机器学习里的数学第 7 课凸优化初步

管老师

七月在线

June, 2016

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
- 凸集分离定理
  - 凸集分离定理
- 共轭凸函数
  - 共轭凸函数
- 凸优化问题举例

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

优化与凸优化简介

优化问题

优化问题的一般形式

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

其中 $f_0(x)$ 为目标函数，条件里的不等式是限制条件.

举例:

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

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

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 $x \in \mathbb{R}^k$

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

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

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 $w_i \leq 0$ 而且 $\sum\limits_{i=1}^{n} w_i = 1$，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.

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.

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

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

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

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

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.

凸集合与凸函数

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

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

  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.

凸集合与凸函数

凸集分离定理

若 C, D 分别为 $\mathbb{R}^n$ 中的两个不交的非空凸集合,i.e.$C \cap D = \emptyset$, 则一定存在向量 $a \in \mathbb{R}^n$ 以及实数 $b \in \mathbb{R}$ 使得任何 $x_C \in C, x_D \in D$ 有 $a^T x_C \leq b$ 以及 $a^T x_D \geq b$.

定理中不等式的几何意义在于 C, D 分别位于超平面 $a^T x = b$ 的两边.

凸集分离定理

凸集分离定理

证明

分情况讨论：如果 C 是一个独点集 $\{p\}$，那么考虑集合 $\{|p-q| : q \in D\}$. 因为点之间的距离是非负数，所以根据确界原理存在一个序列 $\{q_i \in D\}_{i=1}^{\infty}$. 使得

$$\lim_{i\to\infty} |p-q_i| = \inf\{|p-q| : q \in D\}$$

而另一方面有界序列一定存在收敛子序列, 所以可以假定 $\lim_{i\to\infty} q_i = q_\infty$.

(1) 如果 $p \neq q_\infty$, 则令 $a = q_\infty - p$,$b = a^T(p+q_\infty)/2$, 我们得到

$$a^T p = a^T(-\frac{q_\infty - p}{2} + \frac{p+q_\infty}{2}) = -\frac{|a|^2}{2} + b \leq b$$

$$a^T q_\infty = a^T(\frac{q_\infty - p}{2} + \frac{p+q_\infty}{2}) = \frac{|a|^2}{2} + b \geq b$$

继续证明

此时对于任何一个 $q_0 \in D$, 考虑线段 $q_\infty + t(q_0 - q_\infty)$, 因为 $|q_\infty - p| = \inf\{|q - p| : q \in D\}$, 所以

$$f(t) = |q_\infty + t(q_0 - q_\infty) - p|^2 \geq |q_\infty - p|^2 = f(0), \forall t \in (0, 1)$$

所以 $f'(0) \geq 0$, 于是我们有

$$(q_\infty - p)^T (q_0 - q_\infty) \geq 0$$

也就是

$$a^T q_0 \geq a^T q_\infty \geq b$$

.

凸集分离定理

继续证明

(2) 如果 $p = q_\infty$，那么选取序列 $\{p_j\}_{j=1}^{\infty}$，使得 p_j 不在 D 的闭包里面. 那么根据上一种情况我们知道，对每一个 j，都存在着 a_j 使得 $a_j^T q \geq a_j^T p_j$. 因为 a_j 的长度无关紧要，所以可以假设 a_j 都是单位向量。于是有界序列一定存在收敛子序列，所以可以假定 a_j 收敛于一个单位向量 a_∞. 那么对于任何一个 $q \in D$ 我们有

$$
\begin{aligned}
a_\infty^T p &= \lim_{j\to\infty} a_j^T p_j \\
&\leq \lim_{j\to\infty} a_j^T q \\
&= \lim_{j\to\infty} a_j^T q \\
&= a_\infty^T q
\end{aligned}
$$

令 $a = a_\infty$,$b = a_\infty^T p$. 即得所求.

凸集分离定理

继续证明

如果 C, D 是任意凸集合，考虑集合 $C-D=\{x-y: p\in C, q\in D\}$. 那么不难证明 $C-D$ 还是一个凸集合（作业）. 而且零向量 $O\notin C-D$. 于是使用前面证明的独点集与凸集合的分离定理我们有，存在一个向量 a 使得

$$a^T(p-q)\leq q^TO=0$$

所以对于任意的 $p\in C$,$q\in D$, 都有

$$a^Tq\geq a^Tp$$

这时候令 $b=\sup\{a^Tp: p\in C\}$, 既得所证。

共轭函数

共轭函数

如果 $f:\mathbb{R}^n \to \mathbb{R}$ 是一个函数, 那么 f 的共轭函数

$$f^*(y) = \sup_{x\in \mathbf{dom} f} (y^T x - f(x))$$

其中 $f^*(y)$ 的定义域是使得等式右边有界的那些 y.

共轭函数

共轭函数的性质

- 共轭函数 f^* 是一个凸函数
- 如果 g 是 f 的凸闭包，那么 $g^* = f^*$
- 如果 f 是一个凸函数，那么 $f^{**} = f$

谢谢大家!