#机器学习教学基础

3.10 拉格朗日算法基础

#3.10拉格朗日乘子法 张满满

##讲义和规范

•https://www.matongxue.com/madocs/939.html

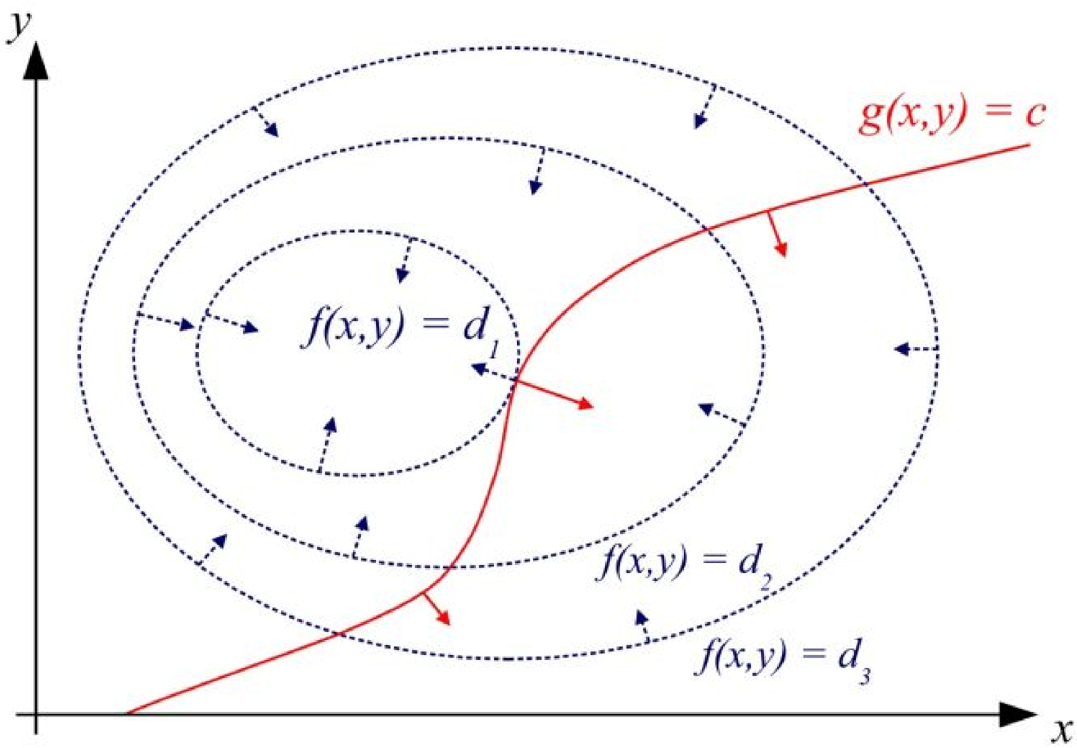
一、拉格朗日乘数法

是通过引入拉格朗日乘子来将含有n个变量和k个约束条件的约束优化问题转化为含有（n+k）个变量的无约束优化问题。

定义：

假设有自变量x和y，给定约束条件g(x,y)=c，要求f(x,y)在约束g下的极值。

我们可以画出f的等高线图，如下图。此时，约束g=c由于只有一个自由度，因此也是图中的一条曲线（红色曲线所示）。显然地，当约束曲线g=c与某一条等高线f=d1相切时，函数f取得极值。
两曲线相切等价于两曲线在切点处拥有共线的法向量。因此可得函数f(x,y)与g(x,y)在切点处的梯度（gradient）成正比。
于是我们便可以列出方程组求解切点的坐标(x,y)，进而得到函数f的极值



二、约束条件的集中形式**

- 等式条件下求最优解

$$
\begin{gathered}
minf(x,y) \\
s.t. \qquad h_i(x,y)=c, i=1,2,3...l
\end{gathered}
$$

