线性规划中的对偶问题

原创 任勇 LOA算法学习笔记 2021-01-08 14:29

01 前言

线性规划的对偶问题在很多领域中都会用到，举个例子，SVM（支持向量机）中就用对偶问题将Soft-Margin和Hard-Margin转换成了相同的形式，并使得其中的样本以内积形式呈现，以便能够使用核函数进行替换。

而我陆续上了运筹学、算法中的最优化，仍然对对偶问题理解不到位，直到上了卜东波老师的算法课。老师很神奇地将内点法（障碍函数法）、罚函数法、和拉格朗日乘子法进行了统一，看作是理想情况下的一种近似。然后从拉格朗日对偶的角度去理解对偶问题。这与最优化或者运筹学教科书上的将对偶问题求解分四种情况，让我们进行死记硬背的方式是完全不同的。（建议想学最优化一定要来听卜老师的算法课）

02 对偶问题举例

生产I、II两种产品，要占用A、B、C设备时间，每件产品机时利润如表所示:

	产品 I	产品 II	每天可用时间
占用 A 机时	0	5	15
占用 B 机时	6	2	24
占用 C 机时	1	1	5
利润	2	1	

如何生产使每天利润最大?转化为线性规划问题:

$$\begin{gathered}\max \quad 2x_1+x_2 \\ s.t. \quad 5x_2 \leq 15 \\ 6x_1+2x_2 \leq 24 \\ x_1+x_2 \leq 5 \\ x1 \geq 0, x_2 \geq 0\end{gathered}$$

现在，该生产厂对外承包。候选的承包商，经过调研得知如下信息:

- 该厂现有三种设备A、B、C，对应的每日可用时 间分别是15小时、24小时和5小时;
- 该厂宣布对外承包前，利用这三种设备生产两种 产品I、II;
- 产品I、II投放市场后的利润分别不低于2、1。

那么，候选承包商应该如何投标才最划算?后续承包商获得的信息如下:

	A	B	C	市场最低利润
产品 I	0	6	1	2
产品 II	5	2	1	1
运行时间	15	24	5	

设备A、B、C的单位承租(投标)价格为 y_1, y_2, y_3

目标: $min \quad 15y_1+24y_2+5y_3$

即求解如下线性规划问题:

$$\begin{aligned} \max \quad & 15y_1 + 24y_2 + 5y_3 \\ s.t. \quad & 6y_2 + y_3 \geq 2 \\ & 5y_1 + 2y_2 + y_3 \geq 1 \\ & y1 \geq 0, y_2 \geq 0, y_3 \geq 0 \end{aligned}$$

该问题的任意一个可行解对应的目标函数值都不小于原问题的目标函数值,而该问题的最优目标函数值等于原问题的最优目标函数值。

可以看作是成产厂与承包商之间的一种对抗。

03 问题的一般情况

3.1 无约束转化为有约束

考虑一般的有约束优化问题：

$$\begin{aligned} \max \quad & f_0 x \\ s.t. \quad & f_i(x) \leq 0 \\ & h_i(x) = 0 \end{aligned}$$

将约束条件转移到目标函数中使得约束优化问题转化为无约束优化问题。

$$min \quad f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) + \sum_{i=1}^{k} I_0(h_i(x))$$

