第六讲、系统优化（上）

彭勤科
系统工程研究所
E-mail: qkpeng@xjtu.edu.cn
Tel: 82667964

2020年5月30日

系统优化概述

- 在系统工程学科中的地位
- 最优化模型举例
- 优化模型的特点
- 最优化问题的基本概念
- 凸集和凸函数
- 优化模型分类
- 典型的优化问题
- 算法设计与分析

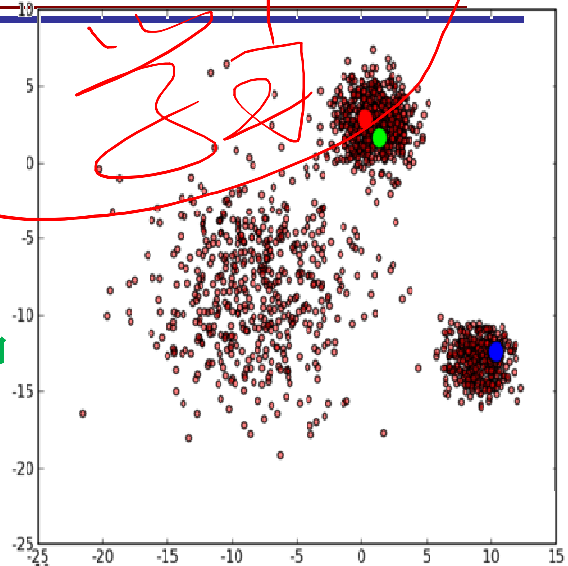
聚类-无监督学习

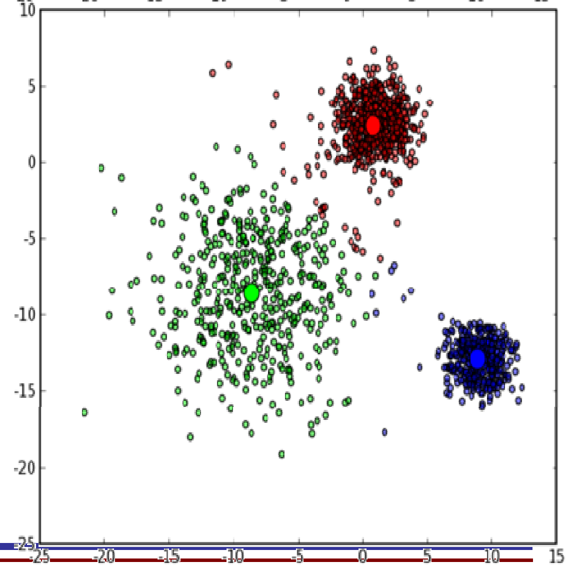
无导师学习

问题：把类标未知的数据按照一定目标分成不同的类。不同的目标函数产生了不同的聚类算法。以*K*-Means为例:

*确定K*个聚类中心，使得每个数据点到聚类中心的平方和最小。但是约束条件是，每一个点只能属于其中一个聚类中心。

$$\begin{cases} \min \quad \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma_{nk}\left\|x_n-\mu_k\right\|^2 \\ s.t. \quad \sum_{k=1}^{K}\gamma_{nk}=1, n\in\{1,2,\cdots,N\} \\ \qquad\quad \gamma_{nk}\in\{0,1\}. \end{cases}$$

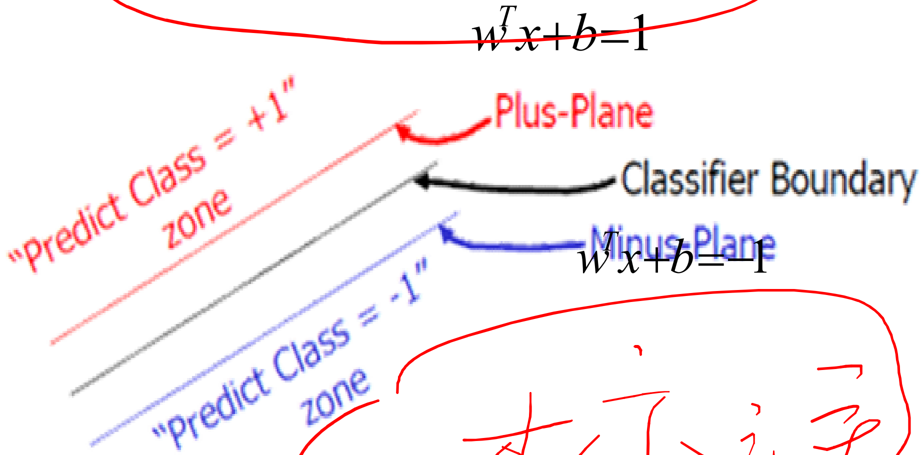




分类-有监督学习

有导师学习

问题：根据标注类别的数据建立分类模型，对未知类别的数据进行分类。不同建模方法导致不同的分类算法。以支持向量机(SVM)为例:



本方法

思路：构建最优的分类(超)平面，使得使得这个(超)平面尽可能将已知数据分开。但是约束条件是，已知的所属关系不能变。

$$
\begin{cases}
\min \ \frac{1}{2}\|w\|^2 \\
s.t. \quad y_n\left(w^T x_n + b\right) \geq 1, \ n \in \{1,2,\cdots,N\} \\
\qquad y_n \in \{-1,1\}, n \in \{1,2,\cdots,N\}
\end{cases}
$$