- 不等式条件下求最优解

$$
\begin{gathered}
minf(x) \\
s.t. \qquad h_j(x)=0, j=1,2,3...l \\
g_i(x)\leq 0, i=1,2,3...l^{'}
\end{gathered}
$$

三、求解不同约束条件问题的最优方法

（1）等式条件下求最优解：

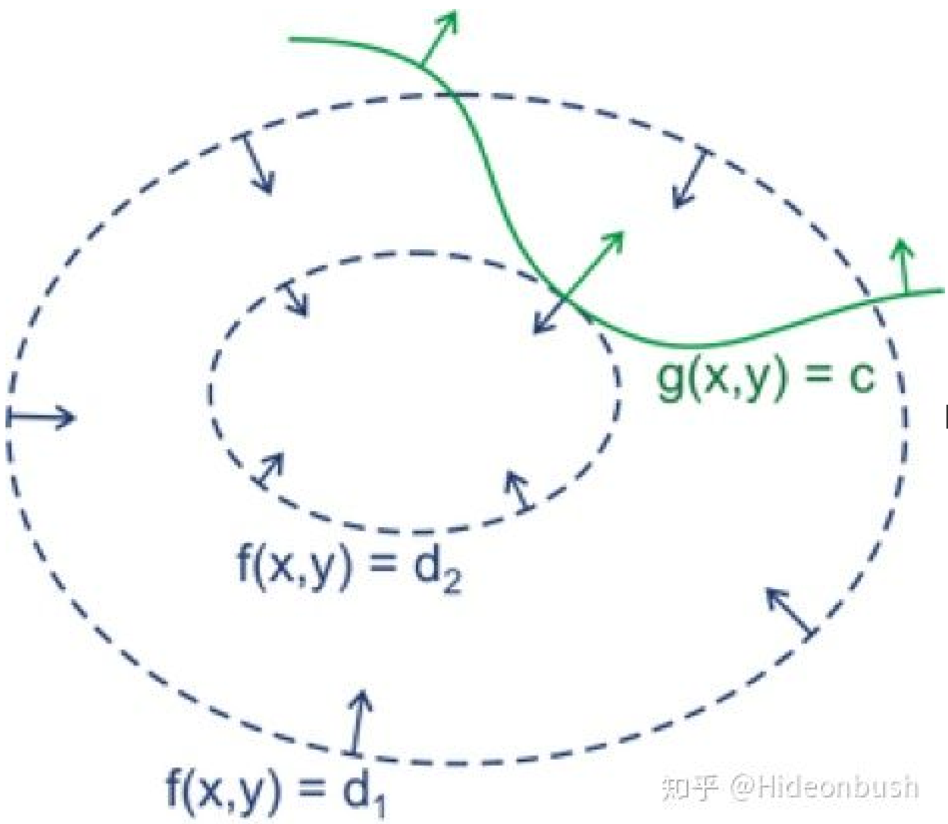
其实，高中的时候已经接触过等式条件机制求最优解的问题，都知道用拉格朗日乘数法可以求得最优解。首先构建一个Lagrange multiplier：

$$L(x,y,\lambda)=f(x,y)+\lambda(h(x)-c)$$

构建出 $L(x,y,\lambda)$ 后，跟原来的问题对比一下，构建出的Lagrange multiplier把原来的等数条件约束情况变成含三个参数 (x,y,λ) 无条件极值的问题，最后可以通过求导数，令极值为零可以求出可行解 x 。

$$\nabla L(x,y,\lambda)=\nabla f(x)+\lambda\nabla(g(x)-c)b$$

（拉格朗日乘子求的解不一定是最优解，其实只是局部最优解，这里称作可行解，只有在凸函数中才能保证最优解）至于为什么可以把原始的等式条件极值问题转化成求 $L(x,y,\lambda)$ 的极值问题？可以观察一下下面的二维图：



