拉格朗日乘子法及KKT条件

预备知识点

有以下预备知识点，其中有页码标注的知识点来源于高等数学-第六版-下册(同济大学)：

1. P98：通过点$M(x_0, y_0, z_0)$且垂直于切平面的直线称为曲面在该点的法线。
2. P105：函数$f(x,y)$在一点(x_0, y_0)的梯度$\nabla f(x_0, y_0)$的方向就是等值线f(x,y)=c在这点的法线方向n。
3. ps：网上搜到的正交概念：曲面在某点的切平面与向量垂直
4. 几何上，两条曲线相切就是他们在这点有一个交点而且切线相同，也就是两个曲线相切他们一定有相同的切线，故梯度平行。知乎搜的问答，有待更新为课本定义。

正文

拉格朗日乘子法可以用于求解带约束条件的优化问题，一般来说，约束条件分为等式约束和不等式约束：

- 等式约束的优化问题，直接应用拉格朗日乘子法去求取最优解。
- 不等式约束的优化问题，转化为在满足KKT约束条件下应用拉格朗日乘子法求解。

等式约束优化

考虑带一个等式约束的优化问题：

$$\begin{gathered}\min_x f(x) \\ s.t.\ g(x) = 0\end{gathered}$$

从几何角度看，有以下结论：

- 对于约束曲面上的任意点x，该点的梯度$\nabla g(x)$正交于约束曲面.（因为，$g(x) = c$的法向量垂直于曲面上切平面，而$g(x,y) = c$的法向量等于$g(x,y)$的梯度，所以梯度正交约束平面。）
- 在最优点x^*，目标函数在该点的梯度$\nabla f(x^*)$正交于约束平面.（因为，若$\nabla f(x^*)$与约束曲面不正交，那么$f(x)$等值线与约束曲面不相切，那么仍可在约束曲面上移动该点使函数值进一步下降。）

由此可知，在最优点x^*，梯度$\nabla g(x)$与$\nabla f(x)$平行，即存在$\lambda \neq 0$使得

$$\nabla f(x^*) + \lambda \nabla g(x^*) = 0 \tag{A.1}$$

λ称为拉格朗日乘子，定义拉格朗日函数：

$$L(x, \lambda) = f(x) + \lambda g(x) \tag{A.2}$$

可以看出：

- 将$L(x, \lambda)$对x的偏导数$\nabla_x L(x, \lambda)$置零，即得式A.1
- 将$L(x, \lambda)$对λ的偏导数$\nabla_\lambda L(x, \lambda)$置零，即得约束条件$g(x) = 0$.

所以，使用拉格朗日函数求偏导数，并使之为零，方程联立起来求解得出的点是原问题的可能极值点，是求等式约束优化问题解的必要条件。至于如何确定所求得的点是否为极值点，在实际问题中往往可根据问题本身的性质来判断。

