

毕业设计理论笔记

钟 威

最后更改时间：March 21, 2013

凸规划问题基础

P_{17} 定义 1.2.1

称 S 是凸集，如果对于任意的 $x_1, x_2 \in S$ 和任意的 $\lambda \in [0, 1]$, 都有：

$$\lambda x_1 + (1 - \lambda)x_2 \in S \quad (1.2.1)$$

P_{17} 定义 1.2.3

任意的 $x_1, x_2 \in S$ 和任意的 $\lambda \in (0, 1)$, 都有：

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \quad (1.2.2)$$

P_{19} 定理 1.2.4

$f(x)$ 是凸函数的充要条件是：对于 S 中的任意一点 \bar{x} ，都有：

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}) \quad (1.2.6)$$

P_{20} 定义 1.2.6

凸规划问题是指最优化问题：

$$\min \quad f_0(x), \quad x \in R^n \quad (1.2.10)$$

$$s.t. \quad f_i(x) \leq 0, \quad i = 1, \dots, m \quad (1.2.11)$$

$$h_i(x) = a_i^T x - b_i = 0, \quad i = 1, \dots, p. \quad (1.2.12)$$

其中 $f_0(x)$ 和 $f_i(x)$ 都是定义在 R^n 上的连续可微凸函数，而 $h_i(x)$ 是线性函数。

P_{20} 引理 1.2.8

若 $f(x)$ 是 R^n 上的凸函数，则对于任意的 $c \in R$ ，水平集：

$$S = \{x | f(x) \leq c, x \in R^n\} \quad (1.2.16)$$

是凸集。

P_{20} 定理 1.2.10

考虑凸规划问题 (1.2.10)→(1.2.12)，若 x^* 是它的局部解，则 x^* 也是它的整体解。

P_{21} 定理 1.2.12

若凸规划问题 (1.2.10)→(1.2.12) 中的目标函数 $f_0(x)$ 是严格凸函数，则该问题的解唯一。

P_{21} 定义 1.2.13

设凸规划问题 (1.2.10)→(1.2.12) 中变量 x 具有式：

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad x_i \in R^{m_i}, \quad i = 1, 2. \quad m_1 + m_2 = n. \quad (1.2.22)$$

所示的分划。称 x_1^* 是该问题关于 x_1 的解，如果存在 $x_2^* \in R^{m_2}$ ，使得 $x^* = (x_1^{*T}, x_2^{*T})^T$ 是该问题的解。

P_{22} 定理 1.2.15

设凸规划问题 (1.2.10)→(1.2.12) 中变量 x 具有式 (1.2.22) 的分划, 记 $f_0(x) = \mathbf{F}_0(x_1, x_2)$ 。若 $\mathbf{F}_0(x_1, x_2)$ 分别是变量 x_1 和 x_2 的严格凸函数, 则该问题对 x_1 的解唯一。

凸规划的对偶理论

P_{23} 定义 1.2.16 对偶问题

引进 Lagrange 函数：

$$\mathbf{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \quad (1.2.35)$$

其中 $\lambda = (\lambda_1, \dots, \lambda_m)^T$ 和 $\nu = (\nu_1, \dots, \nu_m)^T$ 是 Lagrange 乘子向量。由书上 P_{23} 的论述可知, 在 $\lambda \geq 0$ 时：

$$\inf_{x \in R^n} \mathbf{L}(x, \lambda, \nu)$$

是 $f_0(x)$ 的下界。对于要找到最好的下界的问题, 我们把：

$$\max \quad g(\lambda, \nu) = \inf_{x \in R^n} \mathbf{L}(x, \lambda, \nu) \quad (1.2.40)$$

$$s.t. \quad \lambda \geq 0 \quad (1.2.41)$$

称 (1.2.40)→(1.2.41) 为问题 (1.2.10)→(1.2.12) 关于 Lagrange 函数 (1.2.35) 的对偶问题, 或简称为问题 (1.2.10)→(1.2.12) 的对偶问题。称 (1.2.10)→(1.2.12) 为原始问题。

P_{24} 定义 1.2.17 对偶间隙

称原始问题的最优值与对偶问题的最优值之差为原始问题的对偶间隙。

P_{24} 定理 1.2.18 弱对偶定理

原始问题 (1.2.10)→(1.2.12) 的对偶间隙永远取非负, 即原始问题的最优值 p^* 和对偶问题 (1.2.40)→(1.2.41) 的最优值 d^* 有如下关系：

$$\begin{aligned} p^* &= \inf\{f_0(x)\} \\ &\geq \sup\{g(\lambda, \nu)\} \\ &= d^* \end{aligned} \quad (1.2.43)$$