系统可靠性问题

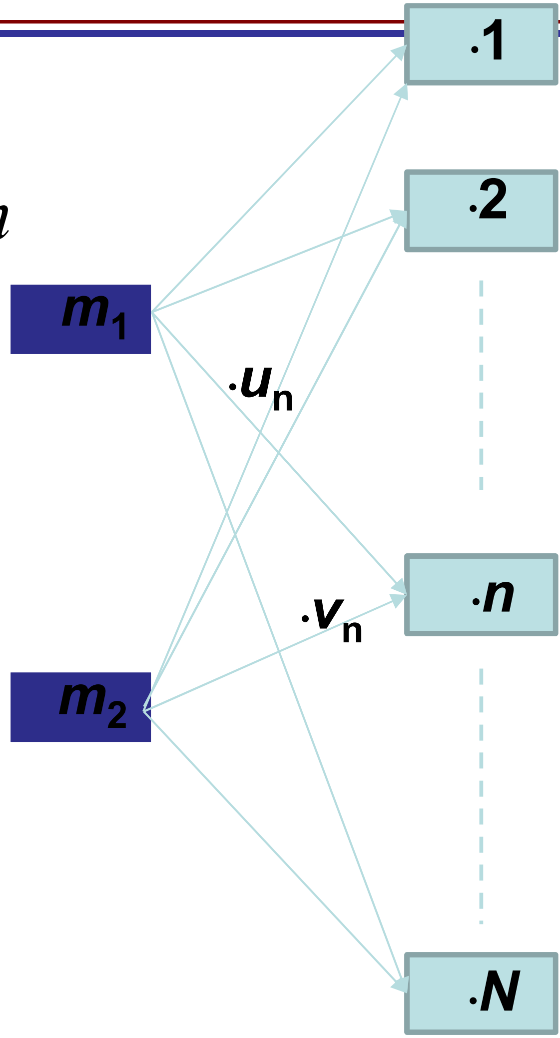
问题：对N个部件串联而成的系统，一部件出现故障时整个系统就不能正常工作，为每个部件配有备用件以提高可靠性，在总费用一定的条件下，如何配置备用件的数量才能使得系统的可靠性最高？

设第n个部件配备u_n个备件时，其正常工作的概率为$P_n(u_n)$，第n个备件的单价为c_n，总经费为C。

$$\begin{cases} \max & \prod_{n=1}^{N} P_n(u_n) \\ s.t. & \sum_{n=1}^{N} c_n u_n \le C, \\ & u_n \ge 0, n \in \{1,2,\cdots,N\}. \end{cases}$$

资源分配问题(两种资源)

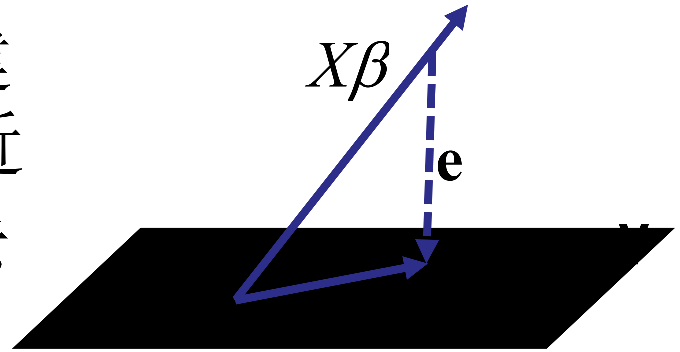
问题：资源总量分别是m_1 和m_2的两种资源，分别给N个用户，给第n个用户分配资源的总量分别为u_n和v_n时产生的收益为$g_n(u_n,v_n)$，如何分配资源使得收益最大？



$$
\begin{cases}
\max \quad \sum_{n=1}^{N} g_n(u_n, v_n) \\
s.t. \quad \sum_{n=1}^{N} u_n = m_1, \\
\qquad \sum_{n=1}^{N} v_n = m_2. \\
\qquad 0 \le u_n \le m_1, 0 \le v_n \le m_2
\end{cases}
$$

最小二乘参数估计

问题：根据若干样本（观测值），建立回归模型$Y=X\beta+\mathbf{e}$，使Y尽可能接近真实值。为此，可以以从不同角度去确定建立样本回归函数的准则函数，也就有了估计回归模型参数的多种方法。我们以最小二乘法则为例，残差平方和最小。(无约束优化模型)



最小二乘的几何解释：理论上响应变量$\boldsymbol{Y}$,与观测变量$\boldsymbol{X}$*系数$\boldsymbol{\beta}$是一致的，但在实际应用中噪声和测量误差的存在，使得$\boldsymbol{Y}$与$\boldsymbol{X\beta}$存在偏差**e**.

$$\min J=\sum_{n=1}^{N} e^2(n)=\mathbf{e}^T\mathbf{e}=(Y-X\beta)^T(Y-X\beta)$$

$$\left.\frac{\partial J}{\partial \beta}\right|_{\beta=\hat{\beta}}=-2X^TY+2X^TX\hat{\beta}=0 \Rightarrow \hat{\beta}=\left(X^TX\right)^{-1}X^TY$$

时间序列形态表示

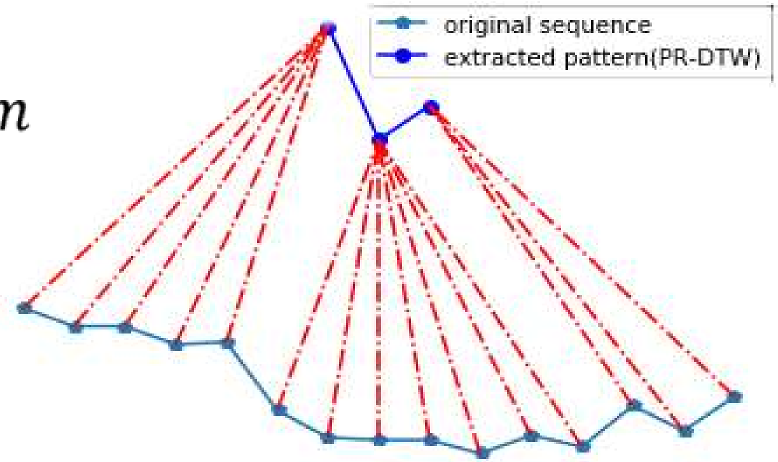
问题：对时间序列进行形态表示（时域上降维）

给定时间序列：$Q=\{q_1,q_2,\dots,q_n\}$,求 $Q^m=\{q_{t_1},q_{t_2},\dots,q_{t_m}\}$

优化模型：

$$\min_{Q^m} \mathrm{f}(Q^m) = \mathrm{DTW}(Q,Q^m)$$

$$\text{s.t.}\begin{cases} Q^m \subseteq Q \\ t_i < t_j \ ,\forall 1 \le i < j \le m \end{cases}$$



