

# 支持向量机（三）：图解KKT条件和拉格朗日乘子法

原创 石头 机器学习算法那些事 2018-11-21

## 前言

支持向量机求解最优化参数的过程中需要用到拉格朗日乘子法和KKT条件，本文用清晰易懂的图解法说明拉格朗日乘子法和KKT条件的含义，希望能够帮助你理解这种最优化思想。

### 最优化问题

求  $f(x, y) = x^2 + y^2$  的最小值

本文讨论各种约束条件下的 $f(x, y)$ 的最小值。

### 函数等高线和梯度

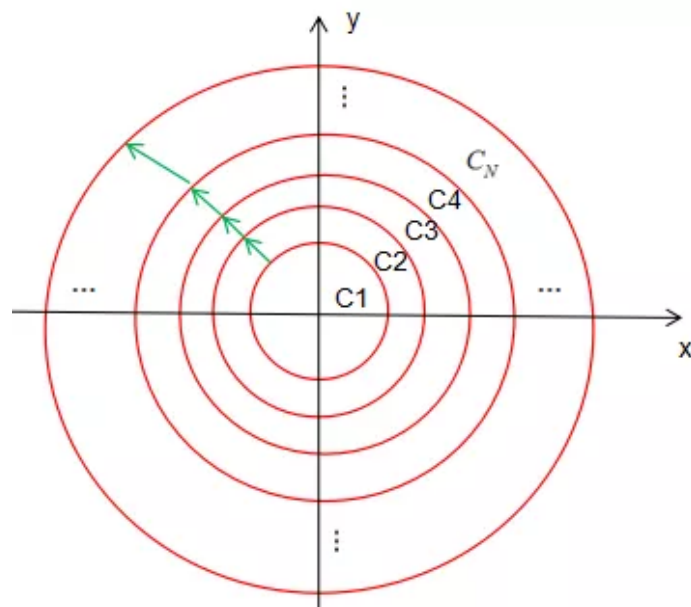
求  $f(x, y) = x^2 + y^2$  的最小值

令  $f(x, y) = C$

因此可以画出 $f(x, y)$ 一系列的等高线

$f(x, y) = C_i, i = 1, 2, \dots$

等高线如下图，同心圆为等高线，圆的半径越大，等高线也越大，即 $f(x, y)$ 值越大。



梯度方向是函数增加最快的方向，即等高线的法向量方向，图示的绿色箭头为梯度方向。

### 无约束条件下的最小值

$$f(x, y) = x^2 + y^2$$

$$\nabla_x f(x, y) = 2 \cdot x$$

$$\nabla_y f(x, y) = 2 \cdot y$$

令  $\nabla_x f(x, y) = 0$ ,  $\nabla_y f(x, y) = 0$ , 得:

$$x = 0, y = 0$$

$$\therefore \min_{x, y} f(x, y) = f(0, 0) = 0$$

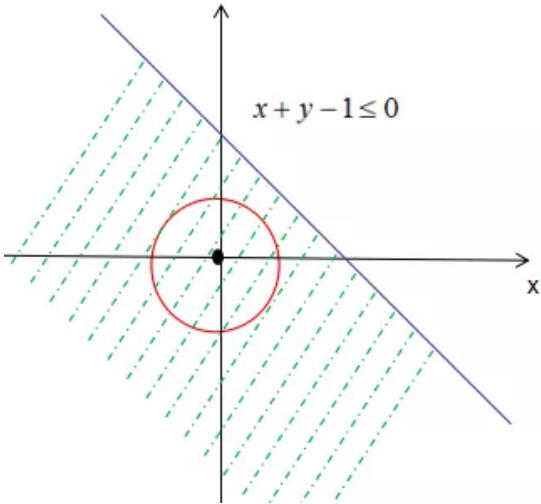
### 单一约束条件下的最小值

(1)

$$\min_{x, y} f(x, y)$$

$$s.t \quad g(x, y) = x + y - 1 \leq 0$$

解法:

