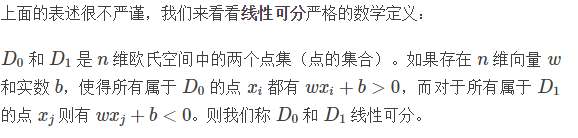
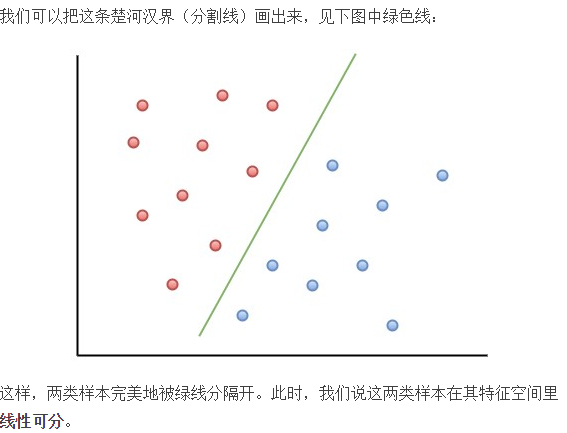
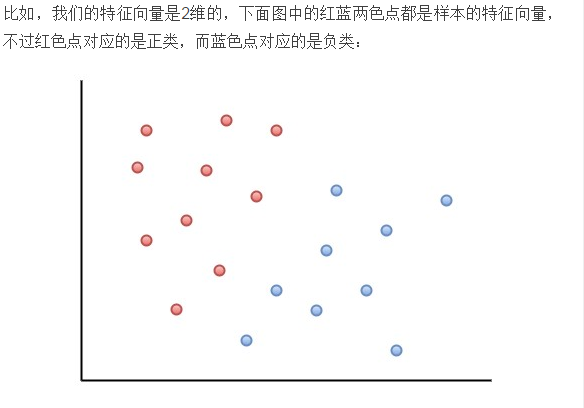
## 第18课：SVM——线性可分 SVM 原理