词语搬运距离

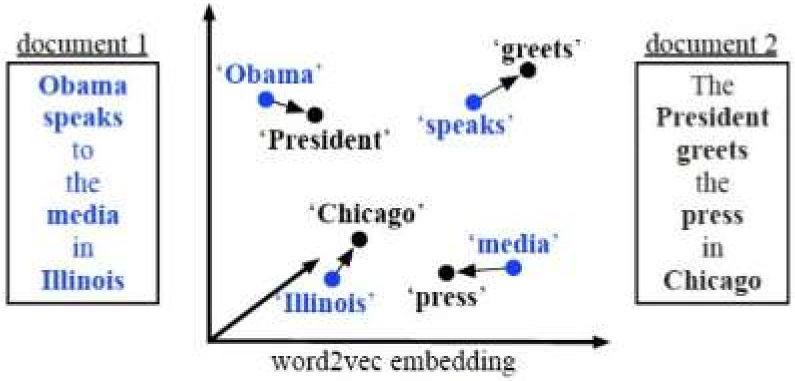
描述：对于计算两个文档之间的距离：用d和d'表示两个文档的$nBOW$向量，允许d中的任意一个词i转移到d'中的任意一个词j, 转移的代价是$c(i,j)$.定义一个转移矩阵$T \in R^{n\times n}$，其中T_{ij}表示单词i有多少的权重要转移到单词j。为了保证d全部转移到d'，必须满足$\sum_{i=1}^{n} T_{ij} = d_i$，$d'$同理

问题：如何找到一个单词匹配方式，使得累加带权重求和距离最小？这个最小距离就是最终两个文本的相似度

$$\min_{T\geq 0} \sum_{i,j=1}^{n} T_{ij} c(i,j)$$

$$s.t. \sum_{j=1}^{n} T_{ij} = d_i \quad \forall i \in \{1,\dots,n\}$$

$$\sum_{i=1}^{n} T_{ij} = d'_j \quad \forall j \in \{1,\dots,n\}$$



对抗生成网络

通过生成网络G和判别网络D不断博弈，进而使G学习到数据的分布，从而达到生成数据的目的。

利用一个判别器去衡量生成的数据和真实数据之间的差距（KL散度），然后不断优化判别器和生成器，最终的平衡点即纳什均衡点。

$$\theta^* = \arg\max_{\theta} \prod_{i=1}^{m} p_G(x^i;\theta) = \arg\max_{\theta} \log \prod_{i=1}^{m} p_G(x^i;\theta)$$

$$= \arg\max_{\theta} \sum_{i=1}^{m} \log p_G(x^i;\theta)$$

$$\approx \arg\max_{\theta} E_{x\sim p_{data}}[\log P_G(x;\theta)]$$

$$= \arg\max_{\theta} \int_x p_{data}(x)\log p_G(x;\theta)dx - \int_x p_{data}(x)\log p_{data}(x)dx$$

$$= \arg\min_{\theta} KL(p_{data} \| p_G)$$

安全攻防-双层优化

问题：安全攻防博弈过程是一类双层优化问题，其中，（上层）攻击者和（下层）防御者各自决策，交替作用，相互对抗，最终达到各自期望收益的最优值，实现纳什均衡。一般地，双层优化问题形如：

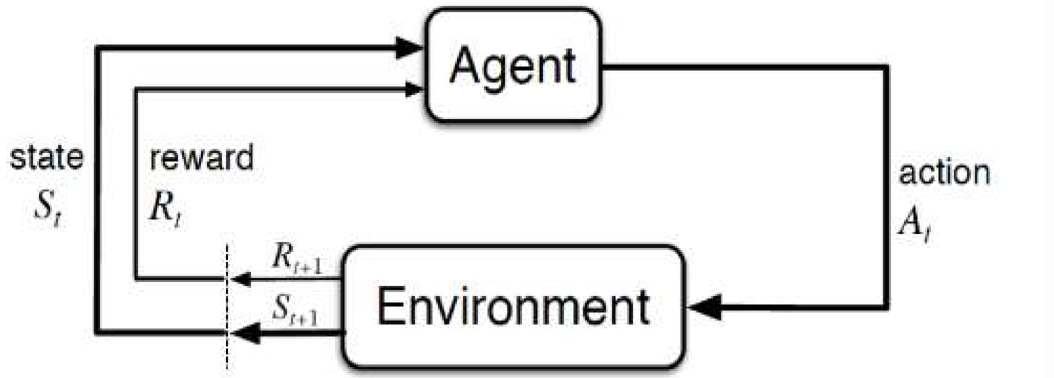
$$
\left\{
\begin{array}{l}
\max\limits_{\boldsymbol{x}\in\mathbb{X}} \mathrm{F}(\boldsymbol{x},\boldsymbol{y}) \\
\text{s.t.}\quad \mathbf{G}(\boldsymbol{x},\boldsymbol{y})\le 0 \\
\qquad\quad \mathbf{H}(\boldsymbol{x},\boldsymbol{y})=0 \\
\boldsymbol{y}\in\arg\min\limits_{\boldsymbol{y}\in\mathbb{Y}} \mathrm{f}(\boldsymbol{x},\boldsymbol{y}) \\
\qquad \text{s.t.}\quad \mathbf{g}(\boldsymbol{x},\boldsymbol{y})\le 0 \\
\qquad\qquad\ \ \mathbf{h}(\boldsymbol{x},\boldsymbol{y})=0
\end{array}
\right.
$$

