个赞可好？感谢！

done 2017年11月23日16:51:58

编辑于 2018-02-09

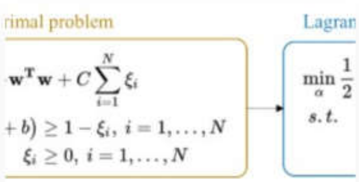
机器学习

文章被以下专栏收录

机器学习笔记
ML/AI/DM的经验、心得、笔记

进入专栏

推荐阅读



支持向量机原理详解(二): 拉格朗日对偶函数，SVM的对偶...

RookieJ

【分类战车SVM】第四话：拉格朗日对偶问题（原来这么简...

第一话：开题话第二话：线性分类第三话：最大间隔分类器第四话：拉格朗日对偶问题（原来这么简单！）第五话：核函数（哦，这太神奇了！）第六话：SMO算法（像Smoke一样简单！）附录：用...

Ansel... 发表于数说工作室

【机器学习】支持向量机SVM（非常详细）

阿泽

18 条评论

切换为时间排序

写下你的评论...

秋念 2017-12-21

仔仔细细的看了两遍，终于搞清楚了原始问题和对偶问题以及转换条件，真的写得非常良心了，赞~~PS：前面原始问题的说明一点也不啰嗦，表示很详细

5

岭大王 (作者) 回复 秋念 2017-12-21

欣慰啊。。。

1

草莓味的工科男 2018-03-07

写的挺好的，感谢!

赞

行者 2018-04-21

讲的挺棒的，刚看完《统计学习方法》，正在疑惑为什么要引入对偶问题来解决原始问题，搜到这个回答，醍醐灌顶，赞一个!

赞

卫星 2018-10-07

写的很棒！思维很缜密啊。楼主非常厉害了

赞

小白Licko 2019-01-05

master面试ing...

“现在能口头推导出SVM的拉格朗日对偶问题，就给你申请助学奖学金”
我为什么没有早点看到这个回答?

赞

知乎用户 2019-02-22

写的好

赞

前来感谢🙏

赞

冰箱 2019-07-18

谢谢！

赞

流線形 2019-07-24

写的好 清晰易懂

赞

陈夏 2019-09-16

写得很清晰，棒棒棒。

赞

Machinist 2019-10-09

看了这个遍没咋看懂，一会在看一遍

赞

Machinist 2019-10-10

转化为一般问题为啥是那个形式

赞

葱油鸭的脑子 2019-10-20

请教一下：最上面求解条件极值时候用max下的α，β，α≥0 的原因是这样可以将问题条件消除从而得到原始问题吗？因此求解原始问题等价于求解max条件下的min？

此外就是对偶问题中为什么最优解≤原问题最优解？意思是对偶问题的最优解不一定是元问题的最优解，有可能是错误的或者什么的吗？

终于有一篇看进去了！谢谢(*°∀°)=3

赞

取数表演艺术家 回复 葱油鸭的脑子 03-17

有假设x不满足原始约束时，P(x)为无穷，当满足原始约束时，P(x)为f(x)。所以原始带有约束问题的最优解，可以转化成无约束问题的最优解。

的确对偶问题的最优解不一定是原问题的最优解。只有强对偶的时候才可以，也就是说满足KKT条件时，等号才成立。

赞

judy 2019-11-04

写的很好，感谢！

赞

东东枪 01-12

感谢

赞

关尔润之 06-08

alpha和alpha_i 的区别是什么

赞

▲ 赞同 221 ▼ 18 条评论 分享 喜欢 收藏 ···