浅谈拉格朗日乘数法及对偶问题在SVM中的应用(一)

原创 马旭淼 LOA算法学习笔记 2021-12-11 23:51

最近在算法课堂上学习了拉格朗日乘数法及其对偶问题,因此我在课后结合课堂所学知识和相关文献资料对拉格朗日乘数法及其相关问题进行了归纳总结。

> 这个系列会分为三部分，第一部分讨论Lagrange乘数法，第二部分讨论Lagrange对偶问题，第三部分讨论Lagrange对偶问题在硬间隔支持向量机(SVM)推导过程中的应用。本文为第一部分的内容。

PART 1 Lagrange乘数法

1.1 简介

对于一般的优化问题，主要可以分为三类：①无约束问题；②等式约束问题；③不等式约束问题。

对于无约束问题(e.g. $minf(x)$)，通常只需要令函数的导函数等于0，即可取得局部或者全局最优解。如果是凸优化问题，则可以保证有全局最优解。对于后二者来说，直接求导数的方法就不再适用，需要其他的方法来辅助求解。

拉格朗日乘数法（Lagrange Multiplier Method），是一种当多元函数中自变量受到一个或多个条件的约束时，求其极值的方法。在优化领域中，拉格朗日乘数法是十分常用的一种方法，同时也是SVM（支持向量机）的理论基础。这种方法适用于求解带等式的约束问题；而在求解带不等式约束的问题时，则需要引入KKT条件。

1.2 等式约束问题

考虑如下等式约束优化问题：

$$\begin{aligned} &min \quad f(x_1, x_2, \ldots, x_m) \\ &s.t. \quad h_i(x) = 0, i = 1, 2 \ldots, n \end{aligned} \tag{1}$$

可以构造如下拉格朗日函数：

$$L(x_1, x_2, \ldots, x_m, \mu_1, \ldots, \mu_n) = f(x_1, x_2, \ldots, x_m) - \sum_{i=1}^{n} \mu_i \cdot h_i(x)$$

如果我们要求上面这个拉格朗日函数的极值，可以通过求其偏导数的方法。即：

$$\nabla_{x_1, \ldots, x_m, \mu_1, \ldots, \mu_n} L(x_1, \ldots, x_m, \mu_1, \ldots, \mu_n) = 0$$

这样就可以构造出$n+m$个方程，用于求解$n+m$个未知数。如式(2)所示：

$$\begin{aligned} \frac{\partial L}{\partial x_j} &= \frac{\partial f}{\partial x_j} - \sum_{i=1}^{n} \mu_i \frac{\partial h_i}{\partial x_j} \\ \frac{\partial L}{\partial \mu_i} &= h_i(x) = 0 \end{aligned} \tag{2}$$

1.3 不等式约束问题

对于不等式约束问题，处理的方法与等式约束大同小异。为简化表达，下面问题的分析将用$\mathbf{x}$来代替$x_1, x_2, \ldots, x_m$。考虑如下问题：

$$\begin{aligned} &min \quad f(\mathbf{x}) \\ &s.t. \quad g_i(\mathbf{x}) \leq 0, i = 1, \ldots, k \end{aligned} \tag{3}$$

$$h_j(\mathbf{x}) = 0, j = 1, \ldots, n$$

对于所有$g_i(\mathbf{x})$，引入松弛变量a_i^2，使得$g_i'(\mathbf{x}) = g_i(\mathbf{x}) + a_i^2 = 0$(引入$a_i^2$而不是$a_i$的原因是松弛变量必须为正数，才能保证原不等式有可能成为等式)。这样，不等式约束问题就变成了等式约束问题，同样可以用上面提到的方法来处理。构造下列拉格朗日函数：

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{a_i^2}) = f(\mathbf{x}) - \sum_{i=1}^{k} \lambda_i \, g_i'(\mathbf{x}) - \sum_{j=1}^{n} \mu_j \, h_j(\mathbf{x}) \tag{4}$$

