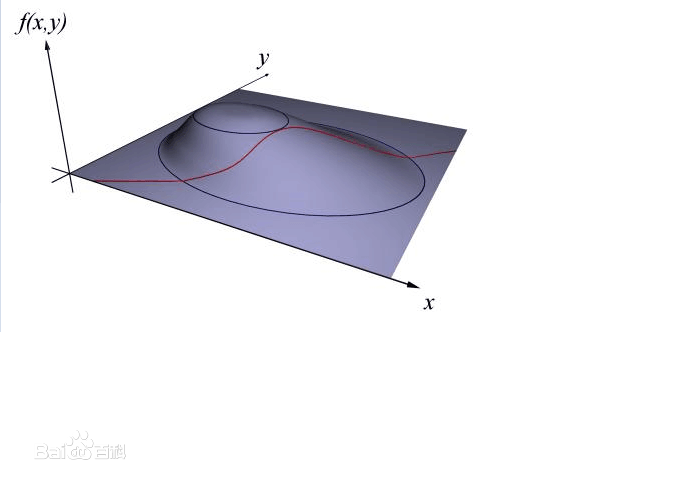
# **一下是copy前人的，这篇写的比较好。直接记录下来了。以便以后回顾。**



这是拉格朗日乘子法的经典理解图

# **机器学习数学原理（5）——广泛拉格朗日乘子法**

这一篇博客针对的是有约束的凸优化问题，主要是为后面的****最优间隔分类器****以及其演化的****SVM（支持向量机，Support Vector Machine）算法****作铺垫。****Andrew Ng****在讲解最优间隔分类器时运用了广泛拉格朗日乘子法但并没有讲的十分的明细，而是直接使用了结论，故笔者专门复习了拉格朗日乘子法并且学习了其在不等式约束情况下的优化（即广泛拉格朗日乘子法）。

这里要感谢博主Poll的博文：[http://www.cnblogs.com/maybe2030/p/4946256.html](http://www.cnblogs.com/maybe2030/p/4946256.html" \t "https://blog.csdn.net/z_x_1996/article/details/_blank)。笔者写的是以博主Poll的博文为骨架，同时加上了自己对其的补充，并且解释了原文中很容易引起误解的或者难理解的公式。当然这里也要感谢****MIT的在线数学课程****，博主Poll的博文正是课程上该内容的笔记。

另外老规矩，由于笔者水平有限，若出现不妥或者错误的地方，欢迎批评指出。

## **1 拉格朗日乘子法的基本思想**

### **1.1 思想概括**

首先我们先要明确这个算法的****目的****是****优化****，对于一个连续的函数来说其实就是求****全局极值****。

现在我们来考虑一下没有约束条件的凸优化问题。对高等数学略有了解的人应该都知道对于一个连续的函数求极值点，其实就应该将函数求导，然后取导数向量为零向量即可。

例如一个n元函数f(x1,x2,…,xn)，记作f(****X****)，当我们需要求其极值时，便分别求f(****X****)对于xi的偏导（i=1~n），然后令这n个偏导为零，便将求极值点的问题转化为n个n元方程构成的方程组求解。

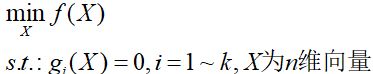
然而往往事情并没有这么简单，上述问题的前提往往在规划问题中过于纯粹。在实际问题中这n个自变量往往不是独立的，即满足一些约束条件。这样一来上述求导的方法便不再使用，因为其求导计算出的极值点****X****往往并不能满足约束条件。

现在我们细细分析一下导致这种差别的原因。不难发现，出现这个问题的本质便是****自变量的相互独立性被约束条件破坏掉****，导致我们并不能任意使用求导后的结果。若想解决这个问题，接下来就需要我们的拉格朗日乘子法。

在上述分析中我们已经知道，如果我们能够还原变量的独立性，换句话说，如果我们能够将约束优化问题转化为非约束问题，那么我们便可以继续使用求偏导这样简洁明了的方法来处理问题了。拉格朗日乘子法的思想便基于此。