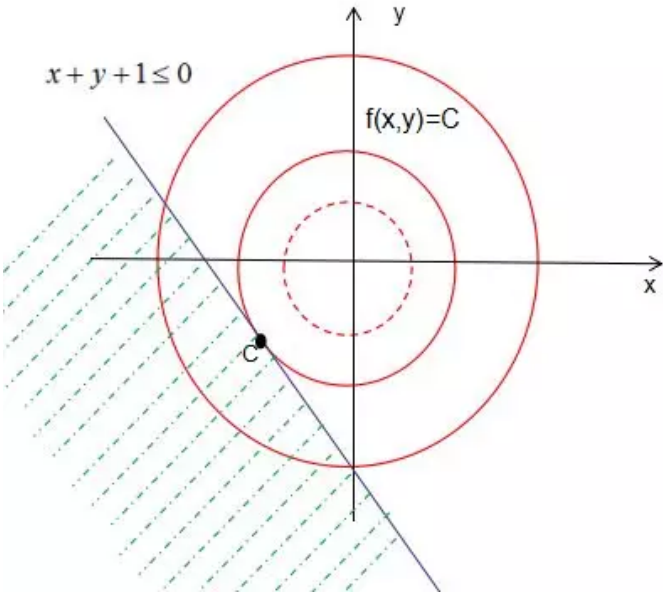


约束区域为绿色虚线区域，因此，目标函数最小值为0。

(2) 修改约束条件：

$$\min_{x,y} f(x,y)$$
$$s.t \quad g(x,y) = x + y + 1 \leq 0$$

图解法：



上图，当f(x,y)的等高线与约束条件的边界直线相切时，f(x,y)有最小值。

拉格朗日乘子法：

构建拉格朗日函数 ( $\alpha \geq 0$ ):

$$L(x, y, \alpha) = f(x, y) + \alpha \cdot g(x, y) \quad (4.1)$$

$$\nabla_x L(x, y, \alpha) = 2 \cdot x + \alpha \cdot \nabla_x g(x, y) \quad (4.2)$$

$$\nabla_y L(x, y, \alpha) = 2 \cdot y + \alpha \cdot \nabla_y g(x, y) \quad (4.3)$$

$$\nabla_\alpha L(x, y, \alpha) = x + y + 1$$

令  $\nabla_x f(x, y) = 0$ ,  $\nabla_y f(x, y) = 0$ ,  $\nabla_\alpha f(x, y, \alpha) = 0$ , 得:

$$x = -\frac{1}{2}, y = -\frac{1}{2}, \alpha = 1$$

$$\min_{x, y} f(x, y) = f\left(-\frac{1}{2}, -\frac{1}{2}\right) = \frac{1}{2}$$

结合图解法理解拉格朗日乘子法的含义:

构建拉格朗日函数:

$$L(x, y, \alpha) = f(x, y) + \alpha \cdot g(x, y)$$

$$\nabla_x L(x, y, \alpha) = \nabla_x f(x, y) + \alpha \cdot \nabla_x g(x, y)$$

$$\nabla_y L(x, y, \alpha) = \nabla_y f(x, y) + \alpha \cdot \nabla_y g(x, y)$$

令  $\nabla_x L(x, y, \alpha) = 0$ ,  $\nabla_y L(x, y, \alpha) = 0$ , 得:

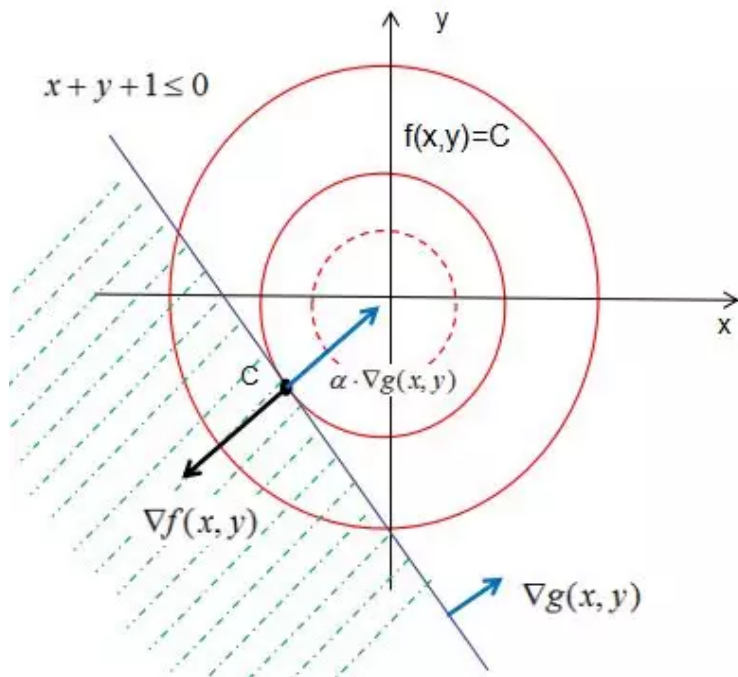
$$\nabla_x f(x, y) = -\alpha \cdot \nabla_x g(x, y) \quad (4.4)$$