图一

蓝色虚线是目标函数 $f(x,y)$ 的等高线，绿色实现 $g(x,y)=c$ 是约束条件。图中，目标函数和条件函数有三种情况：

1、相离

对于相离的情况，我们以前学过，两个函数有交点才说明是两个函数的解，所以相离明显不行。

2、相交

两个函数相交的才是两个函数的解，但是相交得到的一定不是最优值，因为相交意味着肯定还存在其它的等高线在该条等高线的内部或者外部，使得新的等高线与目标函数的交点的值更大或者更小

3、相切

图中，等高线与条件函数相切时候，只有一个交点，是可行解。

当等高线与条件函数在可行解处（相切时候）的梯度平行有下式：

$$\nabla f(x)=-\lambda\nabla(g(x)-c)$$

对上式子移项后，可得到：

$$\nabla L(x,y,\lambda)=\nabla f(x)+\lambda\nabla(g(x)-c)=0$$

该式子恰好与令Lagrange multiplier的导数为零是式子相同。

举一个小栗子：

$$minf(x,y)=x^2+y^2bs.t.\quad xy=3$$

根据上式子可知是一个典型的约束优化问题，约束条件是等式，可以用拉格朗日乘子。

$$L(x,y,\lambda)=x^2+y^2+\lambda(xy-3)$$

原来的条件约束问题，转化为无约束方程组问题。

$$2x + \lambda y = 0$$
$$2y + \lambda x = 0$$
$$xy - 3 = 0$$

求得 $\lambda = \pm 2$ ，当 $\lambda = 2$ 时， $x = \sqrt{3}$ ，$y = \sqrt{3}$ ；当 $\lambda = -2$ ， $x = -\sqrt{3}$ ，$y = -\sqrt{3}$ 。

