AI数学基石——最优化方法篇

目录

- 最优化问题
  - 基本形式
  - 分类
- 线性规划问题
  - 实例
  - 线性规划问题的标准形式
  - 线性规划问题的求解
- 凸集和凸函数
  - 空间里的直线
  - 仿射集
  - 凸集
  - 超平面和半空间
  - 凸函数
- 凸优化问题
  - 定义

01 最优化问题

1.1 基本形式

简单地说，最优化问题就是在一定的约束条件下，求目标函数的最优值。
用数学语言来表述：

$$\begin{aligned} min \quad & f_0(x), x \in \mathbb{R}^n \\ s.t. \quad & f_i(x) \leq 0, i = 1, \cdots, m \\ & h_i(x) = 0, i = 1, \cdots, p \end{aligned}$$

在这里，

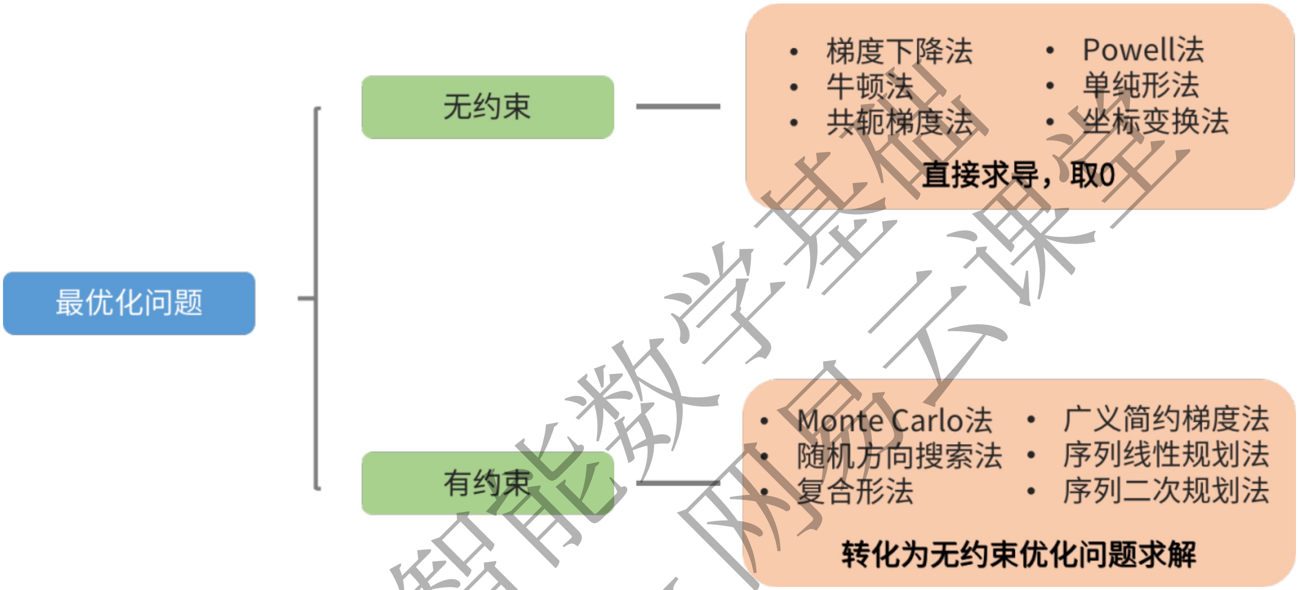
- $x \in \mathbb{R}^n$称为问题的**优化变量**
- 函数$f_o : \mathbb{R}^n \to \mathbb{R}$称为**目标函数**
- 不等式$f_i(x) \leq 0$称为**不等式约束**
- 方程组$h_i(x) = 0$称为**等式约束**。