作为一种优化算法，拉格朗日乘子法的基本思想就是****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****，或者我们可以这么看，****拉格朗日乘子法通过将k个约束条件转化进偏导方程组中的k个等式从而使得原问题不再出现约束****。拉格朗日乘子背后的数学意义是其为约束方程梯度线性组合中每个向量的系数。

解决的问题模型为约束优化问题：



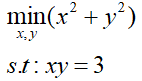
### **1.2 体现思想的数学实例**

首先，我们先以麻省理工学院数学课程的一个实例来作为介绍拉格朗日乘数法的引子。

****【麻省理工学院数学课程实例】****求双曲线xy=3上离远点最近的点。

****解：****

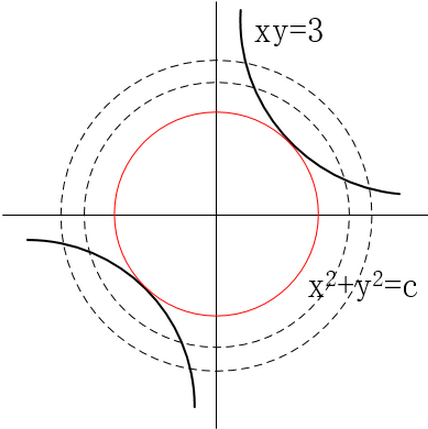
首先，我们根据问题的描述来提炼出问题对应的数学模型，即：



（两点之间的欧氏距离应该还要进行开方，但是这并不影响最终的结果，所以进行了简化，去掉了平方）

根据上式我们可以知道这是一个典型的约束优化问题，其实我们在解这个问题时最简单的解法就是通过约束条件将其中的一个变量用另外一个变量进行替换，然后代入优化的函数就可以求出极值。我们在这里为了引出拉格朗日乘数法，所以我们采用拉格朗日乘数法的思想进行求解。

我们将曲线族****x²+y²=c****画出来，如下图所示，



****当曲线族中的圆与xy=3曲线进行相切时，切点到原点的距离最短****。也就是说，当f(x,y)=c的等高线和双曲线g(x,y)相切时，我们可以得到上述优化问题的一个极值（注意：如果不进一步计算，在这里我们并不知道是极大值还是极小值）。

现在原问题可以转化为求当f(x,y)和g(x,y)相切时，x,y的值是多少？

如果两个曲线相切，那么它们的切线相同，即法向量是相互平行的，▽f//▽g.

由▽f//▽g可以得到，▽f=λ\*▽g。

这时，我们将原有的约束优化问题转化为了一种对偶的无约束的优化问题，如下所示：

| **原问题** | **转化后问题** |
| --- | --- |
| min f(x,y)=x²+y²(s.t. xy=3) | ▽f=λ\*▽g |
| min f(x,y)=x²+y²(s.t. xy=3) | fx=λ\*gx，fy=λ\*gy，xy=3 |
| 约束优化问题 | 无约束方程组问题 |

通过求解右边的方程组我们可以获取原问题的解，

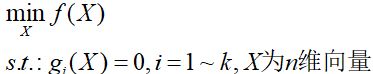
通过求解上式可得，λ=2或者是-2；当λ=2时，(x,y)=(sqrt(3), sqrt(3))或者(-sqrt(3), -sqrt(3))，而当λ=-2时，无解。所以原问题的解为(x,y)=(sqrt(3), sqrt(3))或者(-sqrt(3), -sqrt(3))。