$$\lambda_i \leq 0, \mu_j \in \mathscr{R}$$

同样地，可以通过对其变量求偏导数求解：

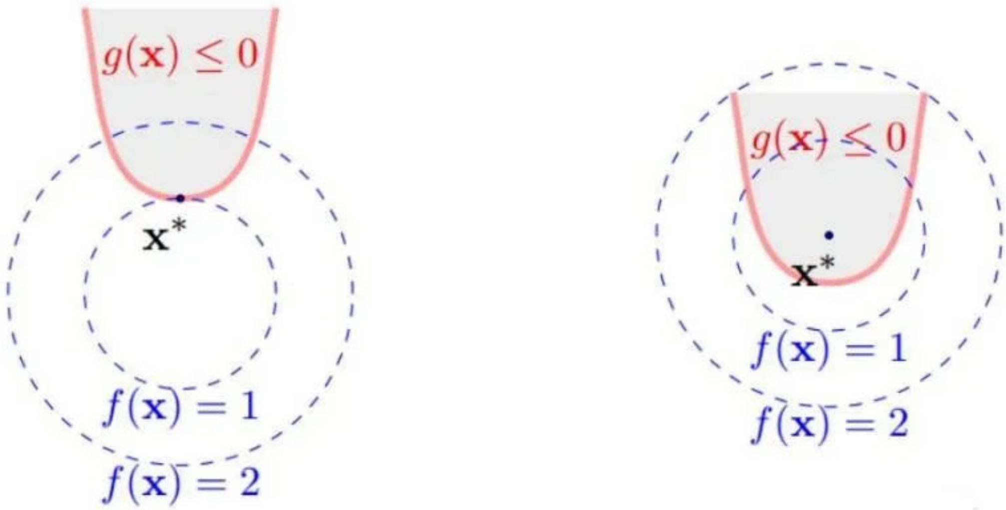
$$\nabla_{\mathbf{x},\lambda,\mu,\mathbf{a}_i^2} L(\mathbf{x}, \lambda, \mu, \mathbf{a}_i^2) = 0$$

若用$\mathbf{x}^*$表示最优解，当x取到最优解时，松弛变量实际上不起作用，则应当有：

$$\nabla_{\mathbf{x}^*,\lambda,\mu} L(\mathbf{x}^*, \lambda, \mu) = 0 \tag{5}$$

1.4 Karush-Kuhn-Tucker条件

对于式(3)所示的问题，考虑不等式约束，可以用下图来描述。



图中可以分为两种情况：

- 如果最优解恰好在边界上，那么问题就相当于只有等式约束，很容易求解。
- 如果最优解在阴影区内部，相当于一定满足约束条件，因此问题就等同于只考虑$f(x)$的最优解，也就是直接求$\nabla f(x) = 0$的解。

综合上述分析以及式(3)(4)(5)，可以归纳总结出Karush-Kuhn-Tucker条件，简称为KKT条件：

1. (不动点条件)$\nabla L(\mathbf{x}^*, \lambda, \mu) = 0$
2. (原可行性条件)$g_i(\mathbf{x}^*) \leq 0, i = 1, \ldots, k; h_j(\mathbf{x}^*) = 0, j = 1, \ldots, n$
3. (对偶可行性条件)$\lambda_i \leq 0, i = 1, \ldots, k$
4. (互补松弛性条件)$\lambda_i^* g_i(\mathbf{x}^*) = 0, i = 1, \ldots, k$

第一个条件也被称为拉格朗日平稳性条件。其意义很直观：当$\mathbf{x}$取到最优解$\mathbf{x}^*$时，其拉格朗日函数在此点处的导数为0，因此该点也被称为不动点。第二个条件为原可行性条件，描述了问题中的所有约束。第三个条件为对偶可行性条件，会在下面的对偶问题中进行解释。最后一个条件为互补松弛性条件，其证明需要用到PART 2介绍的对偶问题。这里先说明结论，互补松弛性条件也可以等价地写成：

$$\begin{aligned} \lambda_i^* < 0 \Rightarrow \quad & g_i(\mathbf{x}^*) = 0 \\ g_i(\mathbf{x}^*) < 0 \Rightarrow \quad & \lambda_i^* = 0 \end{aligned}$$

即对于非零的拉格朗日乘子λ_i，约束条件总是紧的；而约束条件为松则意味着拉格朗日乘子λ_i为零。由此可见，拉格朗日乘子反映了其对应约束条件的松紧，实际上它也是对偶变量，会在PART 2对偶问题的讨论中详细解释。对于一般的任意问题而言，满足KKT条件是使一组解成为最优解的必要条件，也就是说如果一组解为原问题的最优解，则它应当满足KKT条件；当原问题是凸问题的时候，KKT条件也是充分条件。

参考文献

[1]李航. 统计学习方法[M]. 清华大学出版社, 2012.
[2]周志华. 机器学习[M]. 清华大学出版社, 2015.
[3]卜东波.《计算机算法设计与分析》课程讲义[EB/OL].中国科学院大学，2021.