

## 浅谈拉格朗日乘子法



老罗  
数学渣渣又想学算法的程序猿

73 人赞同了该文章

在介绍SVM算法之前，需要一些背景知识，优化理论就是其中的基石。在此之前，我对优化理论都没有很好地理解，今天看了一些知乎上大神们的解释，对拉格朗日乘子法和KKT条件有了一些全新的认识，今天特意mark一下。

感谢

理解 KKT 条件和对偶问题

217 赞同 · 19 评论 文章

和

非线性优化中的 KKT 条件该如何理解？

1505 关注 · 23 回答 问题

马同学的回答。

具体的细节和图画可以到以上两个优秀的回答中看，这里只是记录一些浅显的认识和理解，主要是帮助本人记忆和记录。

说到求解最优解，我们最先想到的应该就是对一个函数进行求导，然后令导数为0，接着就是求解这个方程，得到的就是最优解啦。

例如要求解 $\min_x f(x)$ ，那么就是解方程  $\nabla f(x) = 0$ ，最终的  $x^*$  为最优解。

没错，上面就是没有约束条件下最优解的办法，那么当有约束条件怎么办呢？这时候就要用到拉格朗日乘子法。**其实拉格朗日法就是把一个有约束问题转换成一个无约束问题，此时就需要引入拉格朗日项来起到控制效果。如果满足约束条件，那么这个项就应该有助于最优化。而不满足约束条件，**