不等式约束优化

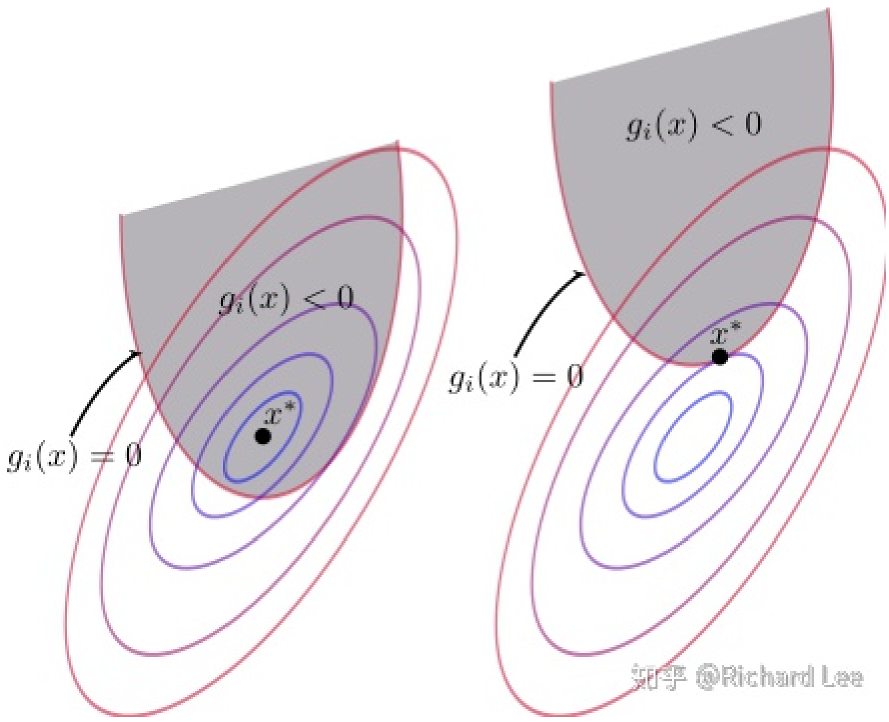
考虑带一个不等式约束的优化问题：

$$\min_x f(x)$$
$$s.t.\ g(x) \le 0$$

构造拉格朗日函数

$$L(x, \lambda) = f(x) + \mu g(x) \tag{A.3}$$

借助一张图来理解，图源于拉格朗日乘子法与对偶问题 - Richard Lee的文章 - 知乎



可以看出，可以根据目标函数$f(x)$的最优解x^*是否在可行域内分成两种情况：

1. x^*在可行域内(上图左边情况)，即$g(x)<0$，那么不等式约束不起作用，直接求$f(x)$的极值即可，等价于将μ置零，然后求$L(x,\mu)$的极值，如下所示：

$$\begin{cases}\nabla_x f(x)=0\\ \mu=0\\ g(x)<0\end{cases} \tag{A.4}$$

2. x^*在可行域边界上(上图右边情况)，即$g(x)=0$，那么不等式约束退化为等式约束，即在边界上取得极小值。由于梯度指向增大的方向，所以从上图右边可以看出，梯度$\nabla g(x)$与$\nabla f(x)$反向相反，故$\mu>0$，最终约束条件：

$$\begin{cases}\nabla_x f(x)+\mu\nabla_x g(x)=0\\ \mu>0\\ g(x)=0\end{cases} \tag{A.5}$$

合并A.4，A.5得到更一般的形式-KKT(Karush-Kuhn-Tucker)条件：

$$\begin{cases}\nabla_x L(x,\lambda)=\nabla_x f(x)+\mu\nabla_x g(x)=0\\ \mu\geq 0\\ g(x)\leq 0\\ \mu g(x)=0\end{cases} \tag{A.6}$$

所以，极值点一定满足KKT条件，即KKT条件是不等式约束优化问题解的必要条件。

等式约束+不等式约束优化

考虑既有等式约束，又有不等式约束的优化问题：

$$\begin{aligned} &\min_{\boldsymbol{x}} f(\boldsymbol{x})\\ s.t.\quad & c_i(\boldsymbol{x})\leq 0,\ \ i=1,2,..,k\\ & h_j(\boldsymbol{x})=0,\ \ j=1,2,..,l \end{aligned} \tag{A.7}$$

构造拉格朗日函数：

$$L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta})=f(\boldsymbol{x})+\sum_{i=1}^{k}\alpha_i c_i(\boldsymbol{x})+\sum_{j=1}^{l}\beta_j h_j(\boldsymbol{x}) \tag{A.8}$$

KKT条件为：

$$\begin{cases}\nabla_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta})=0\\ \alpha_i c_i(x)=0\ ,i=1,2,...k\\ \alpha_i\geq 0\ ,i=1,2,...k\\ c_i(\boldsymbol{x})\leq 0\ ,i=1,2,...k\\ h_i(\boldsymbol{x})=0\ ,j=1,2,...,l\end{cases} \tag{A.9}$$

此外，不再证明的定理：当$f(\boldsymbol{x})$和$c_i(\boldsymbol{x})$为凸函数，$h_j(\boldsymbol{x})$是仿射函数，假设有$\boldsymbol{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$满足KKT条件，则$\boldsymbol{x}^*$一定是极值点，即KKT条件式A.9是约束问题式A.7极值点的充分条件。

引申：

SVM的最优化问题为

$$
\min_{w,b} \frac{1}{2}\|w\|^2 \\
s.t.\ \ y_i(w^T x_i + b) \geq 1\ (i \in [m])
$$

可以看出，svm的目标函数、不等式约束条件是凸函数，所以KKT条件是极值点的充分必要条件。