1. 线性可分和超平面

****二分类****问题就是：给定的各个样本数据分别属于两个类之一，而目标是确定新数据点将归属到哪个类中。



超平面：n 维欧氏空间中维度等于 n-1 的线性子空间。

维欧氏空间（直线）中的超平面为0维（点），2维欧氏空间中的超平面为1维（直线）；3维欧氏空间中的超平面为2维（平面）；以此类推……

在数学意义上，将线性可分的样本用超平面分隔开的分类模型，叫做线性分类模型，或线性分类器。

我们可以想象，在一个样本特征向量线性可分的特征空间里，可能有许多超平面可以把两类样本分开。在这种情况下，我们当然要找最佳超平面。

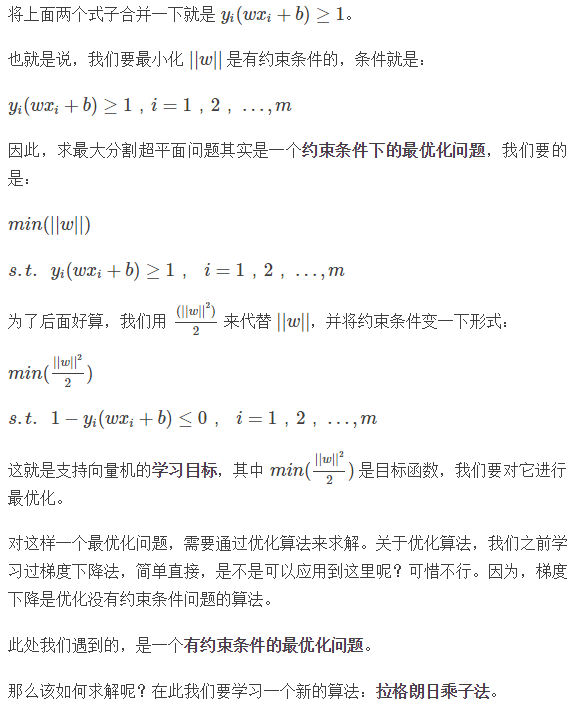