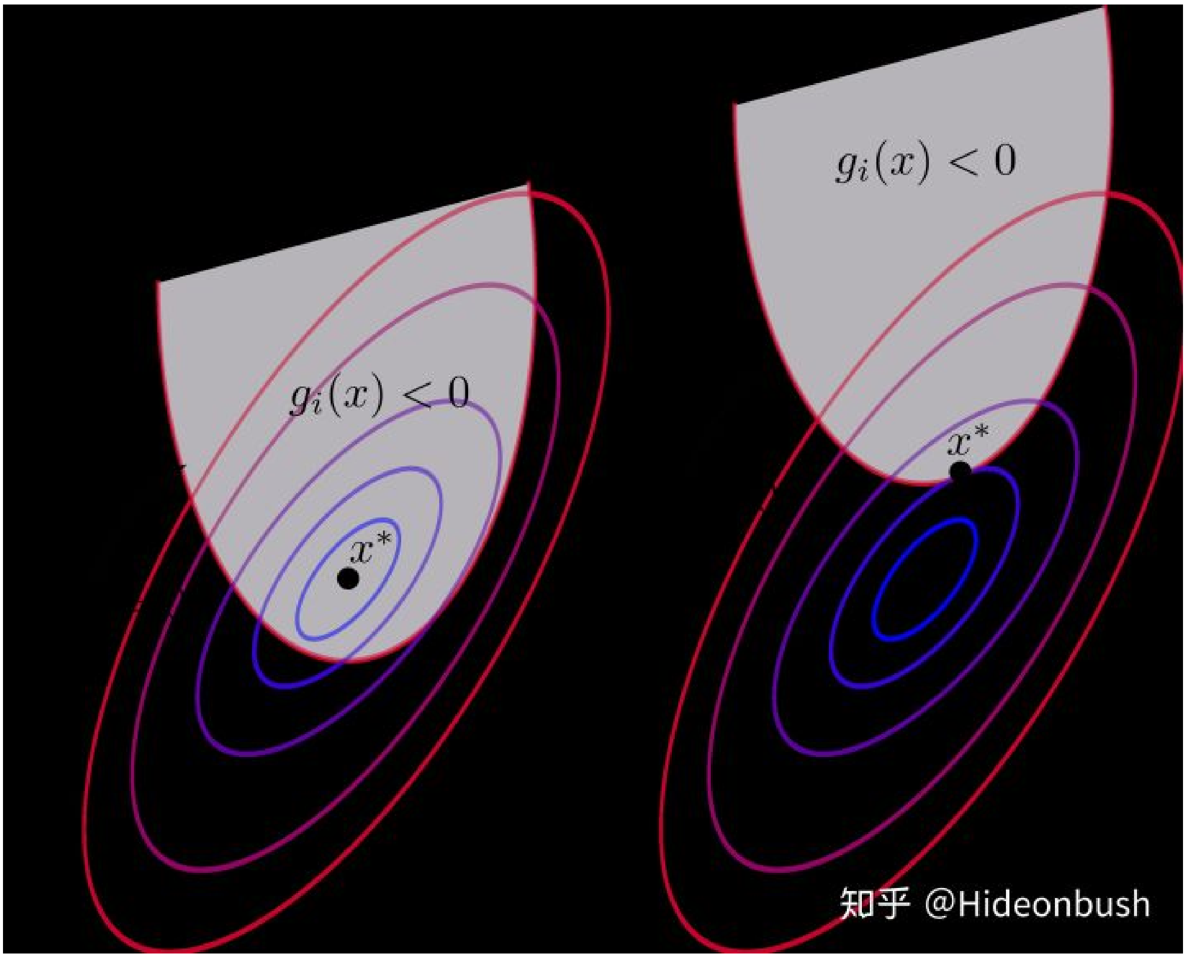
（2）不等式条件下求最优解

上述都是等式条件约束的优化问题，但事实上，很多时候等式约束很难覆盖我们显示的问题，计算成本时候，通常说不能超过多少资金，不能超多多少时间，都是一个范围内，通常面对更多的是不等式条件约束的情况，面对这种情况，可以通过增加KKT条件后，通过拉格朗日乘子求解不等式约束的优化问题。

可以把目标函数和所有的约束条件写成一个式子：

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{l'} \lambda_i g_i(x) + \sum_{j=1}^{l} \mu_j h_j(x)$$

对于不等式条件的情况有两种：即可行解在 $g(x) < 0$ 或者 $g(x) = 0$ 区域内取得（如下图）



图二

- 在 $g(x) < 0$ 内，就是没有限制条件下的最优解，正好满足限制条件，换句话说数约束函数不起作用（如上图左侧）这情况可以直接最小化目标函数，找处可行解 x^* ，所以有：

$$\nabla_{x^*} f(x^*) = 0, \quad \lambda = 0, \quad g(x^*) < 0$$

- 在 $g(x) = 0$ 上，换句话说，就是没有限制条件下的最优解，不能满足限制条件，需要条件起作用，此时 $\lambda \neq 0$，因此可行解在 $g(x) = 0$ 上，就变成等式条件约束的情况，经之前分析，可以证明可行解 x^* 发生在 $\nabla f(x) = -\lambda \nabla g(x)$，即是说，存在一个 λ 使得 $\nabla f(x^*) = -\lambda \nabla g(x^*)$。为确定 λ 范围，回看图一，梯度 $\nabla f(x)$ 方向与梯度 $\nabla g(x)$的方相反且梯度平行，所以，若要使得 $\nabla f(x) = -\lambda \nabla g(x)$ 成立，则 $\lambda > 0$，所以有：

$$g(x^*) = 0, \quad \lambda > 0$$

把两个情况综合考虑得到：

$$\begin{gathered} \lambda > 0 \\ \lambda g(x^*) = 0 \\ g(x^*) \leq 0 \end{gathered}$$

以上就是KKT条件。

考虑不等式条件下求最优解时：

$$\begin{gathered} min f(x) \\ s.t. \quad h_j(x) = 0, j = 1,2,3...l \\ g_i(x) \leq 0, i = 1,2,3...l' \end{gathered}$$

