矩阵论与凸优化第 4 课凸优化初步

管老师

七月在线

Apr, 2017

主要内容

- 优化与凸优化简介
  - 优化问题基本形式
  - 凸优化问题基本形式
  - 凸优化的应用
- 凸集合与凸函数基本概念
  - 凸集合与凸函数的关系
  - 凸集合与凸函数的性质对应
- 凸优化问题举例
  - 混合高斯模型和 EM 算法简介

记号

- 本节课常用数学记号

V, W	向量空间
v, w	向量
$\mathbb{R}^n,\mathbb{R}^m$	实坐标空间
α,β	V 和 W 的基
$T: V \to W$	向量空间 V 到 W 的线性映射
$A_{\alpha,\beta}(T)$	线性映射 T 在 α 和 β 这两组基下的矩阵
$G(v_1, v_2)$	内积空间 V 上的内积
H_α	G 在基 α 下的矩阵形式

优化问题

优化问题的一般形式

$$
\begin{aligned}
\text{最小化：} \quad & f_0(x) \\
\text{条件：} \quad & f_i(x) \leq b_i, i = 1, \cdots, m.
\end{aligned}
$$

其中 $f_0(x)$ 为目标函数，条件里的不等式是限制条件.

优化与凸优化简介

举例：

$Ax=b$

极大似然估计

如果 $L(\mu,\sigma)$ 是一个极大似然估计问题中的似然函数，其中 μ,σ 分别是期望和方差，那么极大似然估计的问题转化为

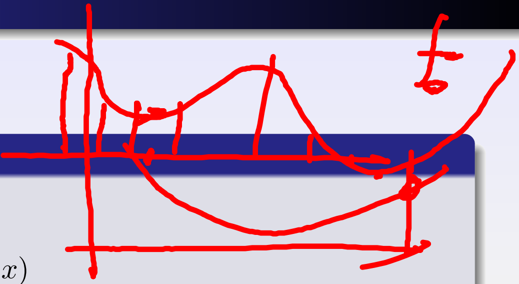
$$\text{最小化：}\quad -L(\mu,\sigma)$$
$$\text{条件：}\quad \sigma \geq 0$$

最小二乘

如果 $A_{n\times k}$ 是一个矩阵,$b\in\mathbb{R}^n$ 是一个向量, 对于 $x\in\mathbb{R}^k$

$$\text{最小化：}\quad f_0(x) = |Ax-b|^2$$

优化与凸优化简介



凸优化问题

凸优化问题的一般形式

$$\begin{aligned} \text{最小化:} \quad & f_0(x) \\ \text{条件:} \quad & f_i(x) \le b_i, i = 1, \cdots, m. \end{aligned}$$

其中 $f_0(x)$ 为目标函数，条件里的不等式是限制条件.

- 凸优化问题的条件,$f_0, f_1, \cdots, f_m$ 都是凸函数.
- 凸优化问题的特点，局部最优等价于全局最优.
- 凸优化问题的求解，几乎总有现成的工具来求解.

优化与凸优化简介

凸优化的应用

- 凸优化问题逼近非凸优化问题，寻找非凸问题的初始点
- 利用对偶问题的凸性给原问题提供下界估计
- 凸优化问题可以给非凸问题带来一些启发

$$\|Ax-b\|^2+\|x\|_0$$

$$\Updownarrow$$

$$\|Ax-b\|^2+\lambda\|x\|_1$$

凸集合与凸函数

凸集合定义

如果一个集合 Ω 中任何两个点之间的线段上任何一个点还属于 Ω, 那么 Ω 就是一个凸集合.i.e.

$$\lambda x_1 + (1-\lambda)x_2 \in \Omega, \forall x_1, x_2 \in \Omega, \lambda \in (0,1)$$

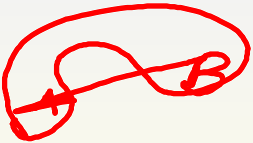
凸函数定义

如果一个函数 f 定义域 Ω 是凸集，而且对于任何两点. 以及两点之间线段上任意一个点都有

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

$\forall x_1, x_2 \in \Omega, \lambda \in (0,1)$

凸集合与凸函数



凸集合与凸函数

函数的上境图

假设 f 是一个定义在 Ω 上的函数，区域 $\{(x,y): y \geq f(x), \forall x \in \Omega\}$ 就是 f 的上境图.