其中，

$$I_-(u) = \begin{cases} 0, & u \leq 0 \\ \infty & u > 0 \end{cases}$$

$$I_0(u) = \begin{cases} 0, & u = 0 \\ \infty & u \neq 0 \end{cases}$$

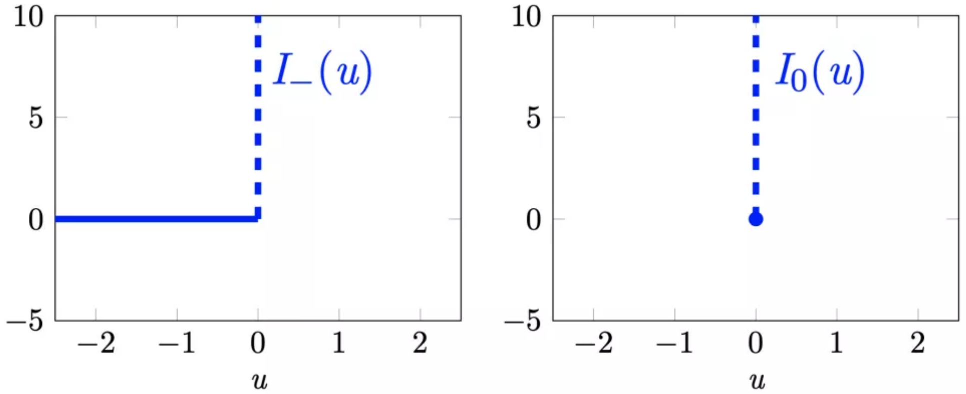


图1 来自卜老师课件

此处强迫 $f_i(x) \leq 0$ ，并且 $g_i(x) = 0$ ，否则目标函数值会非常大。但这样的函数不可求导数，因此找函数去近似这两个函数。

1) 对数障碍函数（内点法）

$$\hat{I}_-(u) = -\frac{1}{t} log(-u), \quad (t > 0)$$

2) 罚函数

$$\hat{I}_-(u) = \begin{cases} u^t & u \geq 0 \quad (t > 1) \\ 0 & otherwise \end{cases}$$

3) 线性函数（拉格朗日）

$$\hat{I}_-(u) = -\lambda\mu \quad (\lambda \leq 0)$$
$$\hat{I}_0(u) = -\nu\mu$$

此处提供了一个下界。

3.2 拉格朗日对偶

用线性函数 $-\lambda\mu \quad (\lambda \leq 0)$ 和 $-\nu\mu$ 去替代，便得到了拉格朗日函数：

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

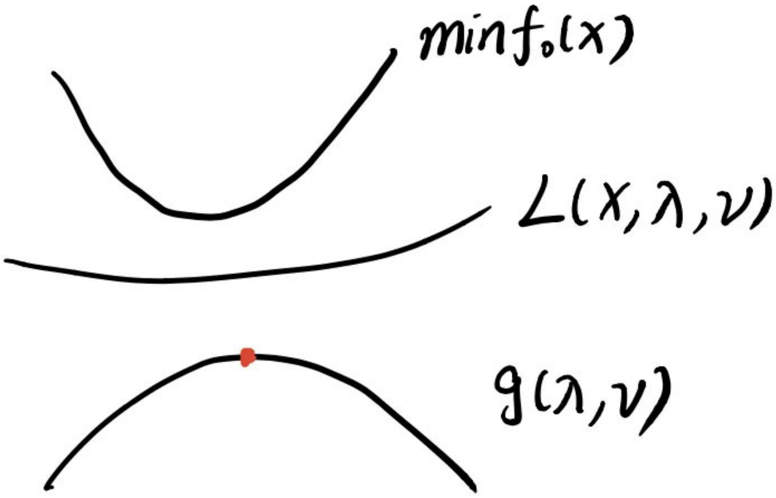
该无约束优化问题等价于原来的约束优化问题。

对每个约束引入一个拉格朗日乘子。若满足原来的约束，此时 $\lambda_i \leq 0, f_i(x) \leq 0, h_i(x) \leq 0$，因此 $L(x, \lambda, \nu) \leq f_0(x)$

考虑 $L(x, \lambda, \nu)$ 关于x的下确界函数 $g(\lambda, \nu)$ （称为拉格朗日对偶函数，为关于拉格朗日函数沟通了变量分别为x和 λ, ν 的函数），

$$f_0(x) \geq L(x, \lambda, \nu) \geq \inf_{x \in D} L(x, \lambda, \nu) = g(\lambda, \nu)$$

拉格朗日函数沟通了变量分别为x和 λ, ν 的两个函数。



找 $g(\lambda, \nu)$ 的最大值，提供了 $f_0(x)$ 最紧的一个下界。既求解: $\max g(\lambda, \nu)$， $g(\lambda, \nu)$ 始终是一个上凸的函数。

3.3 线性规划的拉格朗日对偶

考虑标准线性规划问题，

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & Ax \leq b \\ & x \geq 0 \end{aligned}$$

拉格朗日函数：

$$L(x,\lambda,\nu)=c^Tx-\sum_{i=1}^{m}\lambda_i(a_{i1}x1+\cdots+a_{in}x_n+b_i)-\sum_{i=1}^{n}\nu_ix_i$$

$$c^Tx\geq L(x,\lambda,\nu)\geq \inf_{x\in D}L(x,\lambda,\nu)=g(\lambda,\nu)$$

$$g(\lambda,\nu)=\inf_{x\in D}(c^Tx-\sum_{i=1}^{m}\lambda_i(a_{i1}x1+\cdots+a_{in}x_n+b_i)-\sum_{i=1}^{n}\nu_ix_i)$$

$$=\inf_{x\in D}((\lambda^Tb+(c^T-\lambda^TA-\nu)x)$$

$$=\begin{cases}\lambda^Tb, & if\ c^T=\lambda A+\nu^T\\ -\infty & otherwise\end{cases}$$