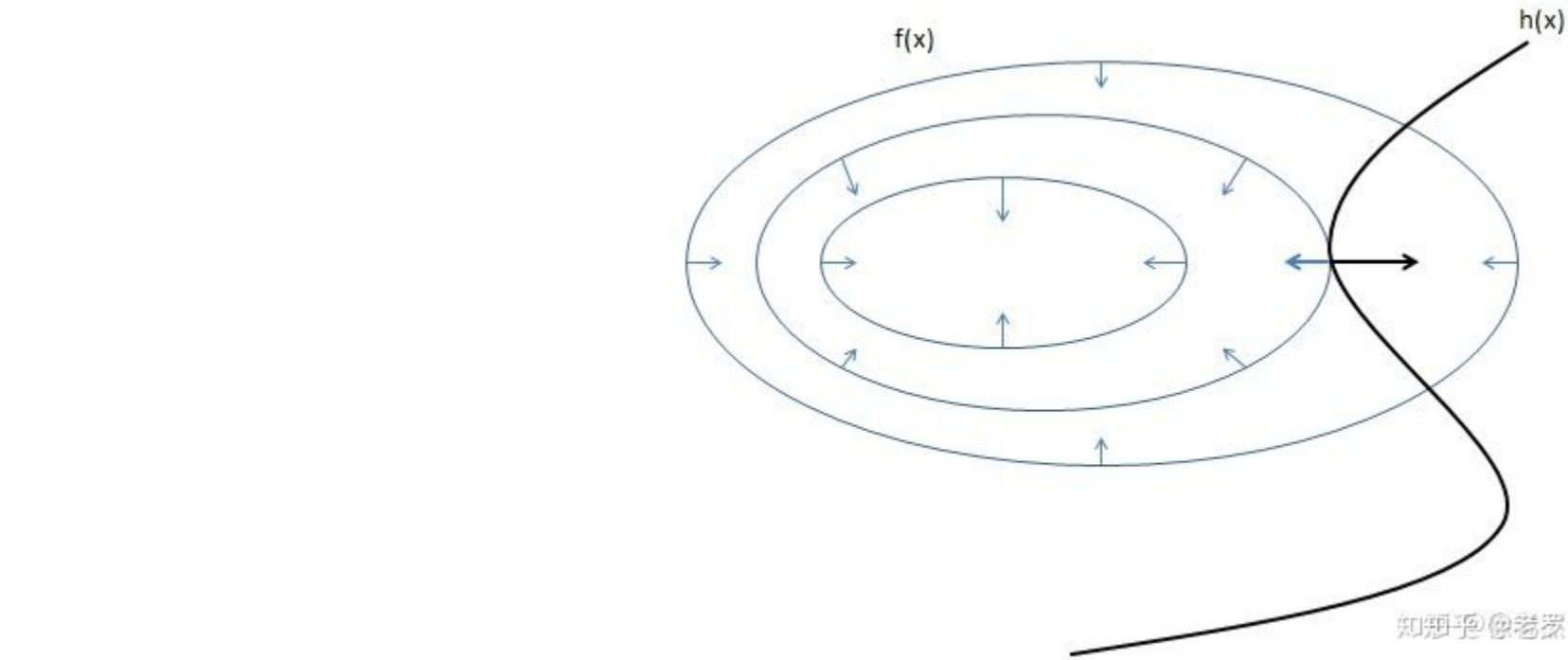


问题描述：

$\min_x f(x)$

$s.t. h(x) = 0$

则优化图如下：



等式约束条件的最优化图，来源：<https://zhuanlan.zhihu.com/p/36621652>

由图可以看到，蓝色线为函数等值线，整个平面是 $x$ 的空间，但是由于加了约束  $h(x) = 0$ ，所以 $x$ 的可行域就只能是黑色那条线，那条线表示的就是  $h(x) = 0$ 。因此找最优解的时候，就只能在黑色线上找。假如没有黑色线的约束，那么函数  $f(x)$  的极值就应该在最小的蓝色椭圆内（还有没有画出来的更小的椭圆，最终应该是一个点），蓝色的箭头就表示它们的梯度反方向（函数值减小的方向）。有了黑色线的约束，最优解应该就在蓝色线和黑色线相切的地方，因为相切，所以两条线的梯度反方向应该是方向相同或者相反。因此满足方程： $\nabla f(x) + \lambda \nabla h(x) = 0$ 。所以对于等式约束的最优化问题，当函数的梯度等于等式约束的梯度的线性组合时，可能可以找到最优解。这就是等式约束的最优化问题的KKT条件。满足KKT条件的点不一定是极值点，但是极值点一定满足KKT条件。

那么拉格朗日乘子法的做法就是定义一个新的拉格朗日函数： $L(x, \lambda) = f(x) + \lambda h(x)$ ，对该函数求导并另其为0，得到的就是上面的方程。这就是等式约束条件的拉格朗日乘子法。拉格朗日法的目的其实就是加多一项控制项，满足条件时，应该帮助最优化的结果；而不满足条件时，就应该破坏最优化的结果。