上层优化：攻击者决策

下层优化：防御者决策

强化学习

强化学习研究的是智能体与环境之间交互的任务。让智能体通过试错，不断地学习在不同的状态下做出最优的动作，即获得最大的未来回报的期望值，从而学习出最优策略。



优化目标：

$$\pi_* = \underset{\pi}{argmax}\, V_\pi(s)$$

$V_\pi(s)$：状态s时，使用策略π时，未来回报的期望值。

π_*：最优策略

食谱问题

- 问题提出

设市场上可以买到n种不同的食品，第j种食品的价格为c_j，每种食品含有m种营养成份，每单位第j种食品含第i种营养成份为a_{ij}。假设每人每天对第i种营养成份的需要量不少于b_i，试确定在保证营养条件下的最经济食谱。

- 模型建立

设每人每天对n种食品的需要量为：$x_1, x_2, \cdots, x_n$， 则相应的伙食费为：

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

保证满足营养最小的需求约束方程为：

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \geq b_i \qquad i = 1,2,\cdots,m$$

$$x_j \geq 0 \qquad j = 1,2,\cdots,n$$

食谱问题（续）

则数学模型为： $\min \quad c_1x_1 + c_2x_2 + \cdots + c_nx_n$

$$
\begin{aligned}
s.t \quad & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \ge b_1 \\
& a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \ge b_2 \\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
& a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \ge b_m \qquad x_j \ge 0 \qquad j = 1,2,\cdots,n
\end{aligned}
$$

如果令 $x = (x_1, x_2, \cdots, x_n)^T \quad c = (c_1, c_2, \cdots, c_n) \quad A = (a_{ij})_{m\times n} \quad b = (b_1, b_2, \cdots, b_m)^T$

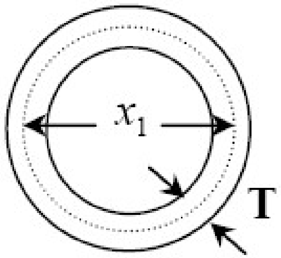
则上面的优化模型可以写成如下矩阵形式：

$$
\begin{aligned}
& \min \ cx \\
& s.t \ \ Ax \ge b, x \ge 0
\end{aligned}
$$

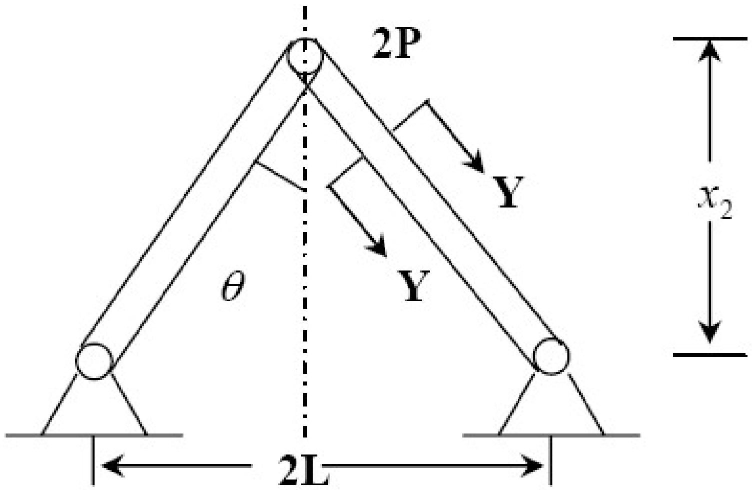
结构设计问题

- 问题提出

设两杆组成的对称衍架如下图所示，已知衍架的跨度为 **2L**，高度 x_2 的上限为 **H**，承受负荷 **2P**，钢管的厚度为 **T**，材料比重 ρ，纵向弹性系数为 **E** 且允许应力 σ_y，试确定钢管的平均内径 x_1 和衍架高度 x_2 并使衍架重量最小。



Y-Y 剖面



结构设计问题(续1)

- 模型建立

衔架的重量为： $G = 2\pi\rho T x_1 (L^2 + x_2^2)^{1/2}$

高度约束： $x_2 \le H$

钢管压应力约束：在负荷 **2P** 的作用下，钢管承受的压力 **F** 为

$$F = \frac{P}{\cos\theta} = \frac{P(L^2 + x_2^2)^{1/2}}{x_2}$$

钢管的横截面面积 **S** 为 $S = \pi T x_1$

因此，钢管的压应力为 $\sigma(x_1, x_2) = \dfrac{P(L^2 + x_2^2)^{1/2}}{\pi T x_1 x_2}$

所以对给定的允许应力 σ_y，要求 $\dfrac{P(L^2 + x_2^2)^{1/2}}{\pi T x_1 x_2} \le \sigma_y$

结构设计问题(续2)

临界应力约束：假设弹性模数为 E，则利用欧拉公式可以算出临界应力σ_l为

$$\sigma_l = \frac{\pi^2 E(x_1^2 + T^2)}{8(L^2 + x_2^2)}$$

因此，要求

$$\frac{P(L^2 + x_2^2)^{1/2}}{\pi T x_1 x_2} \le \frac{\pi^2 E(x_1^2 + T^2)}{8(L^2 + x_2^2)}$$