变量，目标函数，约束条件是最优化问题的三个核心要素。

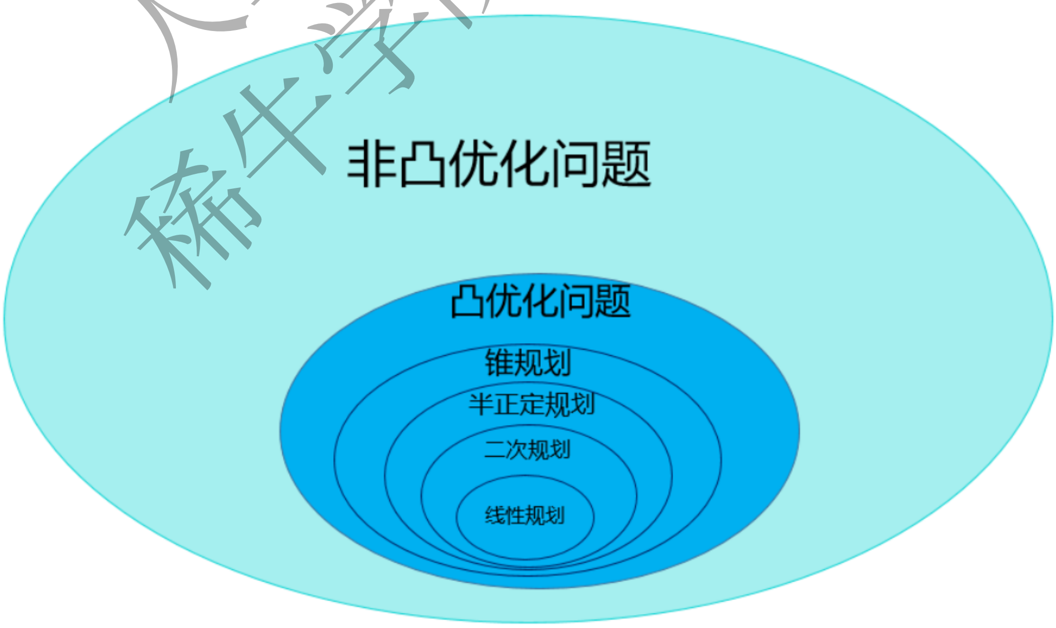
1.2 分类

最优化问题按照有无约束条件，可以分为有约束和无约束的优化问题，而有约束问题又可以分为含等式约束和不含等式约束的优化问题。不同的优化问题有不同的处理策略：

- **无约束的优化问题**：直接求导，使导数为0；
- **含等式约束的优化问题**：主要通过拉格朗日乘数法将含等式约束的优化问题转换成为无约束优化问题；
- **含有不等式约束的优化问题**：主要通过KKT条件将其转化成无约束优化问题。



在有约束的最优化问题里，还可以根据目标函数和约束函数是否为凸函数而分为**凸优化问题**和**非凸优化问题**：



02 线性规划问题

2.1 实例

设某厂计划生产A，B两种产品，它们都需要由机器，人工和材料三种资源共同完成。请问要如何安排生产计划才可以使得利润最大。

	A	B	资源限制
机器（工时）	5	2	120
人工	2	3	90
材料（公斤）	4	2	100
利润（元）	12	8	

解: 设A, B两种产品的生产数量分别为x_1, x_2, 产品总利润为z, 则有:

$$\begin{aligned} max\ z = 12x_1 + 8x_2 \\ s.t. \quad 5x_1 + 2x_2 \leq 120 \\ 2x_1 + 3x_2 \leq 90 \\ 4x_1 + 2x_2 \leq 100 \\ x_1, x_2 \geq 0 \end{aligned}$$

在线性规划中，优化变量也叫决策变量。

2.2 线性规划问题的标准形式

如果目标函数和约束函数都是线性函数，则称该类优化问题为**线性规划问题**。

其一般形式为

$$\max(\min) z = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

$$s.t.\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \geq (=, \leq) b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \geq (=, \leq) b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \geq (=, \leq) b_m \\ x_1, x_2, \cdots x_n \geq (\leq) 0 \end{cases}$$

而线性规划问题的标准形式则需要满足下面的条件：

1. 目标函数为最大值，即 $max\ z = c_1x_1 + c_2x_2 + \cdots + c_nx_n$

2. 约束条件为等式，即

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

3. 约束条件右端项非负，即$b_i \geq 0, i = 1, 2, \cdots, m$
4. 决策变量非负，即$x_1, x_2, \cdots, \geq 0$

简记为

$$\max z = \sum_{j=1}^{n} c_j x_j$$

$$s.t. \begin{cases} \sum_{j=1}^{n} a_{ij} x_j = b_i (i = 1, 2, \cdots, m) \\ x_j \geq 0 (j = 1, 2, \cdots, n) \end{cases}$$

线性规划的标准形式，也可以用矩阵来表示：

$$max\ z = CX$$

$$s.t. = \begin{cases} AX = b \\ X \geq 0 \end{cases}$$

其中

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, C = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}^T, b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{12} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