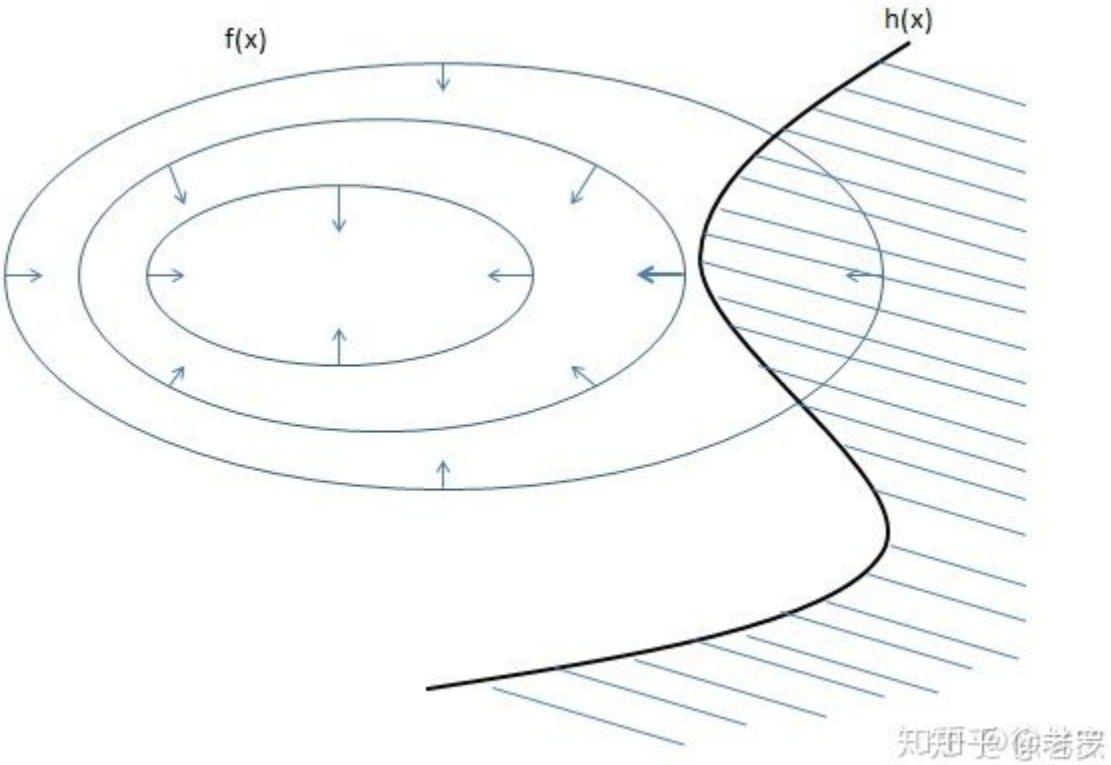
对于不等式约束的最优化问题就得引入KKT条件。



问题描述：

$$\min_x f(x)$$
$$s.t. h(x) \leq 0$$

则优化图如下：



不等式约束条件的最优化图，来源：<https://zhuanlan.zhihu.com/p/36621652>

因为我们要求解的问题是最小化问题，所以下面都会说是梯度的反方向。

蓝色斜线为部分为可行域，也就是  $h(x) \leq 0$ ，其他和上面一样。那么最优解要在这个区域里面找了。那么就可以有两种情况出现：

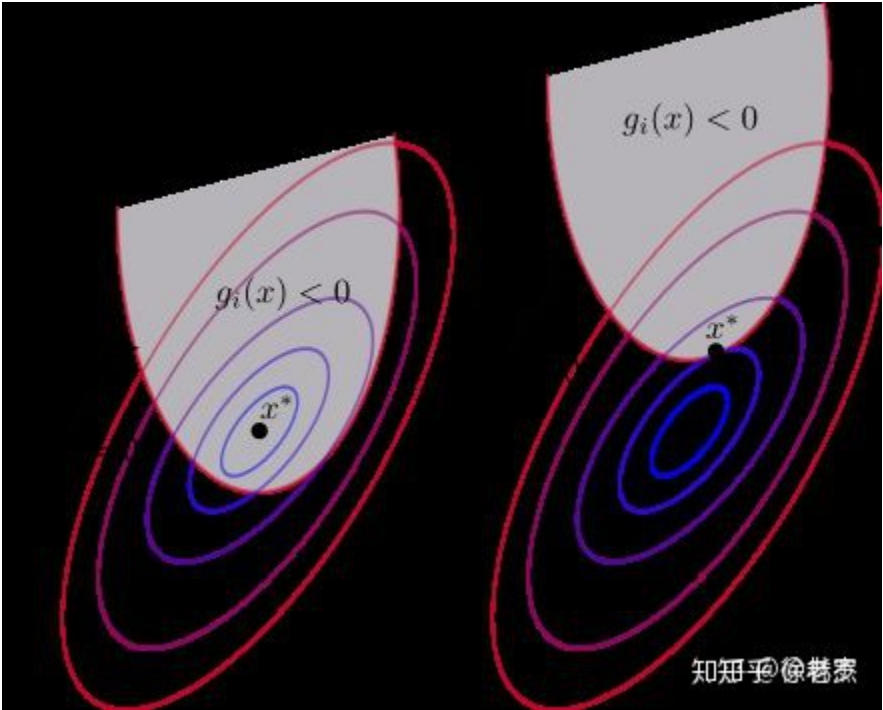
第一种就函数  $f(x)$  极值点就在  $h(x) < 0$  内（这些点也被称为不活跃点），从上图来看，也就是函数  $f(x)$  的收敛点在  $h(x)=0$  的右方，那么就相当于没有约束条件，只需要满足的方程  $\nabla f(x) = 0$  就行了。于是此时对于拉格朗日函数  $L(x, \lambda) = f(x) + \lambda h(x)$ ，其中的  $\lambda = 0$ ，即不需要考虑约束条件。