综上所述，衔架设计问题的最优化模型为

$$\begin{aligned}
\min \quad & 2\pi\rho T x_1 (L^2 + x_2^2)^{1/2} \\
s.t \quad & x_2 \le H \\
& \frac{P(L^2 + x_2^2)^{1/2}}{\pi T x_1 x_2} \le \sigma_y \\
& \frac{P(L^2 + x_2^2)^{1/2}}{\pi T x_1 x_2} \le \frac{\pi^2 E(x_1^2 + T^2)}{8(L^2 + x_2^2)} \\
& x_1 \ge 0, x_2 \ge 0
\end{aligned}$$

水果采购问题

- 问题提出

某单位举行联欢会要采购香蕉、苹果和葡萄，假设市场上香蕉、苹果和葡萄的价格分别为 **4.2** 元/公斤、**2.4** 元/公斤、**2.2** 元/公斤，如果要求采购水果的钱不超过 **300** 元、水果总重量不小于 **10** 公斤、香蕉和苹果总和不小于 **6** 公斤，试确定最好的采购方案。

- 模型建立

设 x_1、x_2、x_3 分别是购买香蕉、苹果和葡萄的重量（公斤），则购买水果的总金额 y_1 和采购水果的总重量 y_2 为

$$y_1 = 4.2x_1 + 2.4x_2 + 2.2x_3$$
$$y_2 = x_1 + x_2 + x_3$$

水果采购问题(续)

采购应该满足的约束为：

资金约束：$4.2x_1 + 2.4x_2 + 2.2x_3 \le 300$

重量约束：$x_1 + x_2 + x_3 \ge 10$， $x_1 + x_2 \ge 6$

非负约束：$x_i \ge 0$， $i = 1,2,3$

所以，优化模型为

$$
\begin{aligned}
&\min\ 4.2x_1 + 2.4x_2 + 2.2x_3 \\
&\max\ x_1 + x_2 + x_3 \\
&s.t\quad 4.2x_1 + 2.4x_2 + 2.2x_3 \le 300 \\
&\qquad\ x_1 + x_2 + x_3 \ge 10 \\
&\qquad\ x_1 + x_2 \ge 6 \\
&\qquad\ x_i \ge 0,\quad i = 1,2,3
\end{aligned}
$$

优化模型的特点

- 优化模型的要素：目标、优化变量、约束集合　$(F(x), x, S)$
  - 目标：用于度量优劣程度的指标（函数）

    (1) $\min \quad c_1x_1 + c_2x_2 + \cdots + c_nx_n$

    (2) $\min \quad 2\pi\rho T x_1 (L^2 + x_2^2)^{1/2}$

    (3) $\min \ 4.2x_1 + 2.4x_2 + 2.2x_3$
    $\max \ x_1 + x_2 + x_3$

  - 变量：表示优化的方案。

    (1) 每人每天对 n 种食品的需要量 $x_1, x_2, \cdots, x_n$

    (2) 钢管的平均厚度 x_1 和衔架高度 x_2

    (3) 购买香蕉、苹果和葡萄的重量 x_1、x_2、x_3

  - 约束集合：可行方案组成的集合。

    由一组线性或非线性方程的等式或不等式组成。

最优化问题的基本概念

在优化问题的三要素$(F(x),x,S)$中，满足约束条件（在约束集合中）的点x称为<u>**可行点**</u>，全体可行点组成的集合S称为<u>**可行集或可行域或约束集合**</u>。如果一个优化问题没有包含约束集合S，的可行集是整个空间，则称此问题为无约束问题。

设$f(x)$为目标函数，x为优化变量，S为可行集，极小化问题记为

$$\min f(x) \qquad (4.1)$$
$$s.t. \quad x \in S$$

定义 **4.1** 对极小化问题（**4.1**），$\overline{x} \in S$，若对$\forall x \in S$，有

$$f(x) \geq f(\overline{x})$$

则$\overline{x}$称为极小化问题（**4.1**）的<u>**最优解或极小解（整体最优解或全局最优解）**</u>

最优化问题的基本概念(续1)

定义 **4.2** 对极小化问题（**4.1**），$\bar{x} \in S$，若存在$\bar{x}$的ε **(>0)**邻域

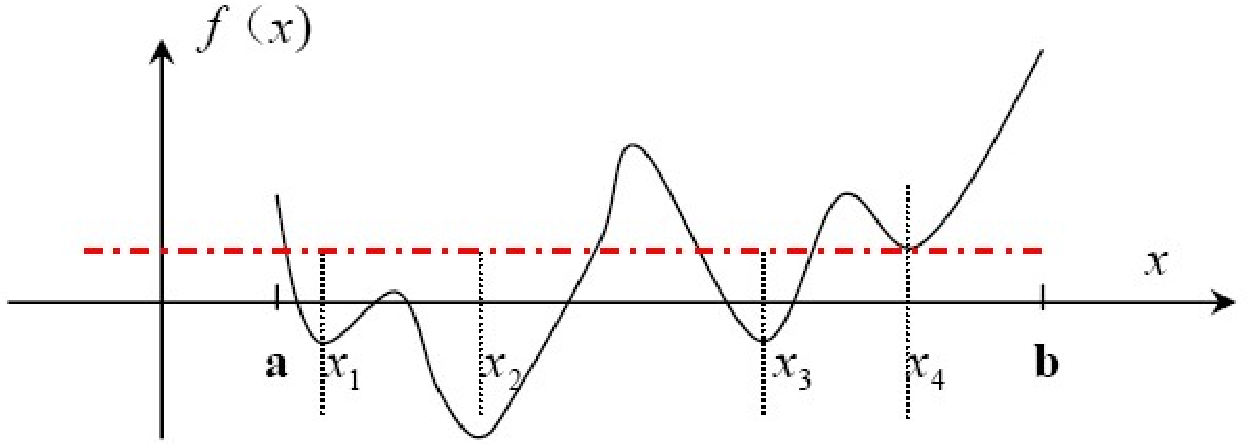
$$N_\varepsilon(\bar{x}) = \{x \mid \|x - \bar{x}\| < \varepsilon, \varepsilon > 0\}$$