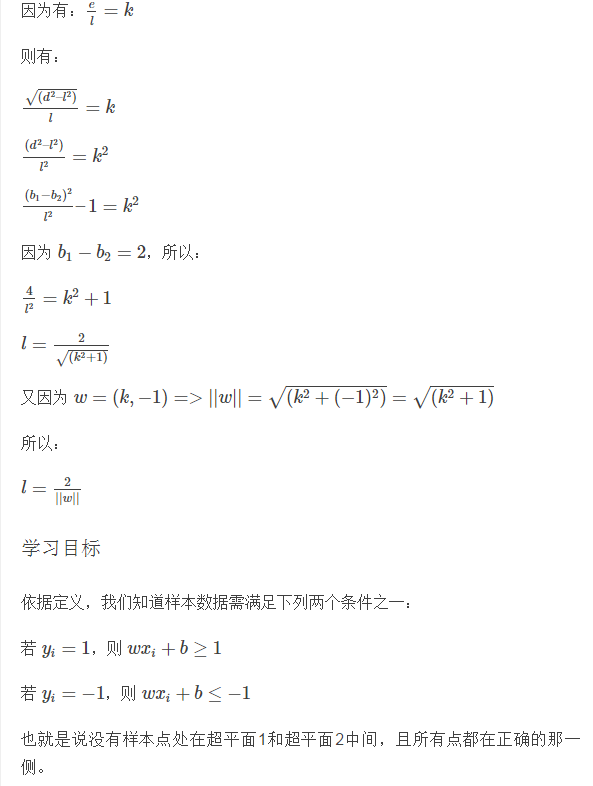
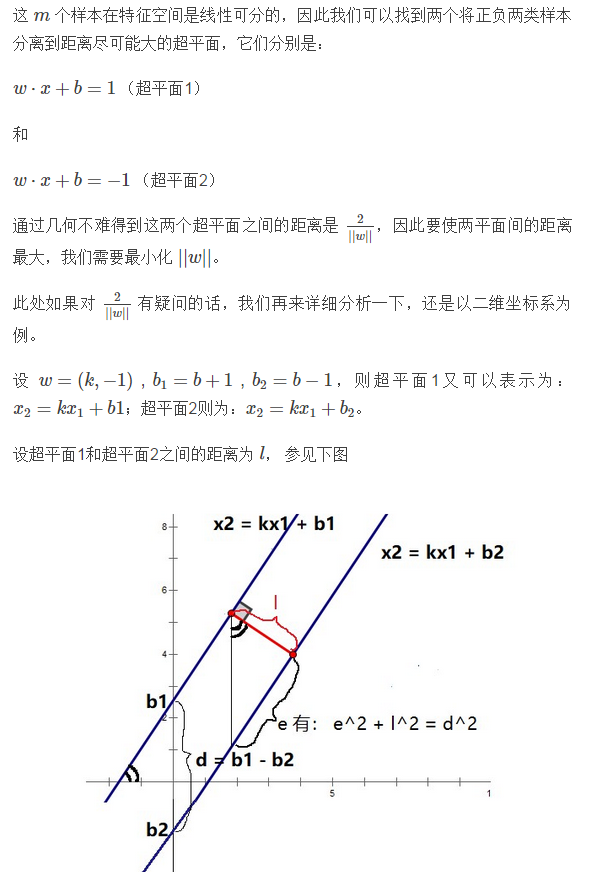
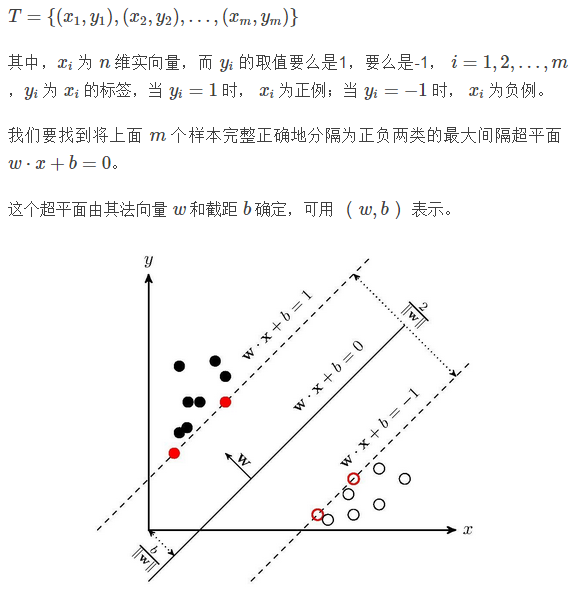
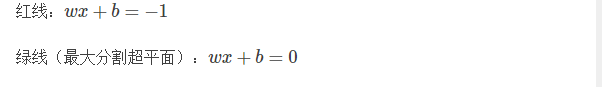
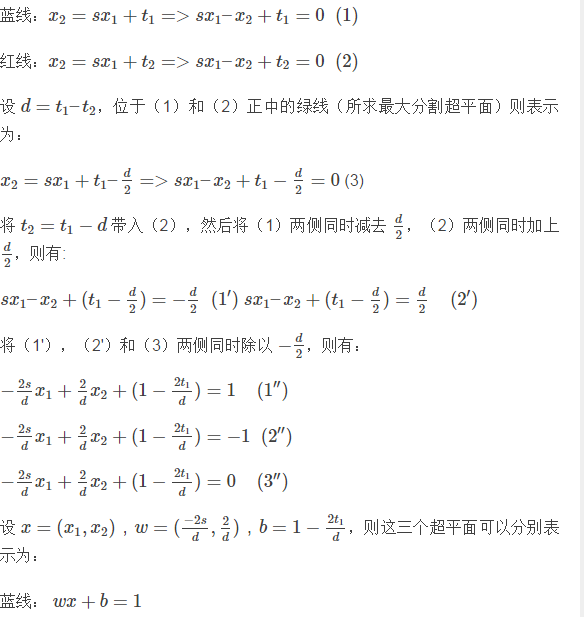
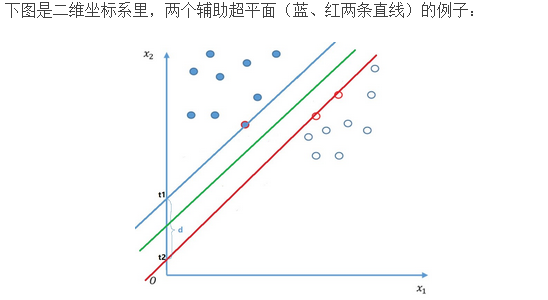
什么样的超平面是最佳的呢，一个合理的策略是：以最大间隔把两类样本分开的超平面，是最佳超平面！