上境图就是函数图像上方的部分区域.

凸集合与凸函数

一个函数是凸函数当且仅当 f 的上境图是凸集合.

凸集合与凸函数有很多相对应的性质可以由这个结论来进行链接。

凸集合与凸函数

凸集合与凸函数

凸组合

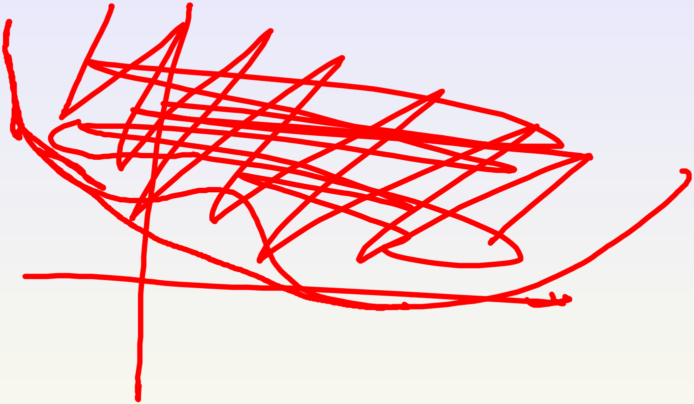
对于任何 n 个点 $\{x_i\}_{i=1}^n$, 以及权重系数 $\{w_i\}_{i=1}^n$. 若权重系数非负 $w_i \geq 0$ 而且 $\sum_{i=1}^{n} w_i = 1$, 则线性组合

$$S = \sum_{i=1}^{n} w_i x_i$$

为一个凸组合.

凸组合的物理意义可以理解成 n 个重量为 w_i 的点的整体重心.

凸集合与凸函数



凸集合与凸函数

集合的凸包

n 个点 $\{x_i\}_{i=1}^n$ 的全部凸组合就构成 $\{x_i\}_{i=1}^n$ 的凸包.

函数的凸闭包

如果 C 是函数 f 的上境图，$\overline{C}$ 是 C 的凸包, 那么以 $\overline{C}$ 为上境图的函数称为 f 的凸闭包.

凸集合与凸函数

集合的凸包的性质

若 $\overline{C}$ 是 C 的凸闭包，那么

- $C \subset \overline{C}$
- C 的支撑平面也是 $\overline{C}$ 的支撑平面，反之亦然.

SVM

函数的凸闭包的性质

若 g 是 f 的凸闭包，那么

- $g \leq f$
- $\inf\{g\} = \inf f$

凸集合与凸函数的对应性质 (凸组合)

凸集合性质

假设 Ω 是一个凸集合，那么 Ω 任何子集的凸包仍包含于 Ω.

凸函数性质：琴生 (Jensen) 不等式

如果 $f:\Omega\to\mathbb{R}$ 是一个凸函数，则对于任何 $\{x_i\in\Omega\}_{i=1}^n$，以及凸组合 $\sum_{i=1}^n w_i x_i$ 都有

$$\sum_{i=1}^n w_i f(x_i) \ge f\Big(\sum_{i=1}^n w_i x_i\Big)$$

话外篇：琴生 (Jensen) 不等式的推论

很多常用的不等式都是琴生不等式的推论，事实上，大部分不等式要么来自于 $x^2 \geq 0$ 要么来自于琴生不等式.