对$\forall x \in S \cap N_\varepsilon(\bar{x})$，有

$$f(x) \geq f(\bar{x})$$

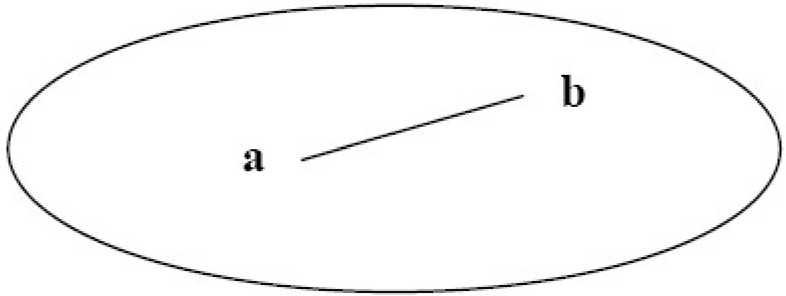
则$\bar{x}$称为极小化问题（**4.1**）的**局部最优解或局部极小解**。

注 **1**：**局部最优解在全局可能是很差的**，如下图$\{\min f(x) \mid a \leq x \leq b\}$



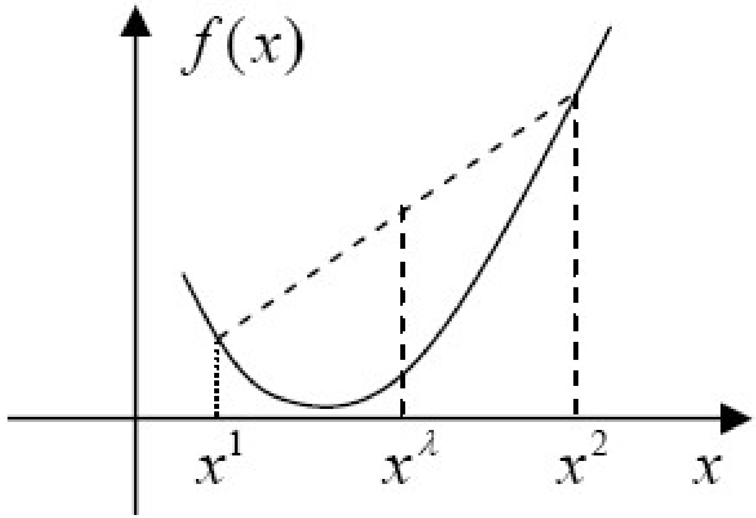
凸集和凸函数

- 凸集概念与性质



- 凸函数概念与性质

$$f(\lambda x^1+(1-\lambda)x^2)\le \lambda f(x^1)+(1-\lambda)f(x^2)$$



凸函数的几何解说（弓弦在弓背上）

凸的优化问题的性质

- 凸优化问题

  $f(x)$是凸函数，S是凸集合

- 凸优化问题最优解的性质
  - 凸优化问题如果存在局部最优解，则该局部最优解必然为全局最优解。
  - 对凸优化问题，全局最优解容易计算。

- 非凸优化问题最优解的性质
  - 局部最优解一般不是全局最优解。
  - 对非凸优化问题，由于大多数优化算法只能保证得到局部最优解，所以全局最优解的计算需要在算法设计上进行特殊的考虑。

优化问题的分类

- 目标函数的数量
  - 单目标优化
  - 多目标优化
- 目标函数和约束函数的性质
  - 线性优化
  - 非线性优化
- 优化变量对时间的依赖性
  - 静态优化
  - 动态优化
- 优化变量定义域
  - 连续优化
  - 离散优化

典型的优化问题 - 线性规划

- 标准形式

$$
\begin{aligned}
\min \quad & \sum_{j=1}^{n} c_j x_j \\
s.t. \quad & \sum_{j=1}^{n} a_{ij} x_j = b_i \quad i = 1,2,\cdots,m \qquad \textbf{(7.1)} \\
& x_j \ge 0, \quad j = 1,2,\cdots,n
\end{aligned}
$$

矩阵标准形式

$$
\begin{aligned}
\min \quad & cx \\
s.t. \quad & Ax = b \\
& x \ge 0
\end{aligned}
$$

