第一章：介绍

（1）课程介绍和典型例子

曹语

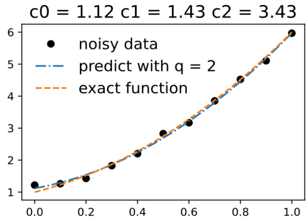
课程主页：https://yucaoyc.github.io/math3806

目录

优化问题无处不在

火车的时刻规划

实验数据处理



自动驾驶

等等问题都涉及复杂的优化过程

优化问题的基本框架

我们可选择的选项：**优化变量/备选解** $\boldsymbol{x} \in \mathbb{R}^n$（或者 $\boldsymbol{x} \in \{1, 2, 3, \cdots, n\}$）

更一般的，可行解的集合被标记为 Ω

我们希望优化的对象：**目标函数 (或费用函数)** $\boldsymbol{x} \mapsto f_0(\boldsymbol{x})$

优化问题：

$$\min_{\boldsymbol{x} \in \Omega} \; f_0(\boldsymbol{x})$$

如果找到 $\boldsymbol{x}^\star$ 满足对于任何的 $\boldsymbol{x}$ 皆有 $f_0(\boldsymbol{x}) \geq f_0(\boldsymbol{x}^*)$，该 $\boldsymbol{x}^\star$ 被称为**最优解**

优化问题的标准形式

$$
\begin{array}{ll}
\text{minimize} & f_0(\boldsymbol{x}) \\
\text{subject to} & \begin{cases} f_i(\boldsymbol{x}) \leqslant 0, & i = 1, \cdots, m \\ h_i(\boldsymbol{x}) = 0, & i = 1, \cdots, p \end{cases} \qquad \text{（刻画可行解的集合 } \Omega\text{）}
\end{array}
$$

惯例： 设不等式和等式约束的右端为零

问题：

- 若原始优化的形式是 $\max f_0(\boldsymbol{x})$，如何变成标准形式?
- 若原始优化的形式是 $\ell \leq f_1(\boldsymbol{x}) \leq r$，如何变成标准形式?

课程介绍

主要介绍凸优化的理论以及重要的优化算法

(I) 凸优化的数学基础

(II) 无约束优化问题

(III) 约束优化问题

基本流程：

具体问题 $\Longrightarrow$ 数学形式 $\Longrightarrow$ 数学理论（即重要定理和一些性质）

$\Longrightarrow$ 回归应用问题 $\Longrightarrow$ 写代码来实践

将使用 Canvas 作为平台：

考核形式： 见课内介绍，以及课程大纲（见 Canvas）

课程资料： 课内 PPT 和代码可见课程主页（Canvas 内有链接）；
作业需要在 Canvas 提交

课本：

- Boyd S, Vandenberghe L. Convex Optimization. Cambridge: Cambridge University Press. 【电子版课本见课程网站的链接】
- （中译本）Stephen Boyd, Lieven Vandenberghe 著王书宁，许鋆，黄晓霖译. 凸优化. 清华大学出版社

联系方式： 请直接通过 Canvas 发消息/或者发邮件至yucao@sjtu.edu.cn

目录

上确界和下确界

优化任务 $\min_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x})$ 本质上是找到某个 $\boldsymbol{x}$ 使得 $f_0(\boldsymbol{x}) \leq f_0(\boldsymbol{y}), \forall \boldsymbol{y} \in \Omega$

问题：是否可能无法找到 x 呢?