算数集合平均不等式

对于正数 $a_1, \cdots, a_n$

$$\frac{\sum_{i=1}^{n} a_i}{n} \geq \left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}}$$

柯西不等式

$$\left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right) \geq \left(\sum_{i=1}^{n} a_i b_i\right)^2$$

凸集合与凸函数

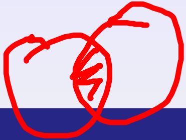
$-\ln x$ 凸

$\frac{1}{x^2} \geq 0$

$$-\ln\left(\frac{\sum a_i}{n}\right) \leq \sum \frac{-\ln a_i}{n}$$

$$\ln\left(\frac{\sum a_i}{n}\right) \geq \sum \frac{\ln a_i}{n}$$

凸集合与凸函数的对应性质 (集合相交)

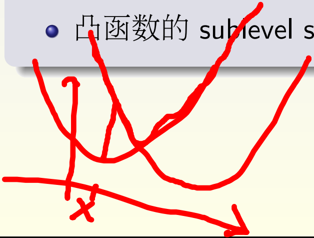


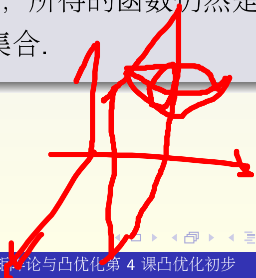
凸集合性质

任意多个凸集合的交集仍是凸集合.

凸函数性质

- 任意多个凸函数的逐点上确界仍是凸函数.
- 固定一个凸函数的若干个变量，所得的函数仍然是凸函数
- 凸函数的 sublevel set 都是凸集合.



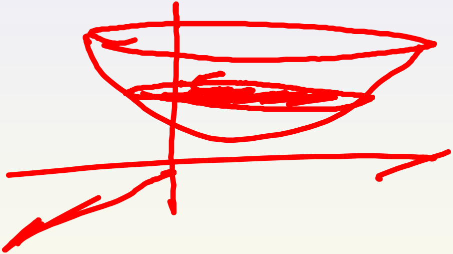


凸集合与凸函数

$f(x)$

$f(x)=c$

$f(x)\le c$



凸集合与凸函数的对应性质 (线性组合)

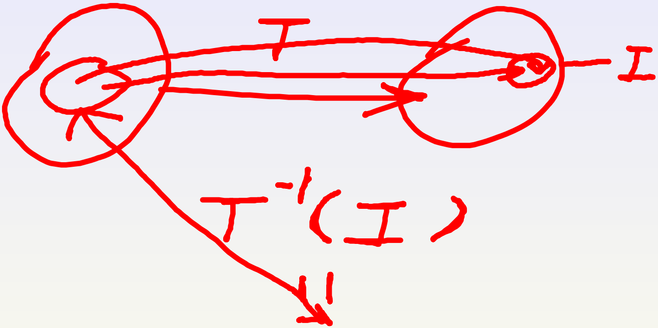
凸集合性质

假设 $T : V \to W$ 是一个线性映射，则

- 若 Ω_V 是 V 中的凸集合，则 $\Omega_W = T(\Omega_V)$ 是 W 中的凸集合.
- 若 Ω_W 是 W 中的凸集合，则 $\Omega_V = T^{-1}(\Omega_W)$ 是 V 中的凸集合.

凸函数性质

- 凸函数的非负线性组合仍是凸函数,$f_1, \cdots, f_k$ 是凸函数，而且 $w_i \geq 0$，则 $\sum_{i=1}^{k} w_i f_k$ 也是凸函数.
- 若 $f : \mathbb{R}^n \to \mathbb{R}$ 是凸函数，$A \in \mathbb{R}^{n\times m}$,$b \in \mathbb{R}^n$，那么复合函数 $g(x) = f(Ax + b)$ 还是凸函数.


T
I
$T^{-1}(I)$

凸集合与凸函数的对应性质 (微分)

凸集合性质

- 若凸集合 Ω 的边界 C 是一个可微曲线，则 C 在任何一点上的切线（平面）都是这个凸集合的支撑线（平面）
- 若凸集合 Ω 的边界 C 是一个二阶可微曲线，则 C 在任何一点上的曲率向量都指向 Ω 内部.

曲率向量的物理意义是加速方向或者受力方向.

凸函数性质

- 若一个凸函数一阶可微，那么凸函数的一阶近似不大于函数本身.i.e.

$$f(x) \geq f(x_0) + (\nabla f(x_0))^T \cdot (x - x_0)$$

- 若一个凸函数二阶可微，那么这个函数的二阶导数 (Henssen 矩阵) 非负（半正定）.