其中$A=(a_{ij})_{m\times n}$为$m\times n$矩阵，x是n维列向量，c是n维行向量，b是m维列向量。

线性规划问题的基本性质

- 基本概念：可行解、基本可行解、最优基本可行解，对偶定理，灵敏度分析，•••。
- 基本性质：线性规划问题的最优解必然在基本可行解（一般为有限个上）达到。

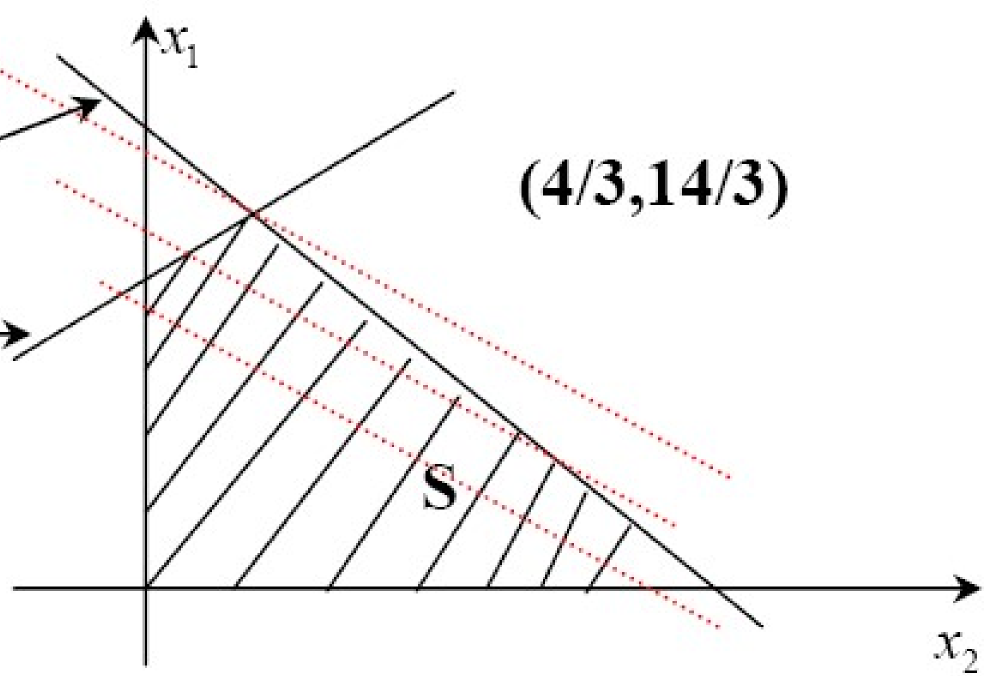
几何解释

$$\min -x_1 - 3x_2$$

$$s.t. \quad x_1 + x_2 \le 6$$

$$-x_1 + 2x_2 \le 8$$

$$x_1 \ge 0, x_2 \ge 0$$



线性规划的基本求解算法

- 标准单纯形法
- 修正单纯形法
- 对偶单纯形法
- 分解算法
- Karmarkar算法
- 内点算法

典型的优化问题 – 非线性规划

- 标准形式

无约束优化问题:

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & x \in E^n \end{aligned} \qquad \textbf{(7.2)}$$

有约束优化问题:

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & g_i(x) \geq 0 \quad i = 1,2,\cdots,m \\ & h_j(x) = 0 \quad j = 1,2,\cdots,l \end{aligned} \qquad \textbf{(7.3)}$$

非线性规划的基本性质

- 基本定义

对目标函数 $f(x)$，称 $\nabla f(x)=(\frac{\partial f(x)}{\partial x_1},\frac{\partial f(x)}{\partial x_2},\cdots,\frac{\partial f(x)}{\partial x_n})^T$ 为 $f(x)$ 在 x 处的梯度，

称矩阵

$$\nabla^2 f(x)=\begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \cdots & \cdots & \cdots & \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

为 $f(x)$ 在 x 处的 **Hessian** 矩阵。

无约束优化最优性条件

性质 **7.1** （一阶必要条件）设函数 $f(x)$ 在 $\bar{x}$ 可微，若 $\bar{x}$ 是问题（**7.2**）的局部最优解，则梯度 $\nabla f(\bar{x})=0$

性质 **7.2** （二阶必要条件）设函数 $f(x)$ 在 $\bar{x}$ 二次可微，若 $\bar{x}$ 是问题（**7.2**）的局部最优解，则梯度 $\nabla f(\bar{x})=0$，且 **Hessian** 矩阵 $\nabla^2 f(\bar{x})$ 是半正定的。

性质 **7.3** （二阶充分条件）设函数 $f(x)$ 在 $\bar{x}$ 二次可微，若 $\nabla f(\bar{x})=0$，且 **Hessian** 矩阵 $\nabla^2 f(\bar{x})$ 是正定的，则 $\bar{x}$ 是局部极小解。

有约束优化最优性条件

对问题（7.3），定义 Lagrange 函数

$$L(x,w,v)=f(x)-\sum_{i=1}^{m}\omega_i g_i(x)-\sum_{j=1}^{l}v_j h_j(x)$$

性质 7.4 （一阶必要条件，K-T 必要条件）在问题（7.3）中，$\bar{x}$是可行解，$I=\{i\mid g_i(\bar{x})=0\}$，$f(x)$和$g_i(x)(i\in I)$在$\bar{x}$可微，$g_i(x)(i\notin I)$在$\bar{x}$连续，$h_j(x)=0$（$j=1,2,\cdots,l$）在$\bar{x}$连续可微，向量集

$$\{\nabla g_i(\bar{x}),\nabla h_j(\bar{x})\mid i\in I,j=1,\cdots,l\}$$

线性无关，若$\bar{x}$是问题（7.3）的局部最优解，则存在$\omega_i(i=1,\cdots,m)$和$v_j(j=1,\cdots,l)$，使得

$$\nabla f(\bar{x})-\sum_{i=1}^{m}\omega_i\nabla g_i\bar{x})-\sum_{j=1}^{l}v_j\nabla h_j(\bar{x})=0$$

$$\omega_i g_i(\bar{x})=0,i=1,\cdots,m$$

$$\omega_i\geq 0,i=1,\cdots,m$$

非线性规划问题的求解算法

- 无约束优化问题求解算法
  - 一维搜索算法：
    - **0.618**法（黄金分割法）
    - **Fibonacci**法
    - 函数逼近法（牛顿法）…。
  - 最速下降法
  - 牛顿法
  - 共轭梯度法
  - 拟牛顿法
  - ●●● ●●● ●●●
- 有约束优化问题求解算法
  - 可行方向法
  - 梯度投影法
  - 割平面法
  - 惩罚函数法
  - ●●● ●●● ●●●