定义拉格朗日函数：

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{l'} \lambda_i g_i(x) + \sum_{j=1}^{l} \mu_j h_j(x)$$

其中 λ_i 是 $g_i(x) \leq 0$ 的格朗日乘子，μ_j 是 $h_j(x) = 0$ 的拉格朗日乘子。

经过之前的分析观察，得知加上不等式约束后可行解 x 需要满足的就是以下的 KKT 条：

$\nabla_x L(x, \lambda, \mu) = 0$ ①

$h_j(x) = 0, \quad j = 1,2,3...l$ ②

$g_i(x) \leq 0, \quad i = 1,2,3...l'$ ③

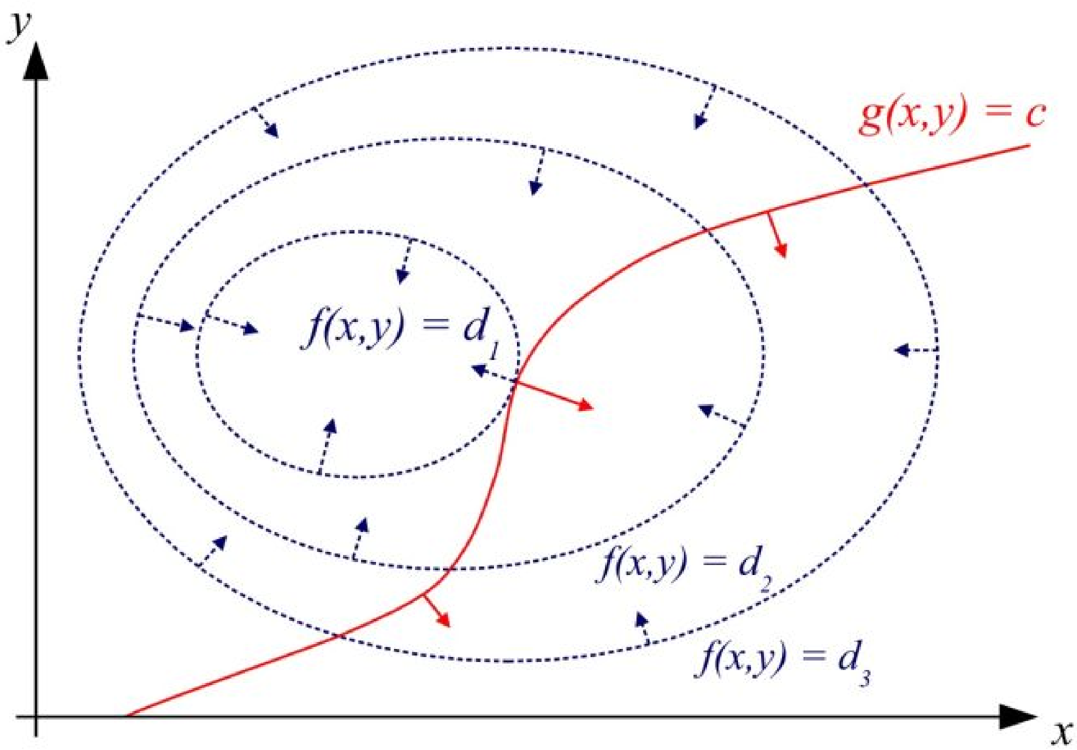
$\lambda_i > 0, \quad i = 1,2,3...l'$ ④

$\lambda_i g_i(x) = 0, \quad i = 1,2,3...l'$ ⑤

定义：

假设有自变量x和y，给定约束条件g(x,y)=c，要求f(x,y)在约束g下的极值。

我们可以画出f的等高线图，如下图。此时，约束g=c由于只有一个自由度，因此也是图中的一条曲线（红色曲线所示）。显然地，当约束曲线g=c与某一条等高线f=d1相切时，函数f取得极值。
两曲线相切等价于两曲线在切点处拥有共线的法向量。因此可得函数f(x,y)与g(x,y)在切点处的梯度（gradient）成正比。
于是我们便可以列出方程组求解切点的坐标(x,y)，进而得到函数f的极值



，使用拉格朗日乘子法解决下面的问题：

very helpful.