例子： $\Omega = (0, 1]$，$f_0(\boldsymbol{x}) = \boldsymbol{x}$。此时我们可以无限让 $\boldsymbol{x}$ 趋紧 0，但是由于 0 不是可行解，因此该优化问题其实无解。

定义 (上确界和下确界)

- 对于任意的一个集合 $C \subset \mathbb{R}$，我们定义其**上确界（标记为 $\sup C$）**为集合 C 的所有上界中最小的值。若 C 有上界，根据上确界性质[1]，$\sup C$ 存在。

- 对于任意的一个集合 $C \subset \mathbb{R}$，我们定义其**下确界（标记为 $\inf C$）**为集合 C 的所有下界中最大的值。若 C 有下界，则 $\inf C$ 存在。

例子： 对于集合 $C = (0, 1]$，$\sup C = 1$, $\inf C = 0$

练习： 请写出集合 $C = \{1\} \cup [2, 3] \cup (4, 5)$ 的上/下确界

[1] Least-upper-bound property

对于上述问题，更一般性的符号为

$$\sup_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x}) \qquad \text{或} \qquad \inf_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x})$$

好处：我们永远可以计算这个数值（有时可能是 $\pm\infty$）

- 若 $\inf_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x}) = -\infty$，或不存在 $\boldsymbol{x} \in \Omega$ 满足 $f_0(\boldsymbol{x}) = \inf_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x})$，则该优化问题无可行解。
- 若存在 $\boldsymbol{x} \in \Omega$ 满足 $f_0(\boldsymbol{x}) = \inf_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x})$，该问题等价于 $\min_{\boldsymbol{x}\in\Omega} f_0(\boldsymbol{x})$。即严格意义下，我们使用 $\min$ 的时候，我们（一般而言）已经假设/验证该问题可以有解。
- 对于 $\sup$ 的情况类似。

目录

最小二乘

最小二乘至少可以追溯到 Gauss 在 1820s 的论文

假设我们有数据点 (x_i, b_i), $i = 0, 1, 2, \cdots, \ell$，我们希望找到一个函数 S 使得误差 $\delta_i = S(x_i) - b_i$ 最小化

问题： 误差点 δ_i 可以写成一个向量的形式，对于向量的大小，欧式距离如何来描述?

(A) $\sum_{i=1}^{\ell} \delta_i$

(B) $\sum_{i=1}^{\ell} |\delta_i|$

(C) $\sum_{i=1}^{\ell} (\delta_i)^2$

如果我们使用欧式距离来描述:

$$\min_{S} \sum_{i=0}^{\ell} \left(S(x_i) - b_i\right)^2$$

函数 S 太过于抽象，我们也许可以考虑使用多项式

考虑 $S(x) = c_0 + c_1 x + \cdots + c_q x^q$

$$\min_{c_0, c_1, \cdots, c_q} \sum_{i=0}^{\ell} \left[\sum_{j=0}^{q} c_j (x_i)^j - b_i\right]^2 = f_0(c_0, c_1, \cdots, c_q)$$

此处参数 $c_0, c_1, \cdots, c_q$ 是我们希望找寻的，它们的所有可能性构成了**备选解**

$$\frac{\partial f_0}{\partial c_k} = 2\sum_{i=0}^{\ell}\left(\sum_{j=0}^{q} c_j\,(x_i)^j \;-\; b_i\right)\; x_i^k = 0 \tag{1}$$

我们引入一些符号来简化一下：$\boldsymbol{A}_{i,j} = x_i^j$

$$\sum_{i=0}^{\ell}\sum_{j=0}^{q}\boldsymbol{A}_{i,j}\boldsymbol{A}_{i,k}c_j = \sum_{i=0}^{\ell}\boldsymbol{A}_{i,k}b_i$$

我们可以更简洁的把方程写成矩阵的形式：

$$\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{c} = \boldsymbol{A}^{\mathsf{T}}\boldsymbol{b} \qquad\Longrightarrow\qquad \boldsymbol{c} = \left(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{b} \tag{2}$$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{00} & \boldsymbol{A}_{01} & \cdots & \boldsymbol{A}_{0q} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{A}_{\ell 0} & \boldsymbol{A}_{\ell 1} & \cdots & \boldsymbol{A}_{\ell q} \end{bmatrix} \qquad \boldsymbol{c} = \begin{bmatrix} c_0 \\ \vdots \\ c_q \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_0 \\ \vdots \\ b_\ell \end{bmatrix}$$