2.3 线性规划问题的求解

求解线性规划问题就是从满足约束条件的方程组中找出一个解，使目标函数达到最大值。主要方法有：

1. 图解法
2. 单纯形法
3. 计算机软件求解

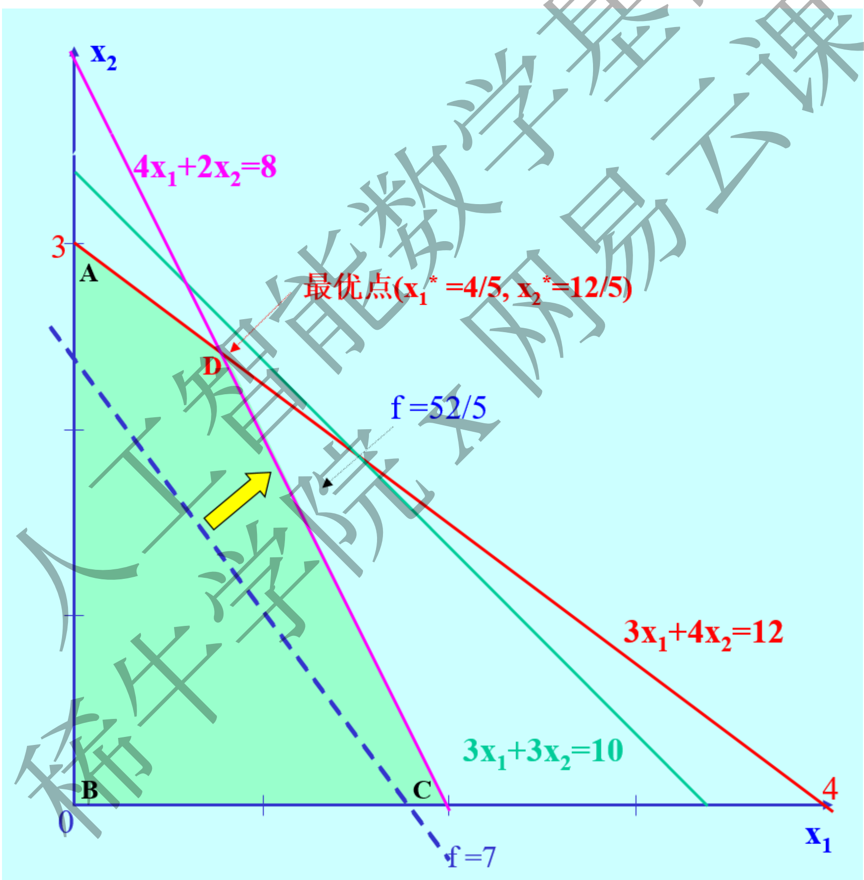
图解法

图解法主要适用于二维的或者三维的线性规划问题。它的基本步骤为：

1. 建立直角坐标系
2. 确定可行域（约束条件所构成的区域）
3. 图示目标函数
4. 最优解的确定 （可行域中使目标函数值达到最优的点）

比如，我们要求解下面这个最优化问题：

$$
\begin{gathered}
max\ z = 8x_1 + 6x_2 \\
s.t.\quad 3x_1 + 4x_2 \le 12 \\
4x_1 + 2x_2 \le 8 \\
3x_1 + 3x_2 \le 10 \\
x_1, x_2 \ge 0
\end{gathered}
$$

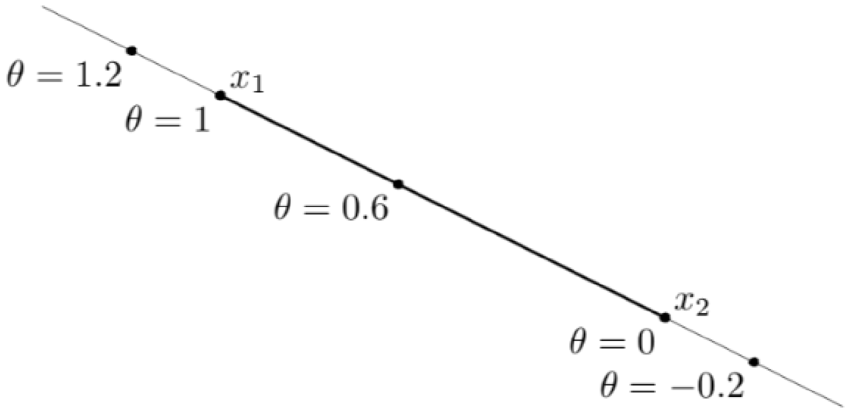


03 凸集和凸函数

2.1 空间里的直线

假设在$\mathbb{R}^n$内存在两个不同的点x_1和x_2，那么穿越两点的直线可以用下式来表示：

$$
y = \theta x_1 + (1-\theta) x_2, \quad \theta \in \mathbb{R}
$$



如果在指定原点和θ的情况下，y就表示通过x_1和x_2这两点的直线上的某一点。

2.2 仿射集

定义

> 对于任意的$x_1, x_2 \in C$及$\theta \in \mathbb{R}$, 有 $\theta x_1 + (1-\theta)x_2 \in C$, 那么集合$C$就是仿射的。