$$\nabla_y f(x, y) = -\alpha \cdot \nabla_y g(x, y) \quad (4.5)$$

由式4.4和式4.5得到  $f(x, y)$  和  $g(x, y)$  的梯度关系:

$$\nabla f(x, y) = -\alpha \cdot \nabla g(x, y) \quad (4.6)$$

由 (4.6) 式的梯度关系, 表示在图形的含义如下图:



如上图，拉格朗日乘子 $\alpha$ 的含义是约束条件边界直线的法向量与目标函数等高线的法向量是共线向量，结合约束边界直线在C点的定义( $g(x,y)=0$ )，即KKT条件，表示为：

$$\begin{cases} \nabla_x L(x, y, \alpha) = \nabla_x f(x, y) + \alpha \cdot \nabla_x g(x, y) = 0 \\ \nabla_y L(x, y, \alpha) = \nabla_y f(x, y) + \alpha \cdot \nabla_y g(x, y) = 0 \\ \alpha \geq 0 \\ \alpha \cdot g(x, y) = 0 \\ g(x, y) \leq 0 \end{cases}$$

### 讨论

本节简单讨论上面为未提及的两种情况：

#### (1) 增加约束条件 $h(x,y)=0$

写成拉格朗日函数：

$$L(x, y, \alpha) = f(x, y) + \alpha \cdot g(x, y) + \beta \cdot h(x, y)$$

$$\nabla_x L(x, y, \alpha) = \nabla_x f(x, y) + \alpha \cdot \nabla_x g(x, y) + \beta \cdot \nabla_x h(x, y)$$

$$\nabla_y L(x, y, \alpha) = \nabla_y f(x, y) + \alpha \cdot \nabla_y g(x, y) + \beta \cdot \nabla_y h(x, y)$$

令  $\nabla_x L(x, y, \alpha) = 0$ ,  $\nabla_y L(x, y, \alpha) = 0$ , 得:

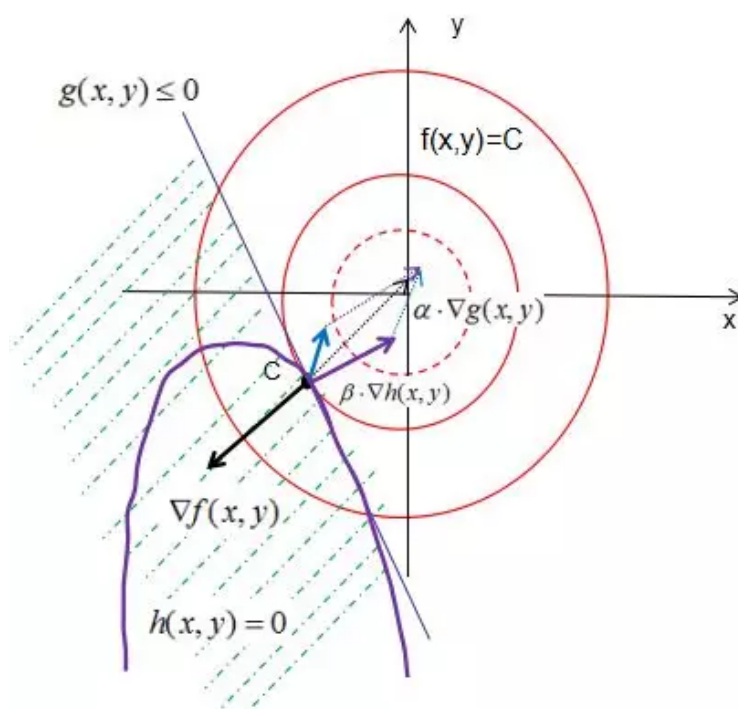
$$\nabla_x f(x, y) = -\alpha \cdot \nabla_x g(x, y) - \beta \cdot \nabla_x h(x, y)$$

$$\nabla_y f(x, y) = -\alpha \cdot \nabla_y g(x, y) - \beta \cdot \nabla_y h(x, y)$$