上述定理有如下推论：

P_{24} 推论 1.2.19

设 \tilde{x} 和 $(\tilde{\lambda}, \tilde{\nu})$ 分别是原始问题 (1.2.10)→(1.2.12) 和对偶问题 (1.2.40)→(1.2.41) 的可行解。若 $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$, 则 \tilde{x} 和 $(\tilde{\lambda}, \tilde{\nu})$ 分别是原始问题和对偶问题的整体解。

P_{24} 定理 1.2.20 Slater 条件

强对偶定理讨论的是对偶间隙为零的情况, 我们需要“约束规格”来保证对偶间隙为零。对于凸规划问题 (1.2.10)→(1.2.12), 最简单的约束规格是满足 Slater 条件：如果存在可行点 x , 使得：

$$f_i(x) < 0, \quad i = 1, \dots, m; \quad a_i^T x - b_i = 0, \quad i = 1, \dots, p \quad (1.2.44)$$

特别地, 当凸规划问题的前 k 个不等式约束为线性约束： $f_i(x) = a_i^T x - b_i \leq 0, \quad i = 1, \dots, k$ 时, 条件可进一步宽松为书上 (1.2.45)→(1.2.46)。

P_{25} 定理 1.2.21 强对偶定理

对于凸规划问题 (1.2.10)→(1.2.12)，若它满足 Slater 条件，则它的对偶间隙为零。进一步，若还知原始问题的最优值可以达到，即存在最优解 x^* ，则对偶问题的最优值也可以达到，即存在对偶问题的整体解 (λ^*, ν^*) 使得：

$$\begin{aligned} f_0(x^*) &= \inf f_0(x) \\ &= \sup\{g(\lambda, \nu) | \lambda \leq 0\} \\ &= \max\{g(\lambda, \nu) | \lambda \leq 0\} \\ &= g(\lambda^*, \nu^*) \end{aligned} \quad (1.2.47)$$

P_{25} 定义 1.2.22 KKT 条件

考虑凸规划问题 (1.2.10)→(1.2.12)，称 x^* 满足 KKT 条件，如果存在分别于约束 (1.2.11) 和约束 (1.2.12) 对应的乘子向量 $\lambda = (\lambda_1^*, \dots, \lambda_m^*)$ 和 $\nu = (\nu_1^*, \dots, \nu_p^*)$ 使得 Lagrange 函数：

$$\mathbf{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \quad (1.2.48)$$

满足：

$$f_i(x^*) \leq 0, \quad i = 1, \dots, m \quad (1.2.49)$$

$$h_i(x^*) = 0, \quad i = 1, \dots, p \quad (1.2.50)$$

$$\lambda_i^* \geq 0, \quad i = 1, \dots, m \quad (1.2.51)$$

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \dots, m \quad (1.2.52)$$

$$\nabla_x \mathbf{L}(x^*, \lambda^*, \nu^*) = \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) = 0 \quad (1.2.53)$$

利用强对偶定理可以证明 KKT 条件是凸规划解的必要条件：

P_{25} 定理 1.2.23

考虑凸规划问题 (1.2.10)→(1.2.12)，并设它满足 Slater 条件。若 x^* 是该问题的解，则 x^* 满足 KKT 条件。

下列定理表明 KKT 条件不仅是凸规划解的必要条件，而且也是充分条件。

P_{26} 定理 1.2.24

考虑凸规划问题 (1.2.10)→(1.2.12)，若 x^* 满足 KKT 条件，则 x^* 是该问题的解。

综合上面两个定理，有：

P_{27} 定理 1.2.25

对于满足 Slater 条件的凸规划问题 (1.2.10)→(1.2.12)，点 x^* 是解的充分必要条件是它满足 KKT 条件。

P_{27} 线性规划问题

$$\min \quad c^T x, \quad x \in \mathbf{R}^n \quad (1.2.57)$$

$$s.t. \quad \bar{A}x - \bar{b} \leq 0 \quad (1.2.58)$$

$$Ax - b = 0 \quad (1.2.59)$$

线性问题对应的 Lagrange 函数应为：

$$\mathbf{L}(x, \lambda, \nu) = c^T x + \lambda^T (\bar{A}x - \bar{b}) + \nu^T (Ax - b) \quad (1.2.60)$$

P_{27} 定理 1.2.26

线性规划问题 (1.2.57)→(1.2.59) 的对偶问题是：

$$\max \quad -\bar{b}^T \lambda - b^T \nu \quad (1.2.61)$$

$$s.t. \quad \bar{A}^T \lambda + A^T \nu + c = 0 \quad (1.2.62)$$

$$\lambda \geq 0 \quad (1.2.63)$$

线性分类机

P_{56} 分类问题

给定训练集

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (R^n \times \mathcal{Y})^l \quad (2.1.5)$$