因此，我们将这样的超平面作为最佳超平面：

两类样本分别分隔在该超平面的两侧；

两侧距离超平面最近的样本点到超平面的距离被最大化了。

这样的超平面又叫做最大间隔超平面。

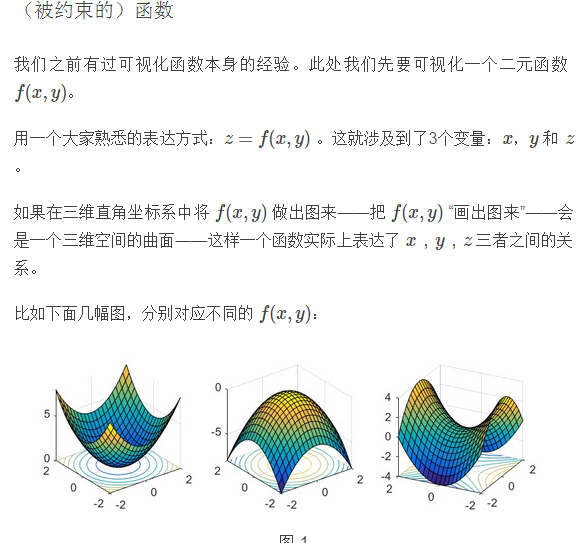


****线性可分支持向量机****就是：以找出线性可分的样本在特征空间中的最大间隔超平面为学习目的的分类模型。

2.线性可分支持向量机

## 第19课：SVM——直观理解拉格朗日乘子法

1. 可视化函数及其约束条件



（函数的）约束条件

函数 f(x,y)f(x,y) 的约束条件为：g(x,y)=0g(x,y)=0。

g(x,y)=0g(x,y)=0 又可以写成：y=h(x)y=h(x) 的形式——它所表达的是 xx 与 yy 两者之间的相互关系！

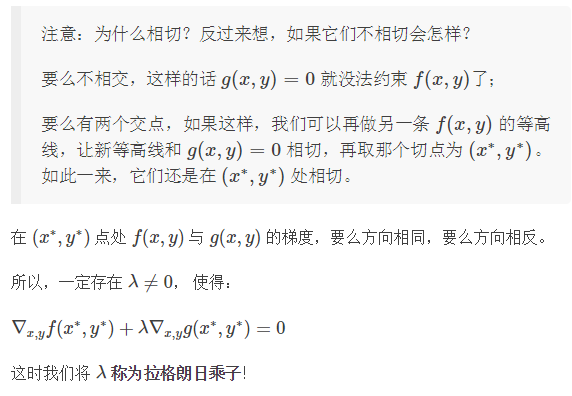
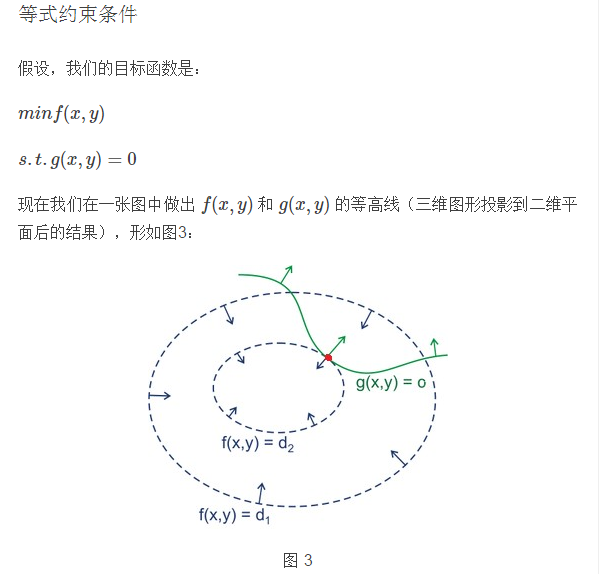
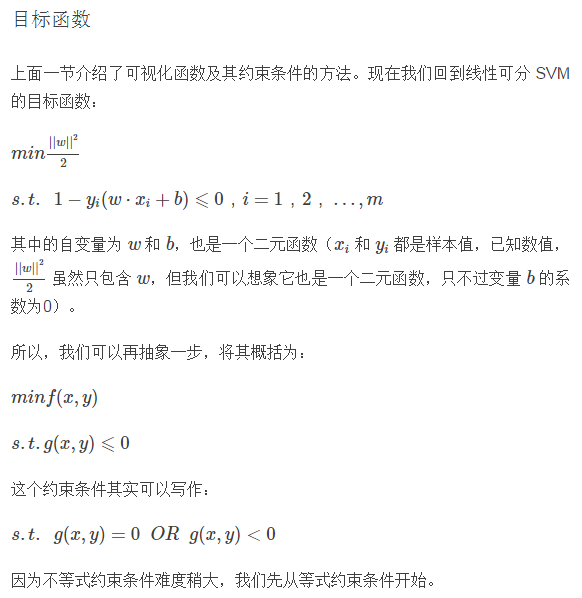
y=h(x)y=h(x)，如果在二维直角坐标中作图，形状应该是一条曲线。

注意：直线也属于广义的曲线；平面也属于曲面。我们说的曲线是包括直线的，曲面则包括平面。

#### 约束条件对函数的约束

那么，一个二维图形对一个三维图形的约束从何体现呢？让我们这样做：

1. 在自己的头脑中建立一个三维直角坐标系，有 xx 轴、yy 轴、zz 轴， 它们互相垂直。
2. f(x,y)f(x,y) 对应图形是三维坐标系里面的一个曲面——一个（可能是奇形怪状的）“体”的“外皮”。
3. 在 z=0z=0 的平面上，把 y=h(x)y=h(x) 的图形（一条曲线）画出来。
4. 将第 3 步做出的那条曲线沿着平行 z 轴的方向平移，它平移过的轨迹也形成了一个曲面，这个曲面和 z=f(x,y)z=f(x,y) 形成的曲面会相交，交叠的部分是一条三维空间中的曲线。
5. 换个角度考虑，第 4 步形成的曲线，其实就是 g(x,y)=0g(x,y)=0 在 z=0z=0 平面上形成的曲线，在 z=f(x,y)z=f(x,y) 形成的曲面上的投影。
6. 拉格朗日乘子法