换句话说，有一个集合C, 如果集合内的任意两点构成的直线仍在集合C内，则称集合C为仿射集。

思考题

假设用$A\vec{x} = \vec{b}$表示的线性方程组有多个解，解的集合用C来表示，记作$C = \{x|Ax = b\}$，其中$A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, 问集合C是仿射集吗?

解

要证一个集合是否为仿射集，主要看集合中的两个元素的线性组合是不是在集合内。
对于任意两个点$x_1, x_2 \in C$，都有$Ax_1 = b, Ax_2 = b$,

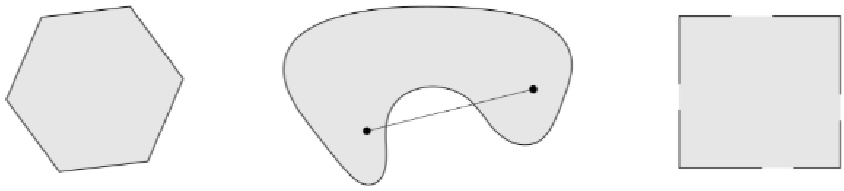
$$A(\theta x_1 + (1-\theta)x_2) = \theta Ax_1 + (1-\theta)Ax_2 = \theta b + (1-\theta)b = b$$

2.3 凸集

凸集和仿射集的定义很类似:

> 对于任意的$x_1, x_2 \in C$且$0 \le \theta \le 1$, 有 $\theta x_1 + (1-\theta)x_2 \in C$, 那么集合$C$就是凸集。

换句话说，对于一个集合$C \subseteq \mathbb{R}^n$，如果集合内的任意两点构成的线段仍在集合C内，则称集合C为凸集。由此可知，仿射集也是凸集。