典型的优化问题 – 多目标优化

- 标准形式

$$\min \begin{bmatrix} f_1(x) \\ f_2(x) \\ \cdots \\ f_k(x) \end{bmatrix}$$

$$s.t.\quad g_i(x) \geq 0 \quad i = 1,2,\cdots,m$$

$$h_j(x) = 0 \quad j = 1,2,\cdots,l$$

- 基本性质：最优解→非劣解→满意解
- 求解算法
  - 加权法
  - 效用函数法
  - 交互式方法
  - 决策支持系统

算法设计与分析简介

- 算法举例

最速下降法

$$x^{k+1} = x^k - \lambda_k \nabla f(x^k)$$

$$\lambda_k = \min_{\lambda \geq 0} f(x^k - \lambda \nabla f(x^k))$$

牛顿法

$$x^{k+1} = x^k - \nabla^2 f(x^k)^{-1} \nabla f(x^k)$$

算法举例(续)

共轭梯度法

（1）给定初始点 $x^{(1)}$，置 $k=1$；

（2）计算梯度 $g_k=\nabla f(x^{(k)})$，若 $\|g_k\|=0$，则停止计算，得点 $\bar{x}=x^{(k)}$；否则，进行下一步；

（3）构造搜索方向，令 $d^{(k)}=-g_k+\beta_{k-1}d^{(k-1)}$
其中当 $k=1$ 时，取 $\beta_{k-1}=0$，当 $k>1$ 时，利用下式计算 β_{k-1}

$$\beta_{k-1}=\frac{\|g_k\|^2}{\|g_{k-1}\|^2}$$

（4）令 $x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)}$ 其中 $\lambda_k=-\dfrac{g_k^T d^{(k)}}{d^{(k)T}Ad^{(k)}}$

（5）若 $k=n$，则令 $\bar{x}=x^{(k+1)}$，否则，置 $k:=k+1$，返回（2）。

算法基本概念

根据上面实例分析可以看出，算法包含三个要素和环节：

(1) 初始点 $x^{(0)}$ 或 $x^{(1)}$；

(2) 如何从点 $x^{(k)}$ 得到点 $x^{(k+1)}$，往往 $x^{(k+1)}$ 是 $x^{(k)}$ 的函数或多个函数组成的复合函数

(3) 算法终止条件，如 $\|g_k\| = 0$ 。

一般而言，优化算法往往是迭代算法：

$$x^0, x^1, x^2, \cdots, x^k, x^{k+1}, \cdots$$

其中 $$x^{k+1} = A(x^k)$$

或更一般 $$x^{k+1} \in A(x^k)$$

算法基本概念(续)

定义 8.1 （**迭代算法**）算法 **A** 是一个如下形式的映射：

$$x^{k+1} = A(x^k) \text{或} x^{k+1} \in A(x^k)$$

例子 **1**：最速下降法是点集映射。

$$x^{k+1} = x^k - \lambda_k \nabla f(x^k) \qquad \lambda_k = \min_{\lambda \geq 0} f(x^k - \lambda \nabla f(x^k))$$

例子 **2**：牛顿法是一个函数。

$$x^{k+1} = x^k - \nabla^2 f(x^k)^{-1} \nabla f(x^k)$$

例子 **3**：而共轭梯度法是一个复合函数。

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)} \qquad d^{(k)} = -g_k + \beta_{k-1} d^{(k-1)}$$

$$\beta_{k-1} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \qquad \lambda_k = -\frac{g_k^T d^{(k)}}{d^{(k)T} A d^{(k)}}$$

算法设计

(1) 映射 $A(x)$

算法设计的重点，前面列举了的单纯行算法、 最速下降法、牛顿法和共轭梯度法，后续将看详细介绍相关算法。

(2) 算法终止条件

当梯度充分小时 $|\nabla f(x^{(k)})| < \varepsilon$

当优化变量变化很小时 $\left\|x^{(k+1)} - x^{(k)}\right\| < \varepsilon$ $\dfrac{\left\|x^{(k+1)} - x^{(k)}\right\|}{\left\|x^{(k)}\right\|} < \varepsilon$

当目标函数值增长非常小时

$|f(x^{(k)}) - f(x^{(k+1)})| < \varepsilon$ $\dfrac{|f(x^{(k)}) - f(x^{(k+1)})|}{|f(x^{(k)})|} < \varepsilon$

算法分析

(1) 收敛性分析

序列$\{x^{(k)}:k=1,2,...,\}$的收敛性分析，高等数学知识。

(2) 收敛速度分析

定义 5：设序列$\{r^{(k)}\}$收敛于r^*，满足

$$0 \le \overline{\lim_{k\to\infty}} \frac{\left\| r^{(k+1)} - r^* \right\|}{\left\| r^{(k)} - r^* \right\|^p} = \beta < \infty$$

的非负数p的上确界称为序列$\{r^{(k)}\}$的收敛阶（或级）。若序列的收敛阶为p，则称该序列是p阶收敛的。

当阶$p=1$且$\beta<1$，则称序列是以收敛比β线性收敛的；若$p>1$或$p=1,\beta=0$，则称序列是超线性收敛的。

收敛速度分析举例

例子 **1**：序列$\{a^k\},0<a<1$是以收敛比a线性收敛的；

例子 **2**：序列$\{a^{(2k)}\},0<|a|<1$是 **2** 阶收敛的；

例子 **3**：最速下降法是线性收敛的；

例子 **4**：牛顿法是 **2** 阶收敛的。

优化理论研究的主要内容、关键问题、主要方法

- 关键问题
  - 规模
  - 非线性
  - **NP**问题
- 主要方法
  - 大系统
  - 并行
  - 随机算法
  - 进化算法等

本讲需要掌握的重点知识

- 优化模型要素（特别是优化模型与决策模型、评价模型的区别），建立优化模型的方法。