拉格朗日乘子和拉格朗日函数

定义拉格朗日函数：L(x,y,λ)=f(x,y)+λg(x,y)L(x,y,λ)=f(x,y)+λg(x,y) —— 其中 λλ 是拉格朗日乘子。

拉格朗日函数把原本的目标函数和其限制条件整合成了一个函数。

拉格朗日函数对 x，yx，y 求偏导：

L′x,y(x,y,λ)=f′x,y(x,y)+λg′x,y(x,y)Lx,y′(x,y,λ)=fx,y′(x,y)+λgx,y′(x,y)

我们令拉格朗日函数对 x，yx，y 求偏导的结果为 00，则有：

f′x,y(x,y)+λg′x,y(x,y)=0fx,y′(x,y)+λgx,y′(x,y)=0

我们刚才已经知道了，f(x,y)f(x,y) 符合 g(x,y)=0g(x,y)=0 约束的极小值的点 (x∗,y∗)(x∗,y∗) 满足：∇x,yf(x∗,y∗)+λ∇x,yg(x∗,y∗)=0∇x,yf(x∗,y∗)+λ∇x,yg(x∗,y∗)=0，用导函数表示就是：f′x,y(x∗,y∗)+λg′x,y(x∗,y∗)=0fx,y′(x∗,y∗)+λgx,y′(x∗,y∗)=0。也就是说我们要求的极小值点正好满足拉格朗日函数对 x、yx、y 求导后，令其结果为 0 形成的导函数。

既然如此，我们当然也就是可以直接引入一个新的变量：λλ —— 拉格朗日乘子，和目标函数、约束条件一起，先构造出拉格朗日函数 L(x,y,λ)L(x,y,λ)。

然后再令 L′x,y(x,y,λ)=0Lx,y′(x,y,λ)=0，之后通过最优化 L′x,y(x,y,λ)=0Lx,y′(x,y,λ)=0，求取 (x∗,y∗)(x∗,y∗) 的值。

另外，而令拉格朗日函数对 λλ 求偏导的结果为 00：L′λ(x,y,λ)=0Lλ′(x,y,λ)=0，就得到了约束条件：g(x,y)=0g(x,y)=0。

拉格朗日函数也可以通过求导变化成约束条件本身。

于是，原本有约束的优化问题，就可以转化为对拉格朗日函数的无约束优化问题了。

不等式约束条件

了解了约束条件是等式的情况，我们再来看约束条件是不等式的情况。

当条件是：g(x,y)⩽0g(x,y)⩽0 时, 我们可以它拆分成：g(x,y)=0ORg(x,y)<0g(x,y)=0ORg(x,y)<0，两种情况来看。

这两种情况下可以先各取一个极小值，再比较一下，哪个更小就用哪个。

拆分后的严格不等式约束条件：g(x,y)<0g(x,y)<0

而对于 g(x,y)<0g(x,y)<0 情况，因为 <0<0 是一个区间，而不再是对应到三维坐标中的某个固定的曲面。

这种情况下，如果 f(x,y)f(x,y) 的极小值就在这个区间里，然后直接对 f(x,y)f(x,y) 求无约束的极小值就好了。

对应到拉格朗日函数就是：令 λ=0λ=0，直接通过令 f(x,y)f(x,y) 的梯度为 00 求 f(x,y)f(x,y) 的极小值。

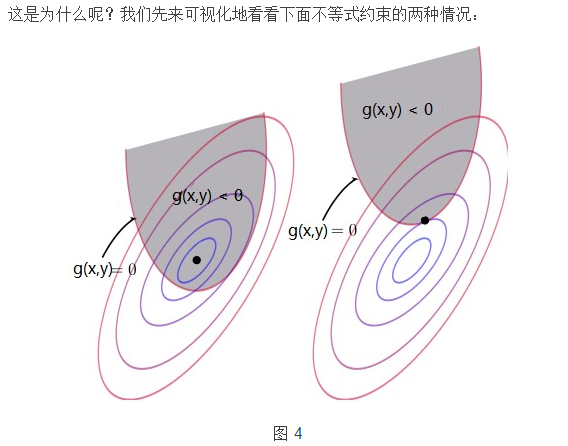
求出极小值 f(x∗,y∗)f(x∗,y∗) 之后，再判断 (x∗,y∗)(x∗,y∗) 是否符合约束条件。