找 $f_0(x)$ 最紧的一个下界，即 $max\ g(\lambda,\nu)$ 也就是:

$$\begin{gathered}\max \quad g(\lambda,\nu)\\ s.t. \quad \lambda\leq 0\\ \nu\geq 0\end{gathered}$$

等价于

$$\begin{gathered}\max \quad \lambda^Tb\\ s.t. \quad \lambda\leq 0\\ \nu\geq 0\\ c^T=\lambda A+\nu^T\end{gathered}$$

合并最后两个约束，得：

$$\begin{gathered}\max \quad \lambda^Tb\\ s.t. \quad \lambda^TA\leq c^T\\ \lambda\leq 0\end{gathered}$$

得到了对偶问题的形式,拉格朗日乘子就是对偶变量。

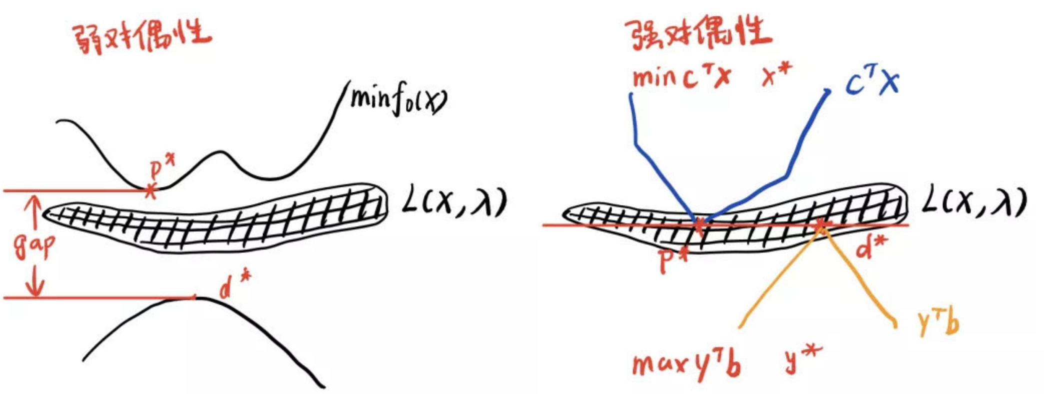
拉格朗日乘子 λ 的几种解释：

1) 对偶变量y

2) 影子价格（边际成本）：约束加强或放松单位大小，目标函数发生多少变化。相当于拉格朗日函数对b求偏导。

3) $\lambda=\frac{\nabla f_0(x)}{\nabla h(x)}$ (目标函数的梯度/约束的梯度)

3.4 弱对偶与强对偶



原问题和对偶问题之间存在下述相互关系：

- **弱对偶性**:原对偶问题任何可行解的目标值都是另一问题最优目标值的界。(推论:原对偶问题目标值相等的一对可行解是各自的最优解)
- **强对偶性**:原对偶问题只要有一个有最优解，另一个就有最优解，并且最优目标值相等。

对线性规划，强对偶性成立。一般来说，大部分凸问题强对偶成立；对于部分非凸问题也成立。

3.5 最优解求解

原问题与对偶问题的互补松弛性：

设 $\hat{X}$ 和 $\hat{Y}$ 分别是原问题和对偶问题的可行解，则它们分别是各自问题最优解的充要条件是满足互补松弛性条件：

$$\hat{Y}(b-AX)=0\ and\ \hat{X}(A^Ty-C)=0$$

即如果原问题某个不等式是松的(不等于0)，则其相应的对偶变量（拉格朗日乘子）必须是紧的(等于0)，反之亦然。(由上面互补松弛条件得 $b^T\hat{Y}=\hat{Y}^TAX=C^TX$ ，根据弱对偶性的推论可知两者分别是各自问题的最优解)

同时考虑原问题和对偶问题，x*和y*是最优解，则满足：

1) 原问题可行

2) 对偶问题可行

3) 互补松弛性

求解过程可以有三种方法：

- 单纯形法：构造初始化满足（1）和（3），保持满足（1）和（3）的情况下改进x,y使得更满足（2）。
- 对偶单纯形法：构造初始化满足（2）和（3），保持满足（2）和（3）的情况下改进x,y使得更满足（1）。
- 内点法：构造初始化满足（1）和（2），保持满足（1）和（2）的情况下改进x,y使得更满足（3）。

