拉格朗日对偶性

对于无约束问题，可以使用一阶的梯度下降方法，二阶的牛顿法或者拟牛顿法，在约束优化问题中，通常需要利用拉格朗日对偶性，将原始问题转换为对偶问题，通过求解对偶问题得到原始问题的解。为什么需要转换呢？原始问题中，目标函数可能是凸函数，现有的优化理论都是针对，凹函数求极小值的。所以可以先转换成与之等价的对偶问题，对偶问题是凹函数，可以求极小值，方便求解，可以使用多种优化理论求解。在最大熵模型和支持向量机中得到了应用。

1、原始问题

假设 $f(x), c_i(x), h_j(x)$ 是定义在 R^n 上连续可微函数。考虑约束优化问题：

$$\min_{x \in R^n} \ f(x)$$

$$s.t. \quad c_i(x) \le 0, \quad i = 1, 2, \ldots, k$$

$$h_j(x) = 0, \quad j = 1, 2, \ldots, l$$

称此约束最有化问题为原始最优化问题或原始问题。针对这个问题，首先，引入广义拉格朗日函数：

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

这里，$x = (x^{(1)}, x^{(2)}, \ldots, x^{(n)})^T \in R^n$，其中，$\alpha_i, \beta_j$ 是拉格朗日乘子，$\alpha_i \ge 0$，在这里 α_i 如果小于0，那么就会导致整个函数无限大，所以需要限制其大于0。假设固定 α，β，考虑 x 的函数，则有，原问题是个凸函数：

$$\theta_P(x) = \max_{\alpha, \beta; \alpha_i \ge 0} L(x, \alpha, \beta)$$

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，存在某个 i 使得 $c_i(w) > 0$ 或者存在某个 j 使得 $h_j(w) \ne 0$，那么就有：

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

因为若某个 i 使约束条件 $c_i(x) > 0$，则可令 $\alpha_i \to +\infty$，若某个 j 使 $h_j(x) \ne 0$，则可令 β_j 使 $\beta_j h_j(x) \to +\infty$，然后再将其他的 α_i，β_j 均取为0，所以原始问题的约束条件必须满足上面的，不然是不成立的。

如果满足上面的约束条件，就有 $\theta_P(x) = f(x)$，为什么呢？$L(x, \alpha, \beta) \le f(x)$ 由于拉格朗日函数的第三项为0，当拉格朗日函数取最大值的时候，也就是当第二项为0的时候最大，所以只要满足 $\alpha_i c_i(x) = 0$ 就能满足条件。

为了求解原始问题，需要极小化，所以转换为极小化问题：

$$\min_x \theta_P(x) = \min_x \max_{\alpha, \beta; \alpha_i \ge 0} L(x, \alpha, \beta)$$

所以原始问题的等价形式就是上面函数的形式，也就是求解极小极大问题。这样依赖，就把原来的最优化问题表示成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设置为 p^*：

$$p^* = \min_x \theta_P(x)$$

2、对偶问题

对偶问题就是将原问题的极小极大的顺序翻转，可以定义对偶问题为：

$$\theta_D(\alpha,\beta) = \min_x L(x,\alpha,\beta)$$

再考虑极大化对偶问题，可以得到，对偶问题是个凹函数

$$\max_{\alpha,\beta;\alpha_i\geq 0} \theta_D(\alpha,\beta) = \max_{\alpha,\beta;\alpha_i\geq 0} \min_x L(x,\alpha,\beta)$$

上面问题也叫广义拉格朗日的极大极小问题，这就称为原*/ 始问题的对偶问题，定义对偶问题的最优值：

$$d^* = \max_{\alpha,\beta;\alpha_i\geq 0} \theta_D(\alpha,\beta)$$

3、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有值，则

$$d^* = \max_{\alpha,\beta;\alpha_i\geq 0} \theta_D(\alpha,\beta) \leq \min_x \max_{\alpha,\beta;\alpha_i\geq 0} L(x,\alpha,\beta) = p^*$$

证明：对任意的 α,β 和 x，有

$$\theta_D(\alpha,\beta) = \min_x L(x,\alpha,\beta) \leq L(x,\alpha,\beta) \leq \max_{\alpha,\beta;\alpha_i\geq 0} L(x,\alpha,\beta) = \theta_P(x)$$

所以以下等式恒成立：

$$\theta_D(\alpha,\beta) \leq \theta_P(x)$$

也就是需要满足前者的最大值小于后者的最小值：

$$\max_{\alpha,\beta;\alpha\geq 0} \theta_D(\alpha,\beta) \leq \min_x \theta_P(x)$$

所以有：

$$d^* = \max_{\alpha,\beta;\alpha_i\geq 0} \theta_D(\alpha,\beta) \leq \min_x \max_{\alpha,\beta;\alpha_i\geq 0} L(x,\alpha,\beta) = p^*$$

推论 C.1 设 x^* 和 α^*,β^* 分别是原始问题和对偶问题的可行解，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，原始问题和最优问题的值相等，$d^*=p^*$，这时可以用解对偶问题替代原始问题。只有在 $d^*=p^*$ 时，二者才能替换。

定理 C.2 考虑原始函数和对偶函数，假设函数 $f(x)$ 和 $c_i(x)$ 是凸函数 这也是为什么要引入对偶函数，因为凸函数不能直接求最小值，$h_j(x)$ 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 对所有的 $c_i(x)<0$，如果是等式的情况下，不需要引入**KKT**条件，则存在 x^* 和 α^*,β^* ，使 x^* 是原始问题的解，α^*,β^* 是对偶问题的解，这个定理描述一定存在解。

$$p^* = d^* = L(x^*,\alpha^*,\beta^*)$$

定理 C.3
考虑原始函数和对偶函数，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 对所有的 $c_i(x)<0$，，则存在 x^* 和 α^*,β^* ，使 x^* 是原始问题的解，α^*,β^* 是对偶问题的解。充分必要条件是，x^*,α^*,β^* 满足下面的KKT条件，这个定理描述解应该满足的条件：

$$\nabla_x L(x^*,\alpha^*,\beta^*) = 0$$

$$\nabla_\alpha L(x^*,\alpha^*,\beta^*) = 0$$

$$\nabla_\beta L(x^*,\alpha^*,\beta^*) = 0$$

$$\alpha^* c_i(x^*) = 0 \quad \text{这个条件满足了才能保证广义拉格朗日函数不等式成立}$$

$$c_i(x^*) \leq 0 \quad \text{有了这一项才需要}KKT\text{条件}$$

$\alpha_i \geq 0$　保证不会出现无穷大

$h_j(x) = 0$　如果只有这一项则不需要KKT条件