由上面两式得到  $f(x, y)$  和  $g(x, y)$  的梯度关系:

$$\nabla f(x, y) = -(\alpha \cdot \nabla g(x, y) + \beta \cdot \nabla h(x, y))$$

图形表示如下:

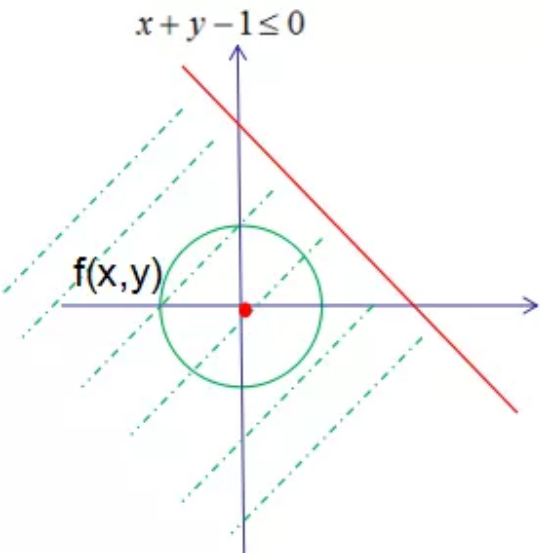


目标函数的梯度可以表示为约束条件的梯度展开的向量空间，系数为拉格朗日乘子  $\alpha$  和  $\beta$ 。若再增加限制条件，结论是一致的。

## (2) 拉格朗日乘子 $\alpha$ 的取值情况

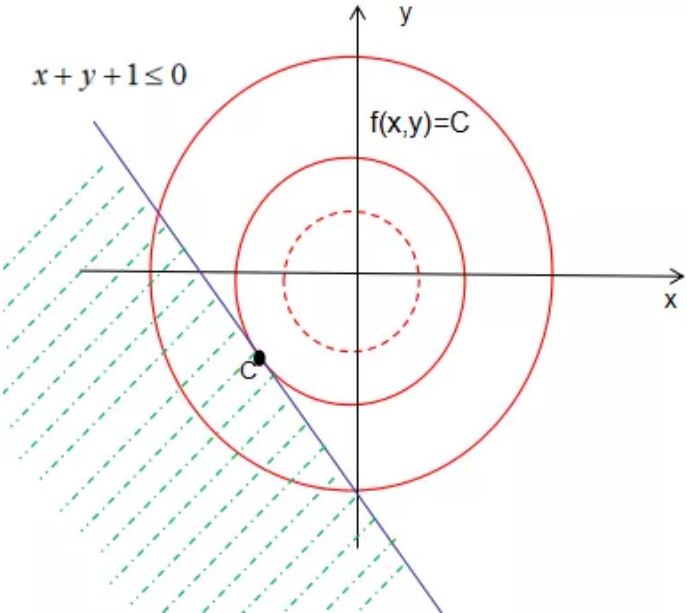
当目标函数的最优解在约束条件的区域内时，拉格朗日乘子  $\alpha = 0$ ；如下图：





最优解是对应的红色实心圆， $\alpha=0$ 。

当目标函数的最优解在约束条件的边界直线内时，拉格朗日乘子 $\alpha>0$ ，如下图：



最优解是对应的黑色实心圆，拉格朗日乘子 $\alpha>0$ 。

参考

<https://www.matongxue.com/madocs/987/>

推荐阅读文章

为什么梯度是函数变化最快的方向

■



-END-



长按二维码关注

机器学习算法那些事  
微信: beautifulife244

砥砺前行 不忘初心