拉格朗日对偶性

预备知识

对于下文涉及到的$\min\limits_{\boldsymbol{x}}$、$\max\limits_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0}$、凸函数、仿射函数等基本概念可参考此博客拉格朗日对偶，从0到完全理解_frostime的博客的 2. **预备的数学知识**

正文

原始问题

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

$$\begin{aligned} &\min_{\boldsymbol{x}} f(\boldsymbol{x}) \\ s.t.\quad &c_i(\boldsymbol{x}) \leq 0,\quad i=1,2,..,k \\ &h_j(\boldsymbol{x}) = 0,\quad j=1,2,..,l \end{aligned} \tag{B.1}$$

称此约束最优化问题为原始最优化问题或原始问题。

构造广义拉格朗日函数：

$$L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}) = f(\boldsymbol{x}) + \sum_{i=1}^{k}\alpha_i c_i(\boldsymbol{x}) + \sum_{j=1}^{l}\beta_j h_j(\boldsymbol{x}) \tag{B.2}$$

先说**结论**：

原最优化问题式$A.7$与$\min\limits_{\boldsymbol{x}}\max\limits_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta})$等价的，即他们有相同的解。

下面开始说明，这两个为什么等价：

将$\max\limits_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta})$这部分，记为$\theta_P(\boldsymbol{x})$。假设给定某个$\boldsymbol{x}$，如果$\boldsymbol{x}$不满足原始问题的约束条件，即存在某个$i$使得$c_i(\boldsymbol{x}) > 0$或者存在某个$j$使得$h_j(\boldsymbol{x}) \neq 0$，那么就有

$$\theta_P(\boldsymbol{x}) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0}[f(\boldsymbol{x}) + \sum_{i=1}^{k}\alpha_i c_i(\boldsymbol{x}) + \sum_{j=1}^{l}\beta_j h_j(\boldsymbol{x})] = +\infty \tag{B.3}$$

相反的，如果$\boldsymbol{x}$满足原始问题的约束条件，那么$\theta_P(\boldsymbol{x}) = f(\boldsymbol{x})$。因此，

$$\theta_P(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}), & \boldsymbol{x}\text{满足原始问题约束} \\ +\infty, & \text{其它} \end{cases} \tag{B.4}$$

所以极小化$\theta_P(\boldsymbol{x})$，即$\min\limits_{\boldsymbol{x}} \theta_P(\boldsymbol{x})$与原始问题等价，将$\min\limits_{\boldsymbol{x}} \max\limits_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta})$成为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值问题。为了方便，定义原始问题的最优值：

$$p^* = \min_x \theta_P(x) \tag{B.5}$$

称为原始问题的值。

对偶问题

再定义广义拉格朗日函数的极大极小问题：

$$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} \min_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}) \tag{B.6}$$

将$\min\limits_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta})$这部分，记为$\theta_D(\boldsymbol{\alpha},\boldsymbol{\beta})$。再来个转换，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：

$$\begin{aligned} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \theta_D(\boldsymbol{\alpha},\boldsymbol{\beta}) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \min_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}) \\ s.t.\ \ \alpha_i \geq 0, \quad i=1,2,...,k \end{aligned} \tag{B.7}$$

式B.7称为原始问题的对偶问题。为了方便，定义对偶问题的最优值：

$$d^* = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} \theta_D(\boldsymbol{\alpha},\boldsymbol{\beta}) \tag{B.8}$$

称为对偶问题的值。

原始问题和对偶问题的关系

定理1：对偶问题给出了原始问题最优解的下界。由:$\min\limits_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}) \leq L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}) \leq \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta})$，可知$\theta_D(\boldsymbol{\alpha},\boldsymbol{\beta}) \leq \theta_P(\boldsymbol{x})$。当原始问题和对偶问题均最优解时，那么$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} \theta_D(\boldsymbol{\alpha},\boldsymbol{\beta}) \leq \min_x \theta_P(x)$，即

$$d^* = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} \min_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}) \leq \min_{\boldsymbol{x}} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}:\alpha_i\geq 0} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}) = p^*$$

定理2：当$f(\boldsymbol{x})$和$c_i(\boldsymbol{x})$为凸函数，$h_j(\boldsymbol{x})$是仿射函数，并且不等式约束是严格可行的，即存在x，对所有i有$c_i(x)<0$，则存在$\boldsymbol{x^*},\boldsymbol{\alpha^*},\boldsymbol{\beta^*}$，分别是原始问题，对偶问题的解，且：

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

定理3：当$f(\boldsymbol{x})$和$c_i(\boldsymbol{x})$为凸函数，$h_j(\boldsymbol{x})$是仿射函数，并且不等式约束是严格可行的，即存在x，对所有i有$c_i(x)<0$，则$\boldsymbol{x^*},\boldsymbol{\alpha^*},\boldsymbol{\beta^*}$分别是原始问题，对偶问题的解的充分必要条件是$\boldsymbol{x^*},\boldsymbol{\alpha^*},\boldsymbol{\beta^*}$满足KKT条件：

$$
\begin{cases}
\nabla_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta}) = 0 \\
\alpha_i c_i(x) = 0\ , i=1,2,...k \\
\alpha_i \geq 0\ , i=1,2,...k \\
c_i(\boldsymbol{x}) \leq 0\ , i=1,2,...k \\
h_i(\boldsymbol{x}) = 0\ , j=1,2,...,l
\end{cases} \tag{A.9}
$$

引申:

SVM的最优化问题为

$$
\min_{w,b} \frac{1}{2}\|w\|^2 \\
s.t.\ \ y_i(w^T x_i + b) \geq 1\ (i \in [m])
$$

可以看出，svm的目标函数、不等式约束条件是凸函数，并且当$\boldsymbol{x}$不是支持向量时，不等式约束是严格执行的，所以强对偶性成立。在强对偶性成立时，将拉格朗日函数分别对原变量和对偶变量求导，再并令导数等于零，即可得到原变量与对偶变量的数值关系。于是，对偶问题解决了，主问题也就解决了。

那么对偶问题如何解决呢？由于svm的目标函数、不等式约束条件是凸函数，所以KKT条件是对偶问题最优解的充分必要条件，找到满足KKT条件的解就能解决对偶问题。

至此，就说明了为什么满足KKT条件的$\boldsymbol{w^*}, \boldsymbol{b^*}, \boldsymbol{\alpha^*}$，这里的$\boldsymbol{w^*}, \boldsymbol{b^*}$就是SVM原始问题的最优解。所以，就可以按照这样一个思想来求最优解：如果所有变量的解满足此最优化问题的KKT条件，那么这个最优化问题的解就得到了。