拆分后的等式约束条件：g(x,y)=0g(x,y)=0

g(x,y)=0g(x,y)=0 可以参照上一小节的做法。

有一点要注意，在仅有等式约束条件时，我们设 f′x,y(x∗,y∗)+λg′x,y(x∗,y∗)=0fx,y′(x∗,y∗)+λgx,y′(x∗,y∗)=0，只要常数 λ≠0λ≠0 就可以了。

但是在以从 g(x,y)⩽0g(x,y)⩽0 中拆分出来的等式为约束条件时，我们要求：存在常数λ>0λ>0, 使得 f′x,y(x∗,y∗)+λg′x,y(x∗,y∗)=0fx,y′(x∗,y∗)+λgx,y′(x∗,y∗)=0

如果是图4左侧的状况，限制条件就又变成了 g(x,y)<0g(x,y)<0，且直接求 f(x,y)f(x,y)极小值就行。

只有是右侧情况时，不等式约束条件才又变成了等式。这种情况，才是我们在拆分了不等式条件后有意义的情况。

在这种情况下，最终找到的符合约束条件的极小值 f(x∗,y∗)f(x∗,y∗) 肯定不是 f(x,y)f(x,y)的极小值。

怎么用数学方式来表现这一点呢？

大家回想一下，****函数梯度的物理意义****：实际上，一个函数在某一点的梯度方向（向量方向）指明了该函数的函数值相对于当前函数值增量的方向。

也就是说：我们在 g(x∗,y∗)g(x∗,y∗) 的梯度方向上前进一个长度：ϵϵ（ϵϵ 很小很小），到达点 (x∗∗,y∗∗)(x∗∗,y∗∗)，则 g(x∗∗，y∗∗)>g(x∗,y∗)g(x∗∗，y∗∗)>g(x∗,y∗)。

注意：此处可以用 z=x+yz=x+y 和 z=−x−yz=−x−y 来直观感受一下。二维平面（比如 z=0z=0 平面）上的 y=xy=x 是这两个函数共同的等高线投影，不过 z=x+yz=x+y 的导数方向为 (1,1)(1,1)，而 z=−x−yz=−x−y 的导数方向为 (−1,−1)(−1,−1)，两个导向量共线，但方向相反，正好分别指向各自上升的区域。

因为 g(x∗,y∗)=0g(x∗,y∗)=0，又根据梯度的物理意义，则 (x∗,y∗)(x∗,y∗) 点处 g(⋅)g(⋅) 函数的梯度方向指向的是 g(x,y)>0g(x,y)>0 的区域。

偏偏我们的约束条件是 g(x，y)<0g(x，y)<0，那说明不等式约束条件所对应的区域与 (x∗,y∗)(x∗,y∗) 点处 g(⋅)g(⋅) 函数的梯度方向是相反的。

我们现在考虑的是图4右侧的情况，此时 f(⋅)f(⋅) 函数带约束条件的极小值点 (x∗,y∗)(x∗,y∗) 就位于 g(x,y)=0g(x,y)=0 对应的曲线上。

而 (x∗,y∗)(x∗,y∗) 点处 f(⋅)f(⋅) 函数的梯度方向正是 f(⋅)f(⋅) 函数本身的增量方向，因此在 g(x,y)<0g(x,y)<0 所在的区域对应的应该是 f(x,y)f(x,y) 的递增方向，如若不然，就应该是图4左侧的情况了，不是吗？