2. ☆ Use the Lagrange multiplier method to solve the following optimization problem:

$$\min_{x} 10 - x_{|1}^2 - x_{|2}^2, \text{subject to } x_{|2} \geq x_{|1}^2, x_{|1} + x_{|2} = 0 \qquad (7)$$

方法一：按照题目要求使用拉格朗日乘子法

这个方法就是按照几个限定的约束条件来解方程，KKT条件整体如下：

$$\left.\frac{\partial L}{\partial X}\right|_{X=X^*}=0 \qquad (1)$$

$$\lambda_j \neq 0, \qquad (2)$$

$$u_k \geq 0, \qquad (3)$$

$$\mu_k g_k(X^*)=0 \qquad (4)$$

$$h_j(X^*)=0 \quad j=1,2,\ldots,p \quad (5)$$

$$g_k(X^*)\leq 0 \quad k=1,2,\ldots q \quad (6)$$

题目的拉格朗日乘子式为L=(10−x21−x22)+λ(x1+x2)+μ(x21−x2)L=(10−x12−x22)+λ(x1+x2)+μ(x12−x2)。

由于python只能求解等式方程，不能求解不等式，所以这里的等式方程有（1）（4）（5）,先取出这三个式子，首先对（1）进行求导，可以获得两个两个式子，再结合（4）（5），这样就有四个方程四个未知数，可以得到理论上的解空间，这样就限定了解的范围,该部分代码如下:

```
rom sympy import *

x1 = Symbol("x1")
x2 = Symbol("x2")
a = Symbol("a")
b = Symbol("b")
f = 10 - x1**2 - x2**2 +a*(x1 + x2) + b*(x1**2 - x2)
fx1 = diff(f,x1)
fx2 = diff(f,x2)
result = solve([fx1,fx2,(x1**2-x2)*b,x1+x2],[x1,x2,a,b])
```

求出来的解可能还不是唯一的，怎么办呢，这时候用另外几个上面没有用到的条件也就是KKT里的非等式解，也就是（2）（3）（7）去获得解空间里符合条件的数值：

```
for i in range(len(result)):
    if result[i][3]>=0 and result[i][0]**2-result[i][1]<=0 and result[i][2]!=0:
        print(result[i])
        print("loss:",10 - result[i][0]**2 - result[i][1]**2 +result[i][2]*
(result[i][1]                + result[i][0]) + result[i][3]*(result[i][0]**2 -
result[i][1]))
```

获得最终解如下，(-1, 1, 6, 4)，('loss:', 8)。

##拉格朗日乘子法的应用

1. 约束最优化在经济学占有很重要的地位。例如一个消费者的选择问题可以被视为一个求效用方程在预算约束下的最大值问题。拉格朗日乘数在经济学中被解释为影子价格，设定在某种约束下，在这里即收入的边际效用。

拉格朗日乘数就是效用函数在最优解处对收入的偏导数，也就是在最优解处增加一个单位收入带来的效用增加，或者说在最优解处有效用衡量收入的价值，称之为收入的边际效用。

在企业生产问题中，拉格朗日乘数用来衡量要素投入变动所带来的收入变动，du/dm=λ,u表示效用函数或生产函数，m表示收入或要素投入。

2. 微观经济学研究消费者行为时，所要阐述的核心问题是消费者均衡的原则。
   所谓消费者均衡指的是一个有理性的消费者所采取的均衡购买行为。进一步说
   它是指保证消费者实现效用最大化的均衡购买行为。
   但人的需要或欲望是无限的，而满足需要的手段是有限的。所以微观经济学
   所说的效用最大化只能是一种有限制的效用最大化。而这种限制的因素就是各种
   商品的价格和消费者的货币收入水平。