其中 $x_i \in R^n, y_i \in \mathcal{Y} = \{1, -1\}, i = 1, \dots, l$ ，寻找 R^n 空间上的一个实值函数 $g(x)$ ，以使用决策函数

$$f(x) = \text{sgn}(g(x)) \quad (2.1.6)$$

推断任意输入 x 对应的输出 y 。

P_{57} 线性可分问题

略

P_{59} 通过最大间隔法将分类问题转化为

$$\max_{\omega, b} \quad \frac{2}{\|\omega\|} \quad (2.2.3)$$

$$s.t. \quad \forall i \in \{i | y_i = 1\} : (\omega \cdot x_i) + b \geq 1 \quad (2.2.4)$$

$$\forall i \in \{i | y_i = -1\} : (\omega \cdot x_i) + b \leq -1 \quad (2.2.5)$$

或者：

$$\min_{\omega, b} \quad \frac{1}{2} \|\omega\|^2 \quad (2.2.8)$$

$$s.t. \quad y_i((\omega \cdot x_i) + b) \geq 1, \quad i = 1, \dots, l \quad (2.2.9)$$

P_{62} 分类问题的对偶问题

通过求解 (2.2.8)→(2.2.9) 的对偶问题而寻求它的解。引入 Lagrange 函数：

$$\mathbf{L}(\omega, b, \alpha) = \frac{1}{2} \|\omega\|^2 + \sum_{i=1}^l \alpha_i (y_i ((\omega \cdot x_i) + b) - 1) \quad (2.2.26)$$

根据定义 1.2.16 给出的对偶问题应该为：

$$\max_{\omega, b} \quad g(\alpha) = \inf_{\omega, b} \mathbf{L}(\omega, b, \alpha) \quad (2.2.30)$$

$$s.t. \quad \alpha \geq 0 \quad (2.2.31)$$

由于 $\mathbf{L}(\omega, b, \alpha)$ 是 ω 的严格凸二次函数，它取最小值时：

$$\nabla_{\omega} \mathbf{L}(\omega, b, \alpha) = \omega - \sum_{i=1}^l y_i x_i \alpha_i = 0 \quad (2.2.32)$$

即

$$\omega = \sum_{i=1}^l \alpha_i y_i x_i \quad (2.2.33)$$

代入式 (2.2.26)：

$$\inf_{\omega, b} \mathbf{L}(\omega, b, \alpha) = \begin{cases} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) + \sum_{j=1}^l \alpha_j, & \text{when } \sum_{i=1}^l y_i \alpha_i = 0 \\ -\infty & \text{else} \end{cases} \quad (2.2.35)$$

故最优化问题：

$$\max_{\alpha} \quad -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j (x_i \cdot x_j) \alpha_i \alpha_j + \sum_{j=1}^l \alpha_j \quad (2.2.27)$$

$$s.t. \quad \sum_{i=1}^l y_i \alpha_i = 0 \quad (2.2.28)$$

$$\alpha_i \geq 0, \quad i = 1, \dots, l \quad (2.2.29)$$

是原始问题 (2.2.8)→(2.2.9) 的对偶问题。在最优化方法的研究中常用等价的最小化问题取代最大化问题，这里也引进与上述问题有相同解集的最小化问题：

$$\min_{\alpha} \quad \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j (x_i \cdot x_j) \alpha_i \alpha_j - \sum_{j=1}^l \alpha_j \quad (2.2.36)$$

$$s.t. \quad \sum_{i=1}^l y_i \alpha_i = 0 \quad (2.2.37)$$

$$\alpha_i \geq 0, \quad i = 1, \dots, l \quad (2.2.38)$$

该问题的矩阵形式可以表示为：

$$\min_{\alpha} \quad W(\alpha) = \frac{1}{2} \alpha^T H \alpha - e^T \alpha \quad (2.2.40)$$

$$s.t. \quad \alpha^T y = 0 \quad (2.2.41)$$

$$\alpha \geq 0 \quad (2.2.42)$$

其中：

$$\begin{aligned} H &= (y_i y_j (x_i \cdot x_j))_{l \times l} \\ e &= (1, \dots, 1)^T \\ \alpha &= (\alpha_1, \dots, \alpha_l)^T \\ y &= (y_1, \dots, y_l)^T \end{aligned} \quad (2.2.39)$$

因为存在可逆矩阵 $Q = (y_1 x_1, \dots, y_l x_l)$ 使得 $H = Q^T Q$ ，所以 H 是半正定矩阵。由于该问题满足 Slater 条件，由定理 1.2.23 知，对于问题的解 α^* 来说，存在 Lagrange 乘子向量 \tilde{b}^* 和 s^* 满足：

$$\mathbf{L}(\alpha, s, \tilde{b}) = W(\alpha) + s(-\alpha) + \tilde{b}(\alpha^T y)$$

