附录 C　拉格朗日对偶性

在约束最优化问题中，常常利用拉格朗日对偶性（Lagrange duality）将原始问题转换为对偶问题，通过解对偶问题而得到原始问题的解．该方法应用在许多统计学习方法中，例如，最大熵模型与支持向量机．这里简要叙述拉格朗日对偶性的主要概念和结果．

1．原始问题

假设 $f(x)$，$c_i(x)$，$h_j(x)$ 是定义在 $\mathbf{R}^n$ 上的连续可微函数．考虑约束最优化问题

$$\min_{x\in\mathbf{R}^n} f(x) \tag{C.1}$$

$$\text{s.t.}\quad c_i(x)\leqslant 0,\quad i=1,2,\cdots,k \tag{C.2}$$

$$h_j(x)=0,\quad j=1,2,\cdots,l \tag{C.3}$$

称此约束最优化问题为原始最优化问题或原始问题．

首先，引进广义拉格朗日函数（generalized Lagrange function）

$$L(x,\alpha,\beta)=f(x)+\sum_{i=1}^{k}\alpha_i c_i(x)+\sum_{j=1}^{l}\beta_j h_j(x) \tag{C.4}$$

这里，$x=(x^{(1)},x^{(2)},\cdots,x^{(n)})^{\mathrm{T}}\in\mathbf{R}^n$，$\alpha_i,\beta_j$ 是拉格朗日乘子，$\alpha_i\geqslant 0$．考虑 x 的函数：

$$\theta_P(x)=\max_{\alpha,\beta:\alpha_i\geqslant 0} L(x,\alpha,\beta) \tag{C.5}$$

这里，下标 P 表示原始问题．

假设给定某个 x．如果 x 违反原始问题的约束条件，即存在某个 i 使得 $c_i(w)>0$ 或者存在某个 j 使得 $h_j(w)\neq 0$，那么就有

$$\theta_P(x)=\max_{\alpha,\beta:\alpha_i\geqslant 0}\left[f(x)+\sum_{i=1}^{k}\alpha_i c_i(x)+\sum_{j=1}^{l}\beta_j h_j(x)\right]=+\infty \tag{C.6}$$

因为若某个 i 使约束 $c_i(x)>0$，则可令 $\alpha_i\to+\infty$，若某个 j 使 $h_j(x)\neq 0$，则可令 β_j 使 $\beta_j h_j(x)\to+\infty$，而将其余各 α_i，β_j 均取为 0．

相反地，如果 x 满足约束条件式 (C.2) 和式 (C.3)，则由式 (C.5) 和式 (C.4) 可知，$\theta_P(x)=f(x)$．因此，

$$\theta_P(x)=\begin{cases} f(x), & x\text{满足原始问题约束} \\ +\infty, & \text{其他} \end{cases} \tag{C.7}$$

所以如果考虑极小化问题

$$\min_x \theta_P(x)=\min_x \max_{\alpha,\beta:\alpha_i\geqslant 0} L(x,\alpha,\beta) \tag{C.8}$$

它是与原始最优化问题 (C.1) ～ (C.3) 等价的，即它们有相同的解．问题 $\min\limits_x \max\limits_{\alpha,\beta:\alpha_i\geqslant 0} L(x,\alpha,\beta)$ 称为广义拉格朗日函数的极小极大问题．这样一来，就把原始最优化问题表示为广义拉格朗日函数的极小极大问题．为了方便，定义原始问题的最优值

$$p^*=\min_x \theta_P(x) \tag{C.9}$$

称为原始问题的值．

2．对偶问题

定义

$$\theta_D(\alpha,\beta)=\min_x L(x,\alpha,\beta) \tag{C.10}$$

再考虑极大化 $\theta_D(\alpha,\beta)=\min\limits_x L(x,\alpha,\beta)$，即

$$\max_{\alpha,\beta:\alpha_i\geqslant 0} \theta_D(\alpha,\beta)=\max_{\alpha,\beta:\alpha_i\geqslant 0} \min_x L(x,\alpha,\beta) \tag{C.11}$$

问题 $\max\limits_{\alpha,\beta:\alpha_i\geqslant 0} \min\limits_x L(x,\alpha,\beta)$ 称为广义拉格朗日函数的极大极小问题．