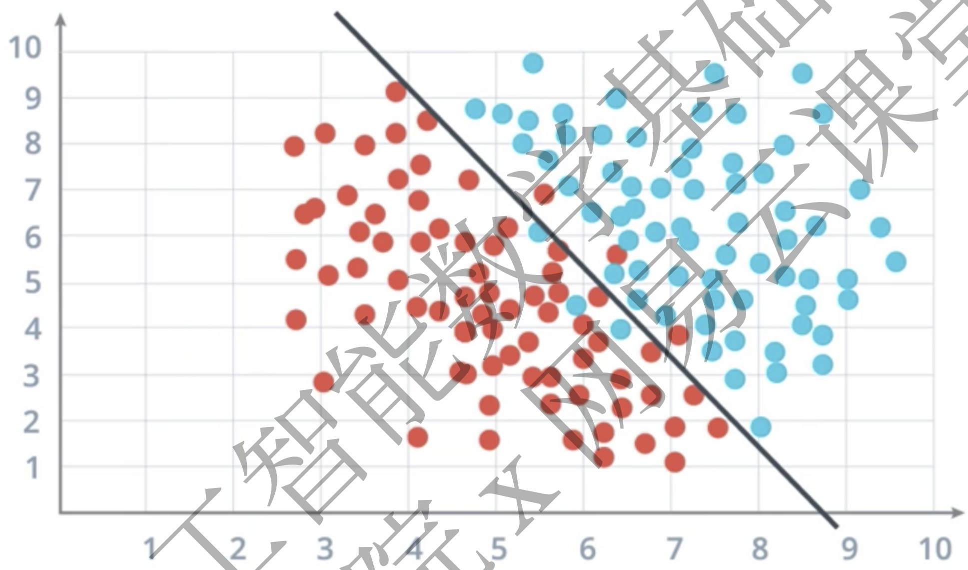


2.4 超平面和半空间

我们知道，二维空间里的直线在数学上，是以下线性方程

$$ax + by = d$$

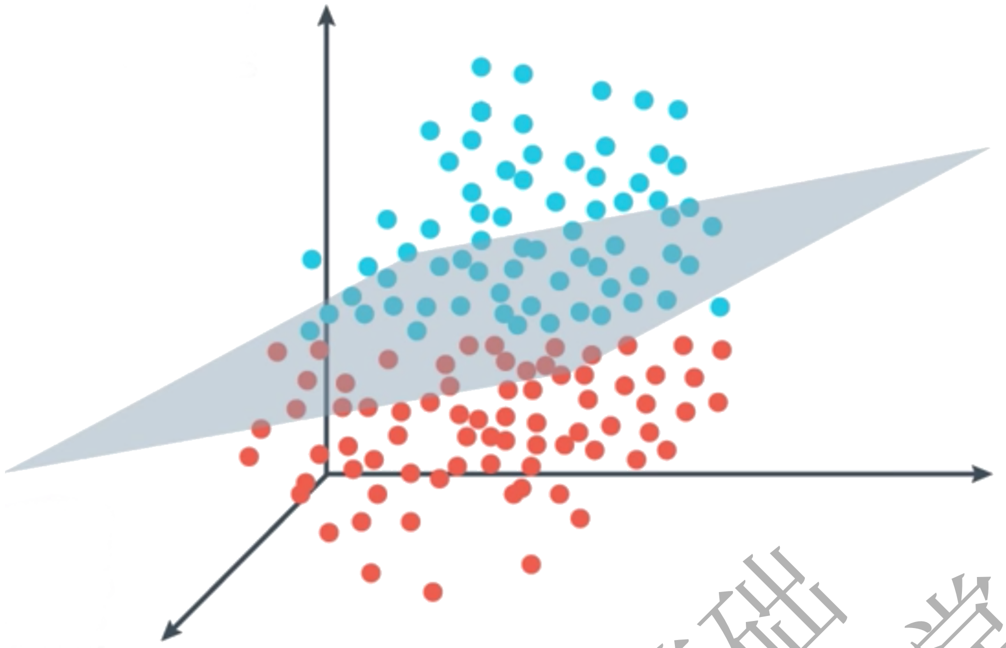
的所有解(x, y)在空间里形成的集合。



三维空间里的平面，则是以下线性方程

$$ax + by + cz = d$$

的所有解(x, y, z)在空间里形成的集合。



<以上图片来自于 Medium上的文章>

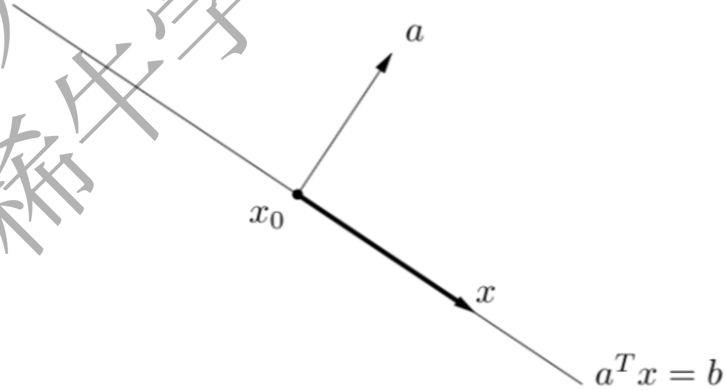
超平面为空间中的平面在n维空间里的推广，表示为

$$a_1x_1+\cdots+a_nx_n=d$$

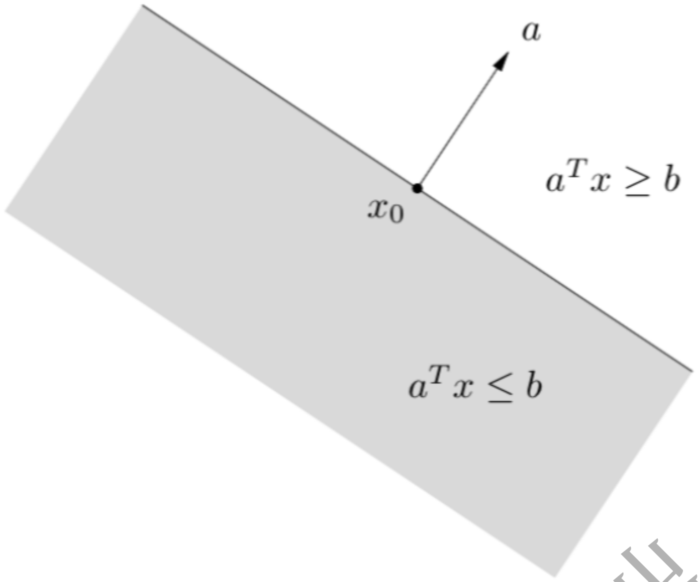
将$\mathbf{x}$和$\mathbf{a}$写成向量的形式，就可以得到超平面H的向量表达式:

$$H=\{\mathbf{x}\in\mathbb{R}^n|\mathbf{a}^T\mathbf{x}=d\}$$

对于平面中的两点$\mathbf{x}$和$\mathbf{x_0}$, 有$\mathbf{a}^T\mathbf{x}=d$和$\mathbf{a}^T\mathbf{x_0}=d$，将两式相减，得到$\mathbf{a}^T(\mathbf{x}-\mathbf{x_0})=0$ 于是，得到向量a与平面正交。超平面既是仿射集也是凸集。