由 KKT 条件可以得出：

$$\alpha^{*T} y = 0, \quad \alpha^* \geq 0 \quad (2.2.47)$$

$$s^* \geq 0, \quad s^{*T} \alpha^* = 0 \quad (2.2.48)$$

$$H\alpha^* - e + \tilde{b}^* y - s^* = 0 \quad (2.2.49)$$

由式 (2.2.48) 和 (2.2.49) 可知：

$$H\alpha^* - e + \tilde{b}^* y \geq 0 \quad (2.2.50)$$

令 $\omega^* = \sum_{i=1}^l \alpha_i^* y_i x_i$ ，则式 (2.2.50) 等价于：

$$y_i((\omega^* \cdot x_i) + \tilde{b}^*) \geq 1, \quad i = 1, \dots, l \quad (2.2.51)$$

这表明 (ω^*, \tilde{b}^*) 是原始问题 (2.2.8)→(2.2.9) 的可行解。更进一步，由 (2.2.33) 和 (2.2.47)→(2.2.49) 可得：

$$\begin{aligned} \frac{1}{2} \|\omega^*\|^2 &= \frac{1}{2} \alpha^{*T} H \alpha^* \\ &= \frac{1}{2} \alpha^{*T} H \alpha^* - \alpha^{*T} (H \alpha^* + \tilde{b}^* y - e - s^*) \\ &= -\frac{1}{2} \alpha^{*T} H \alpha^* - \tilde{b}^* \alpha^{*T} y + e^T \alpha^* + s^{*T} \alpha^* \\ &= -\frac{1}{2} \alpha^{*T} H \alpha^* - \tilde{b}^* \mathbf{0} + e^T \alpha^* + \mathbf{0} \\ &= -\frac{1}{2} \alpha^{*T} H \alpha^* + e^T \alpha^* \end{aligned} \quad (2.2.52)$$

因此，原始问题 (2.2.8)→(2.2.9) 在 (ω^*, \tilde{b}^*) 点的目标函数值和对偶问题 (2.2.40)→(2.2.42) 的最优值相同。

P_{63} 定理 2.2.7

考虑线性可分问题，则对偶问题 (2.2.27)→(2.2.29) 必有解 $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$ 。

P_{63} 定理 2.2.8

最优化问题 (2.2.36)→(2.2.38) 是凸二次规划。

P_{63} 定理 2.2.9

若 $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$ 是问题 (2.2.36)→(2.2.38) 的任意解, 则原始问题 (2.2.8)→(2.2.9) 的唯一解是:

$$\omega^* = \sum_{i=1}^l \alpha_i^* y_i x_i \quad (2.2.43)$$

$$b^* = y_j - \sum_{i=1}^l \alpha_i^* y_i (x_i \cdot x_j) \quad (2.2.44)$$

或者

$$\omega^* = \sum_{i=1}^l \alpha_i^* y_i x_i \quad (2.2.45)$$

$$b^* = -\frac{\left(\omega^* \cdot \sum_{i=1}^l \alpha_i^* x_i\right)}{2 \sum_{i=1}^l \alpha_i^*} \quad (2.2.46)$$

P_{65} 算法 2.2.10 线性可分支持向量分类机

(1)

给定训练集合 $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (R^n \times \mathcal{Y})^l$

(2)

构造并求解凸二次规划问题:

$$\min_{\alpha} \quad \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j (x_i \cdot x_j) \alpha_i \alpha_j - \sum_{j=1}^l \alpha_j \quad (2.2.57)$$

$$s.t. \quad \sum_{i=1}^l y_i \alpha_i = 0 \quad (2.2.58)$$

$$\alpha_i \geq 0, \quad i = 1, \dots, l \quad (2.2.59)$$

得到解 $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$

(3)

计算 $\omega^* = \sum_{i=1}^l \alpha_i^* y_i x_i$, 选取 α^* 的一个正分量 α_j^* , 据此计算:

$$b^* = y_j - \sum_{i=1}^l \alpha_i^* y_i (x_i \cdot x_j) \quad (2.2.60)$$

(4)

构造分划超平面 $(\omega^* \cdot x) + b^* = 0$, 由此求得决策函数:

$$f(x) = \text{sgn}(g(x)) \quad (2.2.61)$$

其中,

$$g(x) = (\omega^* \cdot x) + b^* = \sum_{i=1}^l y_i \alpha_i^* (x_i \cdot x) + b^* \quad (2.2.62)$$