目标函数还可以更简洁地写为

$$
\begin{aligned}
f_0(c_0, c_1, \cdots, c_q) &= \sum_{i=0}^{\ell} \left[\sum_{j=0}^{q} c_j (x_i)^j - b_i \right]^2 \\
&= \sum_{i=0}^{\ell} \left(\sum_{j=0}^{q} \boldsymbol{A}_{i,j} c_j - b_i \right)^2 \\
&= \langle \boldsymbol{Ac} - \boldsymbol{b}, \boldsymbol{Ac} - \boldsymbol{b} \rangle \\
&= \| \boldsymbol{Ac} - \boldsymbol{b} \|^2
\end{aligned}
$$

一般的最小二乘问题

$$\min_{\boldsymbol{c}} \; f_0(\boldsymbol{c}) = \|\boldsymbol{A}\boldsymbol{c} - \boldsymbol{b}\|^2$$

其中 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 是矩阵，$\boldsymbol{c} \in \mathbb{R}^n$ 是优化变量，$\boldsymbol{b} \in \mathbb{R}^m$ 是给定的数据

方程的解为

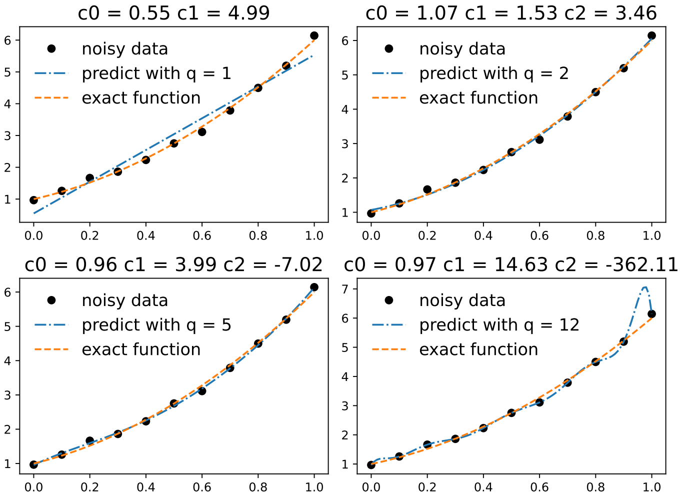
$$\boldsymbol{c} = \left(\boldsymbol{A}^\top \boldsymbol{A}\right)^{-1} \boldsymbol{A}^\top \boldsymbol{b}$$

讨论

问题：最小二乘拟合中维度 m 需要满足 $m \geq n$，即数据量需要不小于参数的个数。从直觉上为什么需要？比如我们有 2 个数据点 $(m = 2)$，使用二次函数去拟合 $(n = 3)$，会出现什么问题？

测试例子

真解为 $S(x) = 3x^2 + 2x + 1$；我们基于真值对数据做了扰动



目录

例子

最小二乘拟合问题中，备选解在空间 $\mathbb{R}^n$ 中没有任何的限制，而现实中，可选的解可能还具有其它约束 $\Longrightarrow$ 我们将以**线性规划问题**为例介绍

问题：假如你有 10 元钱作为本金，想购买可乐与冰棍并运到山上卖给游客来挣钱。倘若购买 2 元一听的可乐，每瓶可乐可以挣 1 元钱；或者购买 3 元一根的冰棍，每个冰棍可以挣 3 元钱。你应该如何选择策略呢?

例子

最小二乘拟合问题中，备选解在空间 $\mathbb{R}^n$ 中没有任何的限制，而现实中，可选的解可能还具有其它约束 $\Longrightarrow$ 我们将以**线性规划问题**为例介绍