只有当 f(x,y)f(x,y) 梯度方向和 g(x,y)<0g(x,y)<0 区域所在方向相同，也就是和 (x∗,y∗)(x∗,y∗)点处 g(⋅)g(⋅) 函数梯度方向相反，那么 f(x,y)f(x,y) 的约束条件极小值才会出现在 g(x,y)=0g(x,y)=0 的曲线上。

所以，在这种情况下，存在常数 λ>0λ>0，使得 f′x,y(x∗,y∗)=−λg′x,y(x∗,y∗)fx,y′(x∗,y∗)=−λgx,y′(x∗,y∗)，进一步导出：

f′x,y(x∗,y∗)+λg′x,y(x∗,y∗)=0，λ>0fx,y′(x∗,y∗)+λgx,y′(x∗,y∗)=0，λ>0

KKT 约束条件

我们将上面拆分开的严格不等式和等式两种情况再整合起来:

如果严格不等式成立时得到 f(⋅)f(⋅) 函数的极小值，则有 λ=0λ=0，这样才能将拉格朗日函数直接转换为原始函数。则有 λg(x,y)=0λg(x,y)=0。

如果等式成立时得到 f(⋅)f(⋅) 函数的约束条件极小值，则必然存在 λ>0λ>0，且同时 g(x,y)=0g(x,y)=0, 因此也有 λg(x,y)=0λg(x,y)=0。

于是，对于不等式约束条件 g(x,y)⩽0g(x,y)⩽0，最终的约束条件变成了：

g(x,y)⩽0g(x,y)⩽0;

λ⩾0λ⩾0;

λg(x,y)=0λg(x,y)=0

这样由1变3的约束条件，叫做 KKT 约束条件。

目标函数转化为拉格朗日函数

当然，KKT 条件不是单独成立的，它是拉格朗日乘子法的一部分！

当我们在约束条件下求解函数最优化问题，且约束条件为不等式时（问题描述如下）：

Optimizef(x,y)Optimizef(x,y)

s.t.g(x,y)⩽0s.t.g(x,y)⩽0

我们针对目标函数生成拉格朗日函数为：

L(x,y,λ)=f(x,y)+λg(x,y)L(x,y,λ)=f(x,y)+λg(x,y)

同时有 KKT 条件：

g(x,y)⩽0g(x,y)⩽0

λ⩾0λ⩾0

λg(x,y)=0λg(x,y)=0

假设最优化结果对应点为 (x∗,y∗)(x∗,y∗)，则当目标函数为：

minf(x,y)minf(x,y)

s.t.g(x,y)⩽0s.t.g(x,y)⩽0

时，若存在 λ≠0λ≠0，使得 f′x,y(x∗,y∗)+λg′x,y(x∗,y∗)=0fx,y′(x∗,y∗)+λgx,y′(x∗,y∗)=0，则 λ>0λ>0。

而当目标函数为：

maxf(x,y)maxf(x,y)

s.t.g(x,y)⩽0s.t.g(x,y)⩽0

时，若存在 λ≠0λ≠0，使得 f′x,y(x∗,y∗)+λg′x,y(x∗,y∗)=0fx,y′(x∗,y∗)+λgx,y′(x∗,y∗)=0，则 λ<0λ<0。

或者写作：此时，若存在 λ≠0λ≠0，使得 f′x,y(x∗,y∗)−λg′x,y(x∗,y∗)=0fx,y′(x∗,y∗)−λgx,y′(x∗,y∗)=0，则λ>0λ>0。

注意：上面几个式子要注意：

L(x,y,λ)=f(x,y)+λg(x,y)L(x,y,λ)=f(x,y)+λg(x,y) 是拉格朗日函数的定义

λ⩾0λ⩾0 是之前我们推导出来的 KKT 条件；

存在常数 λ>0λ>0, 使得 f′x,y(x∗,y∗)+λg′x,y(x∗,y∗)=0fx,y′(x∗,y∗)+λgx,y′(x∗,y∗)=0 或者 f′x,y(x∗,y∗)−λg′x,y(x∗,y∗)=0fx,y′(x∗,y∗)−λgx,y′(x∗,y∗)=0 是推导 KKT 条件过程中经历的一个步骤——这个步骤中涉及到的是在 (x∗,y∗)(x∗,y∗) 这个确定点处 ff 函数和 gg 函数的梯度。