超平面可以将空间分为两个半空间，表示为$\{\mathbf{x}|\mathbf{a}^T\mathbf{x}\leq b\}$或$\{\mathbf{x}|\mathbf{a}^T\mathbf{x}\geq b\}$，其中$a\neq\mathbf{0}$。半空间$\{\mathbf{x}|\mathbf{a}^T\mathbf{x}\geq b\}$表示的是方向为$a$的所有的点的集合，而半空间为$\{\mathbf{x}|\mathbf{a}^T\mathbf{x}\leq b\}$，其是方向为$-a$上的点的集合。因为，半空间任意点的线段在半空间内，而任意连线的直线不在半空间内，半空间是凸集，但不是仿射集。



2.5 凸函数

定义

> 函数$f : \mathbb{R}^n \to \mathbb{R}$的定义域$dom f$是凸集，并且$\forall x, y \in dom f$ 和$\forall\theta \quad 0 \leq \theta \leq 1$有
>
> $$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$
>
> 则称函数f是凸的。

凸函数可以理解为割线总在函数上方的函数。

04 凸优化问题

4.1 定义

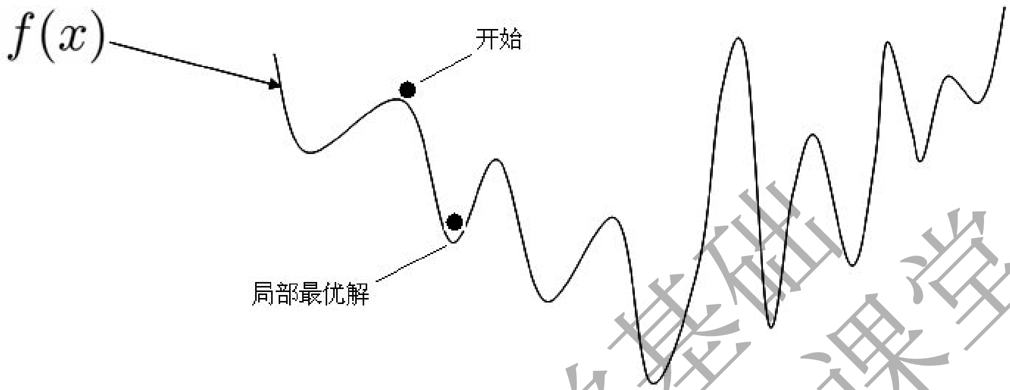
$$min \quad f_0(x), x \in \mathbb{R}^n$$

$$
\begin{aligned}
s.t. \quad & f_i(x) \leq 0, i = 1, \cdots, m \\
& h_i(x) = 0, i = 1, \cdots, p
\end{aligned}
$$

上述优化问题中，如果$f_i(x)$是凸函数，$h_j(x)$为仿射函数的优化问题称为凸优化问题。

凸优化问题的重要性质

1. 凸优化问题的可行域为凸集
2. 凸优化问题中局部最优解同时也是全局最优解。



凸优化之所以如此重要，是因为：

1. 应用非常广泛，机器学习中很多优化问题都要通过凸优化来求解；
2. 很多非凸优化问题，可以转化为凸优化问题来解决；
3. 凸优化问题目前具有成熟求解方法，而其他优化问题则未必。

本章要点总结

- 最优化问题就是在一定的约束条件下，求目标函数的最优值
- 目标函数和约束函数都是线性函数的优化问题就叫线性规划问题
- 目标函数和不等式约束是凸函数，等式约束是仿射函数的优化问题就叫凸优化问题
- 凸优化问题中局部最优解同时也是全局最优解
- 仿射集，凸集，超平面，半空间的定义以及几何含义

参考资料

- 《运筹学 第三版》 清华大学出版社
- 《Convex Optimization》by Stephen Boyd
- Longfei Han的博客
- CMU《Convex Optimization》的课件 by Ryan Tibshirani