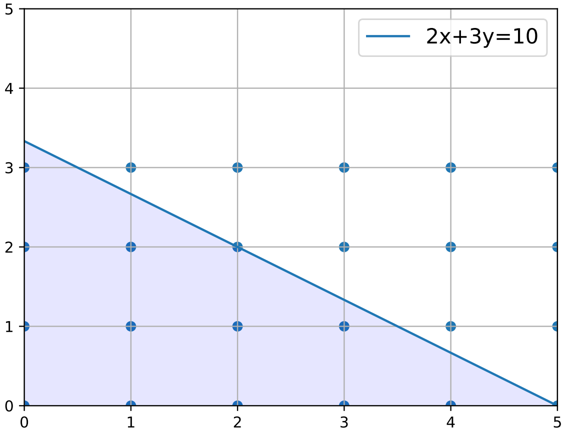
问题：假如你有 10 元钱作为本金，想购买可乐与冰棍并运到山上卖给游客来挣钱。倘若购买 2 元一听的可乐，每瓶可乐可以挣 1 元钱；或者购买 3 元一根的冰棍，每个冰棍可以挣 3 元钱。你应该如何选择策略呢?

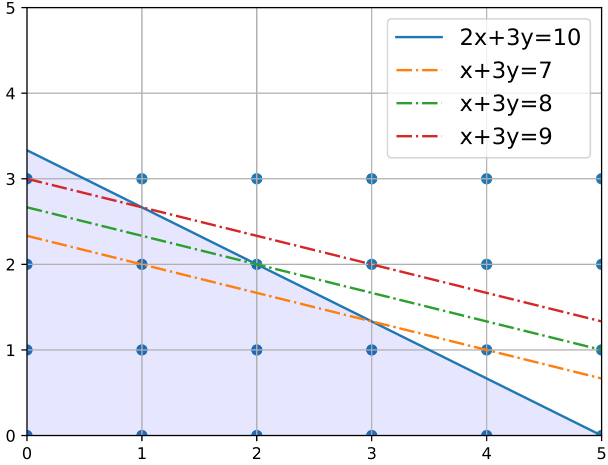
倘若你购买 x 听可乐，与 y 根冰棍，则

$$
\begin{aligned}
\text{优化目标:} \qquad & \max\ x + 3y; \\
\text{限制条件:} \qquad & 2x + 3y \leq 10.
\end{aligned}
$$

限制条件

由于 x, y 都是整数，我们仅需考虑格点。该问题的可行解是什么？





最优策略是购买 3 根冰棍，即使没有用完所有的钱；最高可挣 9 元。

（实数空间的）线性规划的一般形式

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{G}\boldsymbol{x} \preceq \boldsymbol{h} \qquad (\text{不等式约束}) \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \qquad (\text{等式约束}) \end{aligned}$$

- 其中 $\boldsymbol{c}, \boldsymbol{x} \in \mathbb{R}^n$，$\boldsymbol{G} \in \mathbb{R}^{m\times n}$, $\boldsymbol{h} \in \mathbb{R}^m$, $\boldsymbol{A}^{p\times n}$, $\boldsymbol{b} \in \mathbb{R}^p$；
- 符号 $\boldsymbol{G}\boldsymbol{x} \preceq \boldsymbol{h}$ 的含义是对于任意的 i，我们皆有 $\sum_{i=1}^m \boldsymbol{G}_{i,j} x_j \leq h_i$。

线性规划问题一般而言不具有解析解，但我们有高效的算法来近似求解；将在后面章节介绍。

目录

拓展：Chebyshev 逼近问题

最小二乘中，我们使用欧式距离。另一个合理的设定为

$$\min_{c}\left(\max_{i=1,2,\cdots,N}\left|a_i c - b_i\right|\right)$$

其中为简化，我们仅考虑 $a_i, c, b_i \in \mathbb{R}$。

该问题等价于某个线性规划问题：

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & a_i c - t \le b_i, \qquad i = 1, 2, \cdots, N \\ & -a_i c - t \le -b_i, \qquad i = 1, 2, \cdots, N. \end{aligned}$$

优化变量为 (t, c)。

不同向量的范数和其数据拟合问题

可见对于向量大小的不同衡量标准，优化问题的形式会变得不太一样

目标： 回顾/介绍一下向量的 L^p 范数（包含欧式距离和最大距离）

对于一个向量 $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^\top$

- 数量积 $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_i x_i y_i$（是使用尖括号还是圆括号都可以）
- $\|\boldsymbol{x}\|_2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle^{1/2} = (\sum_i x_i^2)^{1/2}$ 为范数（模）