凸集合与凸函数

凸集合与凸函数的对应性质 (光学投影)

凸集合性质

- 若 Ω 是凸集合，那么 Ω 在任何一个平面上的投影仍是凸集合（平行光源投影）
- 若 $\Omega \subset \mathbb{R}^n$ 是凸集合，那么

$$\Omega_{\hat{n}} = \{(x_1/x_n, \cdot, x_{n-1}/x_n, 1) : (x_1, \cdots, x_n) \in \Omega \text{且} x_n \neq 0\}$$

  也是凸集合.(点光源投影)
- 若 $\Omega \subset \mathbb{R}^n$ 是一个凸集合，那么椎体 $tx : x \in \Omega, t \in \mathbb{R}_+$ 也是个凸集合.(点光源)

凸集合与凸函数的对应性质 (光学投影)

凸函数性质

- 若 $f(x,y)$ 是凸函数, 那么 $g(x)=\inf\limits_{(x,y)\in\Omega} f(x,y)$, 也是凸函数
- 若 $f:\mathbb{R}^n\to\mathbb{R}$ 是凸函数，那么 $g(x,t)=tf(x/t):\mathbb{R}^{n+1}\to\mathbb{R}$ 也是个凸函数.

优化问题举例:EM 算法简介与混合高斯模型

潜在变量 (latent variable)

一般指在所处理的问题中不能直接观测的变量

举例

某地希望统计当年高中入学男女学生身高分布，并分别建立正态分布模型:

- 女学生: $M \sim N(\mu_0, \sigma_0)$
- 男学生: $W \sim N(\mu_1, \sigma_1)$

但是测量身高时忘记了记录学生性别。此时，学生性别这一变量就成为了一个潜变量。

混合高斯模型举例

请问如何根据混合身高数据，估计男女学生比例，分别建立男生和女生的身高正态分布模型，并利用此模型来估计每一个样本的性别?

混合高斯模型要求对每一个高斯模型的参数进行估计, 并估计单个样本服从每一个模型的概率。这既是一个统计参数估计问题, 也是一个聚类问题。

混合高斯模型举例

用 τ_0,τ_1 分别表示男女生站总体的比例，那么混合身高数据则服从如下分布:

$$p(x|\tau,\mu,\sigma) = \tau_0 * N(\mu_0,\sigma_0) + \tau_1 * N(\mu_1,\sigma_1)$$

此时所需估计的参数为 $\tau_0,\tau_1,\mu_0,\mu_1,\sigma_0,\sigma_1$ 思考题: 如何证明这个函数是一个概率密度函数?

优化问题举例:EM 算法简介与混合高斯模型

作为一个有明确的概率密度函数表达式的参数估计问题，我们很自然的想到使用极大似然估计的方法:

极大似然估计

$$\tau, \mu, \sigma = \text{argmax} \sum_i p(x_i | \tau, \mu, \sigma)$$

但是这个的计算及其复杂, 复杂的原因在于 ln 里面有两个不同的密度函数。所以如果我们能知道每一个样本的分类，并且能够把这两个密度函数分开就好了。

优化问题举例:EM 算法简介与混合高斯模型

EM 算法

- E 步骤: 对每一个样本的分类进行估计，并且把密度函数分开到不同的 ln 里面去
- M 步骤: 参数估计

EM 算法:E 步骤

- 后验估计分类概率
- 利用凸函数的性质分离变量

优化问题举例:EM 算法简介与混合高斯模型

EM 算法:M 步骤

简单的参数估计。

优化问题举例:EM 算法简介与混合高斯模型

block 样本

block 内容

补充内容

凸优化：参考资料

参考资料

- 非常清晰完整的经典教材：凸优化，Stephen Boyd，Lieven Vandenberghe

作业

- 教材：凸优化 (英文版页码与题号) p60:2.32.10,2.12,2.16,2.23,2.26; p113:3.1,3.2,3.12,3.21,3.36,3.39; p189:4.2,4.8,4.10,4.21 p273:5.1,5.4,5.11,5.225.24

谢谢大家!