可以将广义拉格朗日函数的极大极小问题表示为约束最优化问题：

$$\max_{\alpha,\beta} \theta_D(\alpha,\beta)=\max_{\alpha,\beta} \min_x L(x,\alpha,\beta) \tag{C.12}$$

$$\text{s.t.} \quad \alpha_i\geqslant 0,\quad i=1,2,\cdots,k \tag{C.13}$$

称为原始问题的对偶问题．定义对偶问题的最优值

$$d^*=\max_{\alpha,\beta:\alpha_i\geqslant 0} \theta_D(\alpha,\beta) \tag{C.14}$$

称为对偶问题的值．

3．原始问题和对偶问题的关系

下面讨论原始问题和对偶问题的关系．

定理 C.1　若原始问题和对偶问题都有最优值，则

$$d^* = \max_{\alpha,\beta:\alpha_i\geqslant 0} \min_x L(x,\alpha,\beta) \leqslant \min_x \max_{\alpha,\beta:\alpha_i\geqslant 0} L(x,\alpha,\beta) = p^* \tag{C.15}$$

证明　由式 (C.12) 和式 (C.5)，对任意的 α,β 和 x，有

$$\theta_D(\alpha,\beta) = \min_x L(x,\alpha,\beta) \leqslant L(x,\alpha,\beta) \leqslant \max_{\alpha,\beta:\alpha_i\geqslant 0} L(x,\alpha,\beta) = \theta_P(x) \tag{C.16}$$

即

$$\theta_D(\alpha,\beta) \leqslant \theta_P(x) \tag{C.17}$$

由于原始问题和对偶问题均有最优值，所以，

$$\max_{\alpha,\beta:\alpha_i\geqslant 0} \theta_D(\alpha,\beta) \leqslant \min_x \theta_P(x) \tag{C.18}$$

即

$$d^* = \max_{\alpha,\beta:\alpha_i\geqslant 0} \min_x L(x,\alpha,\beta) \leqslant \min_x \max_{\alpha,\beta:\alpha_i\geqslant 0} L(x,\alpha,\beta) = p^* \tag{C.19}$$

■

推论 C.1　设 x^* 和 α^*,β^* 分别是原始问题 (C.1) ～ (C.3) 和对偶问题 (C.12) ～ (C.13) 的可行解，并且 $d^* = p^*$，则 x^* 和 α^*,β^* 分别是原始问题和对偶问题的最优解．

在某些条件下，原始问题和对偶问题的最优值相等，$d^* = p^*$．这时可以用解对偶问题替代解原始问题．下面以定理的形式叙述有关的重要结论而不予证明．

定理 C.2　考虑原始问题 (C.1) ～ (C.3) 和对偶问题 (C.12) ～ (C.13)．假设函数 $f(x)$ 和 $c_i(x)$ 是凸函数，$h_j(x)$ 是仿射函数；并且假设不等式约束 $c_i(x)$ 是严格可行的，即存在 x，对所有 i 有 $c_i(x)<0$，则存在 x^*,α^*,β^*，使 x^* 是原始问题的解，α^*,β^* 是对偶问题的解，并且

$$p^* = d^* = L(x^*,\alpha^*,\beta^*) \tag{C.20}$$

定理 C.3　对原始问题 (C.1) ～ (C.3) 和对偶问题 (C.12) ～ (C.13)，假设函数 $f(x)$ 和 $c_i(x)$ 是凸函数，$h_j(x)$ 是仿射函数，并且不等式约束 $c_i(x)$ 是严格可行的，则 x^* 和 α^*,β^* 分别是原始问题和对偶问题的解的充分必要条件是 x^*,α^*,β^* 满足下面的 Karush-Kuhn-Tucker (KKT)条件：

$$\nabla_x L(x^*,\alpha^*,\beta^*) = 0 \tag{C.21}$$

$$\nabla_\alpha L(x^*,\alpha^*,\beta^*) = 0 \tag{C.22}$$

$$\nabla_\beta L(x^*,\alpha^*,\beta^*) = 0 \tag{C.23}$$

$$\alpha_i^* c_i(x^*) = 0, \quad i = 1,2,\cdots,k \tag{C.24}$$

$$c_i(x^*) \leqslant 0, \quad i = 1,2,\cdots,k \tag{C.25}$$

$$\alpha_i^* \geqslant 0, \quad i = 1,2,\cdots,k \tag{C.26}$$

$$h_j(x^*) = 0 \quad j = 1,2,\cdots,l \tag{C.27}$$

特别指出，式 (C.24) 称为 KKT 的对偶互补条件．由此条件可知：若 $\alpha_i^* > 0$，则 $c_i(x^*) = 0$．