第二种情况就是函数  $f(x)$  极值点在  $h(x) > 0$ ，从上图来看，也就是函数  $f(x)$  的收敛点在  $h(x)=0$  的左方，很容易就可以看出此时极值点在边界上（ $h(x) = 0$ ，这些点也被称为活跃点），这时候就相当于上面的等式约束条件下的最优化问题，满足的方程为

$$\begin{aligned} \nabla f(x) + \lambda \nabla h(x) &= 0 \\ h(x) &= 0 \end{aligned}$$

↑

此时对于拉格朗日函数  $L(x, \lambda) = f(x) + \lambda h(x)$ ，其中的  $\lambda > 0$ 。因为这种情况下的  $\nabla f(x)$ 和  $\nabla h(x)$  方向必相反，所以要满足上面的方程，那么  $\lambda$  必须大于0。如下图，对于可行域的约束，如果  $f(x)$  的梯度方向是往白色区域去的话，那就成为了第一种情况，极值点在不活跃点内。如果是第二种情况，就  $f(x)$  的梯度方向也就只能往白色区域外（也就是蓝色椭圆那边去），这时候极值点就只能出现在不活跃点，由于相切，所以两个梯度方向必定相反。



两种不同的情况，图中的g(x)就是h(x)。来源：<https://zhuanlan.zhihu.com/p/36621652>

结合上面两种情况，简化上面的式子，则可以写出拉格朗日函数和KKT条件：

$$\begin{aligned} L(x, \lambda) &= f(x) + \lambda h(x) \\ s.t. \lambda &\geq 0, \\ h(x) &\leq 0, \\ \lambda h(x) &= 0. \end{aligned}$$

再推广到多个约束条件：

$$\begin{aligned} \min_x f(x) \\ s.t. h_i(x) &= 0 \quad i = 1, 2, \dots, m \\ g_j(x) &\leq 0 \quad j = 1, 2, \dots, n \end{aligned}$$

则拉格朗日函数为



引入的KKT条件为：

$$g_j(x) \leq 0$$

$$\mu_j \geq 0$$

$$\mu_j g_j(x) = 0$$

最后总结一下，对于有约束的最优化问题，使用拉格朗日乘子法来把问题转换成无约束问题进行求解，也就是问题的最优解必须满足KKT条件，KKT条件简单来说，就是函数的梯度方向由等式约束条件的梯度方向线性组合加上不等式约束条件的活跃点（也就是不等式约束条件等号成立的时候）的梯度的线性组合。由于该KKT条件的存在，所以才有拉格朗日乘子法。而拉格朗日法的目的其实就是加多一项控制项，满足条件时，应该帮助最优化的结果；而不满足条件时，就应该破坏最优化的结果。

编辑于 2020-11-29 22:02

[模式识别](#)   [机器学习](#)

写下你的评论...

5 条评论

默认

最新



erty

总结的很棒👍👍

2019-08-09

👍 2



无用良品

说人话的好文章5555我们老师只会堆公式

10-29

👍 赞



马上科普

请问这篇文章可以转载吗？谢谢

2021-09-07

👍 赞

▲ 赞同 73



💬 5 条评论

🔗 分享

♥️ 喜欢

★ 收藏


📄 申请转载



可以

2021-09-07

1



zerozero

感谢!

2020-08-28

赞



推荐阅读

拉格朗日  
乘子法

优化-拉格朗日乘子法

马铁荀



拉格朗日乘子法

Richard Kung



拉格朗日乘子法

李腾

发表于数形结合

拉格朗日乘子法

上周我们对线性svm的最大间隔进行了推导，得到以下结论  
 $\max \frac{1}{2} W^T$  W subject to  $y^{(i)}(W^T X^{(i)} + b) \geq 1, i=1,2,3,4,.....n$  (subject 的意思是约束于, n 是 样本数量)...

董长佐