性质：

- $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$
- Cauchy-Schwarz 不等式：$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2$
- 三角不等式：$\|\boldsymbol{x} + \boldsymbol{y}\|_2 \leq \|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2$

L^p 范数

把范数中下标 2 改为任何的 $p \in [1, \infty)$:

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \quad \text{p-范数}$$

例子: $\boldsymbol{x} = [1, 3]$，请计算 $\|\boldsymbol{x}\|_5$, $\|\boldsymbol{x}\|_{10}$, $\|\boldsymbol{x}\|_{20}$，你有什么发现?

L^p 范数

把范数中下标 2 改为任何的 $p \in [1, \infty)$：

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \quad \text{p-范数}$$

例子： $\boldsymbol{x} = [1, 3]$，请计算 $\|\boldsymbol{x}\|_5$, $\|\boldsymbol{x}\|_{10}$, $\|\boldsymbol{x}\|_{20}$，你有什么发现?

答案：对于这个例子，当 $p \to \infty$, $\|\boldsymbol{x}\|_p \to 3$

L^p 范数

把范数中下标 2 改为任何的 $p \in [1, \infty)$:

$$\boxed{\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}} \quad \text{p-范数}$$

例子: $\boldsymbol{x} = [1, 3]$，请计算 $\|\boldsymbol{x}\|_5$, $\|\boldsymbol{x}\|_{10}$, $\|\boldsymbol{x}\|_{20}$，你有什么发现?

答案: 对于这个例子，当 $p \to \infty$, $\|\boldsymbol{x}\|_p \to 3$

更一般的，当 $p \to \infty$,

$$\boxed{\lim_{p\to\infty} \|\boldsymbol{x}\|_p \overset{\text{现象/结论}}{=} \max_i |x_i| \overset{\text{定义为}}{=} \|\boldsymbol{x}\|_\infty}$$

更一般情况的范数

$\|\cdot\|$ 是一个函数：输入是一个向量 $\boldsymbol{x}$，输出 $\|\boldsymbol{x}\|$ 是一个非负实数。

定义 (向量的范数)

我们说它是一个范数如果它满足如下条件：

- $\|\boldsymbol{x}\| \geq 0$ ($\|\boldsymbol{x}\| = 0$ 当且仅当 $\boldsymbol{x} = 0$)
- $\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$ 对于任何一个实数 α
- $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

我们使用下标来注明这个范数是什么意义下的。

该量本质是用来描述向量的大小。

不同向量的范数 $\to$ 不同的优化问题

假设我们具有数据 a_i, b_i，我们希望使用线性函数去拟合 $b = ca$，

- 若我们使用 L^∞ 范数，我们希望

$$\min_c \max_{i=1,2,\cdots,N} |a_i c - b_i|,$$

  该问题等价为线性规划问题；

- 若我们使用 L^2 范数，我们希望

$$\min_c \sum_{i=1,2\cdots,N} |a_i c - b_i|^2$$

  该问题是典型的最小二乘问题。

例子 1 和例子 2 之间具有微妙关联。

目录

总结

- 掌握优化问题的框架，以及一些基本术语
- 理解最小二乘拟合的框架，以及能推导出最优解的表达式
- 理解线性规划的问题来源以及其一般形式

阅读作业 & 参考资料：

- 课本第 1 章

参考答案

- 第5页：
  - $\max f_0(\boldsymbol{x})$ 可以等价写成 $\min -f_0(\boldsymbol{x})$；两者的最优解相同。
  - $\ell \leq f_1(\boldsymbol{x}) \leq r$ 等价于

$$f_1(\boldsymbol{x}) - r \leq 0, \qquad \text{且} \qquad -f_1(\boldsymbol{x}) + \ell \leq 0.$$

- 第10页：$\sup C = 5$, $\inf C = 1$
- 第13页：C
- 第17页：参数量多于数据量将注定可以拟合成功，但其效果不一定好；会出现拟合的函数不唯一