通过举上述这个简单的例子就是为了体会拉格朗日乘数法的思想，****即通过引入拉格朗日乘子(λ)将原来的约束优化问题转化为无约束的方程组问题****。

## **2 等式约束下的优化问题**

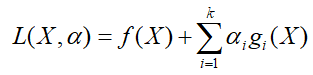
### **2.1 数学模型**

首先明确一下数学问题，等式约束下的优化问题其实就是前面提到的模型：

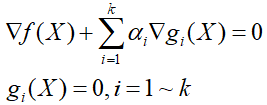


### **2.2 拉格朗日乘子法**

现在有k个约束条件，现在定义：



对X与α求偏导后有：



共（n+k）个方程。

当L取满足上述条件然后取得极值后，我们可以发现有

****min(L) = min(f)****

也就是说我们只要求出上述方程组，其解便是f的极值点。

则我们便将对于f的约束优化问题转化成了对于L的非约束优化问题。

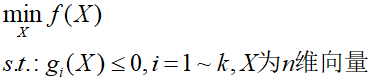
## **3 不等式约束下的优化问题**

前面的是对等式约束下的优化问题进行的求解，由于是等式约束，其实可以看做是在一个超曲面约束的空间面上移动点，找到使得目标函数最小的点。

但是在规划问题中，我们常常遇到的条件是****不超过什么什么****，****不低于什么什么****，这样一来，我们不得不面对不等式约束下的优化问题。如果说等式约束是在一个超曲面约束的空间面上移动点，那么不等式约束的优化问题便是在几个超曲面分割后的一个区域内移动点，故问题的复杂度大大增加。

### **3.1 数学模型**

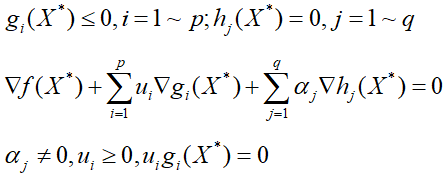
这里我们直接来看数学模型吧！



由于后面等式约束问题我们可以直接转化成无约束问题，所以不再设置等式约束项。

### **3.2 KKT条件**

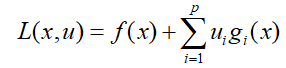
****KKT条件****是指在满足一些有规则的条件下, 一个非线性规划(Nonlinear Programming)问题能有最优化解法的一个必要和充分条件. 这是一个****广义化拉格朗日乘子法的成果****. 一般地, 一个最优化数学模型的列标准形式参考开头的式子, 所谓 Karush-Kuhn-Tucker 最优化条件，就是指上式的最优点****X\*****必须满足下面三个条件:



KKT条件第一项是说最优点****X\*****必须满足所有等式及不等式限制条件, 也就是说最优点必须是一个可行解, 这一点自然是毋庸置疑的。第二项表明在最优点X\*, ∇f必须是∇gi和∇hj的线性組合, μi和αj都叫作拉格朗日乘子。 所不同的是不等式限制条件有方向性, 所以每一个μi都必须大于或等于零, 而等式限制条件没有方向性，所以αj没有符号的限制, 其符号要视等式限制条件的写法而定。

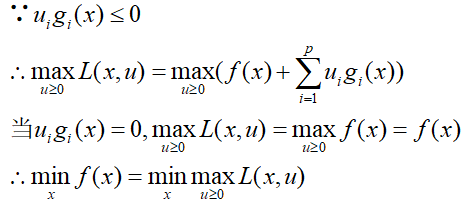
我们来看一下****KKT条件****的由来：

令拉格朗日算子为



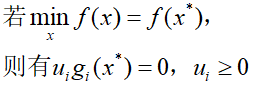
其中ui>=0。

那么有

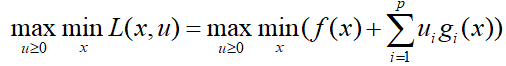


****这里的“当”并不是一个假设，而是max达成需要满足的条件。****

由上述推导不难发现：



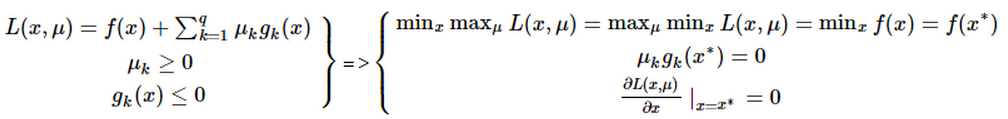
接下来我们来看一下其****对偶问题（注意min与max运算符的顺序！）****：



这里我们知道当我们求了min过后，无论x取什么值，后面的部分都是小于等于0的，所以max的时候一定是将后面一项变为0，所以有：

IMG_268

将上面结论整理一下我们便有了(下图来源于Poll博文)



即通过f(x)偏导为0获得了n个方程（n为x维数），然后又通过ug(x)=0获得了q个方程，即通过增加q个ui来将不等式约束问题转化为（n+q）个方程式。

***以下是对KKT讲的最好的一个***