04 对偶的应用-SVM中的对偶

Soft-Margin\ SVM的原问题形式如下：

$$\min \quad \frac{1}{2}||w||^2+C\sum_{i=1}^n\xi_i$$

$$s.t. \quad 1-\xi_i-y_i(w^Tx_i+b)\le 0 \quad \forall i=1\dots n$$

$$-\xi_i\le 0 \quad \forall i=1\dots n$$

拉格朗日函数

$$L(w,b,\xi,\alpha,\beta)=\frac{1}{2}||w||^2+C\sum_{i=1}^n\xi_i+\sum_{i=1}^n\alpha_i(1-\xi_i-y_i(w^Tx_i+b))-\sum_{i=1}^n\beta_i\xi_i$$

$$=\frac{1}{2}||w||^2+C\sum_{i=1}^n\xi_i+\sum_{i=1}^n\alpha_i-\sum_{i=1}^n\alpha_i\xi_i-\sum_{i=1}^n\alpha_iy_iw^Tx_i-b\sum_{i=1}^n\alpha_iy_i-\sum_{i=1}^n\beta_i\xi_i$$

$$\alpha_i\ge 0,\beta_i\ge 0$$

对原问题的变量w,b, ξ_i 求下确界即极小值，相当于对它们求偏导等于零。

$$\frac{\partial L(w,b,\xi,\alpha,\beta)}{\partial w}=w-\sum_{i=1}^{n}\alpha_i y_i x_i=0$$

$$\frac{\partial L(w,b,\xi,\alpha,\beta)}{\partial b}=\sum_{i=1}^{n}\alpha_i y_i=0$$

$$\frac{\partial L(w,b,\xi,\alpha,\beta)}{\partial \xi_i}=C-\alpha_i-\beta_i=0$$

将这三个式子带入拉格朗日函数，得拉格朗日对偶函数：

$$g(\alpha,\beta)=\inf_{w,b,\xi\in D}L(w,b,\xi,\alpha,\beta)$$

$$=\frac{1}{2}||\sum_{i=1}^{n}\alpha_i y_i x_i||^2+\sum_{i=1}^{n}\alpha_i\xi_i+\sum_{i=1}^{n}\beta_i\xi_i+\sum_{i=1}^{n}\alpha_i-\sum_{i=1}^{n}\alpha_i\xi_i$$

$$-\sum_{i=1}^{n}\alpha_i y_i(\sum_{j=1}^{n}\alpha_j y_j x_j)^T x_i-b\sum_{i=1}^{n}\alpha_i y_i-\sum_{i=1}^{n}\beta_i\xi_i$$

$$=\frac{1}{2}||\sum_{i=1}^{n}\alpha_i y_i x_i||^2-\sum_{i=1}^{n}\alpha_i y_i(\sum_{j=1}^{n}\alpha_j y_j x_j)^T x_i+\sum_{i=1}^{n}\alpha_i$$

$$=-\sum_{i=1,j=i}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j+\sum_{i=1}^{n}\alpha_i$$

由第三个式子 $C-\alpha_i-\beta_i=0$ 得 $C\geq\alpha_i\geq 0$
因此，最终的对偶函数整理为：

$$\max_{\alpha}\quad g(\alpha,\beta)=\sum_{i=1}^{n}\alpha_i-\sum_{i=1,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$s.t\quad C\geq\alpha_i\geq 0\quad ,i=1,\ldots,n$$

$$\sum_{i=1}^{n}\alpha_i y_i=0$$

经过观察，可以发现SVM对偶问题形式中，样本以 $x_i^T x_j$ 这种内积形式出现，为使用核函数 $\Phi(x)$ 提供了基础。之后只需要将内积替换为核函数即可，大大简化了计算。

05 总结

本文总结了卜老师课堂讲解的关于对偶问题的理解，从拉格朗日对偶的角度可以对对偶问题有更深的理解，而不再需要去背公式。最后以推导SVM的对偶形式为例展示了把一个问题转化为其对偶问题的作用。文中引用到了卜老师课件中的部分内容。如果需要对线性规划有更为深刻的理解，强烈建议来听卜老师关于线性规划部分的课程。