

# Support Vector Machine 05 Slate & KKT Condition

Chen Gong

18 November 2019

首先，我们整理一下前面得到的有关约束优化的模型。我们可以描述为：

$$\begin{cases} \min f(x) \\ \text{s.t. } m_i(x) \leq 0, i = 1, 2, \dots, M \\ n_j(x) = 0, j = 1, 2, \dots, N \end{cases} \quad (1)$$

其中，

$$D = \left\{ \text{dom } f \bigcap_{i=1}^M \text{dom } m_i \bigcap_{j=1}^N \text{dom } n_j \right\} \quad (2)$$

我们将模型进行简化可得：

$$\begin{cases} \min f(x) \\ \text{s.t. } m_i(x) \end{cases} \implies G = \{(m, f) | x \in D\} = \{(\mu, t) | x \in D\} \quad (3)$$

那么，我们的优化目标为：

$$p^* = \inf\{t | (\mu, t) \in G, \mu \leq 0\} \quad (4)$$

$$g(\lambda) = \inf\{t + \lambda\mu | (\mu, t) \in G\} \quad (5)$$

通常来说，凸优化问题，不一定是强对偶问题。往往都是凸优化问题需要加上一些限定条件才可以构成强对偶问题。比如说 slate condition，但是这些条件往往都是充分非必要的。这样的条件有很多种，slate condition 只是其中一种，类似的还有 KKT condition。

## 1 Slate Condition

下面简述一下 Slate Condition，详细的证明过程就不做过多的描述。也就是  $\exists \hat{x} \in \text{relint } D, \text{ s.t. } \forall i = 1, 2, \dots, m, m_i(\hat{x}) \leq 0$ 。而 relint 的意思就是，relative interior，相对内部的意思。

而对于绝大部分的凸优化问题，通常 Slate 条件是成立的。而放松的 Slate 条件为：假设  $M$  中有  $k$  个仿射函数， $M - k$  个仿射。而 SVM 是一个典型的凸二次规划问题，也就是目标函数  $f$  是凸函数， $m_i$  是仿射函数， $n_j$  为仿射。那么在几何上是什么意思呢？也就是限制至少有一个点在坐标系的左边，限制直线不是垂直的，这里需要结合 Support Vector Machine 04 中的几何解释来看。

## 2 KKT Condition

在上文中我们知道了 Convex 和 Slater Condition 可以得到强对偶关系，也就是  $d^* = p^*$ 。但是这只是一个充分非必要条件。同样的在满足 KKT Condition 的情况下，我们也可以得出是一个强对偶问题，并且这是一个充分必要的条件。

我们在来回顾一下模型的原问题：

$$\begin{cases} \min f(x) \\ \text{s.t. } m_i(x) \leq 0, i = 1, 2, \dots, M \\ n_j(x) = 0, j = 1, 2, \dots, N \end{cases} \quad (6)$$

而拉格朗日形式的表达为：

$$\mathcal{L}(x, \lambda) = f(x) + \sum_i \lambda_i m_i(x) + \sum_j \eta_j n_j(x) \quad (7)$$

对于对偶问题，我们可以描述对应的  $g(\lambda, \eta) = \min_x \mathcal{L}(x, \eta, \lambda)$ ； $d^* \leftarrow \lambda^*, \eta^*$ 。所以对偶问题 (Dual Prob) 也就是：

$$\begin{cases} \max_{\lambda, \eta} g(\lambda, \eta) \\ \text{s.t. } \lambda_i \geq 0, i = 1, 2, \dots, M \end{cases} \quad (8)$$

下面进行 KKT 条件的推导：

首先一定需要满足的是，在可行域以内。所以，一定会有： $m_i(x^*) \leq 0, n_i(x^*) = 0, \lambda^* \geq 0$ 。并且还需要满足：

$$\begin{aligned} d^* &= \max_{\lambda, \eta} g(\lambda, \eta) = g(\lambda^*, \eta^*) \\ &= \min_x \mathcal{L}(x, \lambda^*, \eta^*) \\ &\leq \mathcal{L}(x, \lambda^*, \eta^*), \quad \forall x \in D \\ &= \mathcal{L}(x^*, \lambda^*, \eta^*) \\ &= f(x^*) + \sum_i \lambda_i m_i(x^*) + \sum_j \eta_j n_j(x^*) \\ &= f(x^*) + \sum_i \lambda_i m_i(x^*) \end{aligned} \quad (9)$$

上式中的  $f(x^*)$  也就是  $p^*$ ，用因为  $\lambda_i m_i(x^*) \leq 0$  是必然存在的。所以， $d^* \leq f(x^*)$ 。这就是弱对偶关系，如果是强对偶关系，就需要我们需要在两个小于或等于号那取等才行。

第一，对于  $\forall i = 0, 1, 2, \dots, M$ ，都有  $\sum_i \lambda_i m_i = 0$ 。

第二， $\min \mathcal{L}(x, \lambda^*, \eta^*), \quad \forall x \in D = \mathcal{L}(x^*, \lambda^*, \eta^*)$ 。也就是：

$$\frac{\partial \mathcal{L}(x, \lambda^*, \eta^*)}{\partial x} \Big|_{x=x^*} = 0 \quad (10)$$

所以，KKT 条件就已经完成了，我们总结一下，KKT 条件分成 3 个部分。

1. 可行条件：也就是需要满足定义域的条件， $m_i(x^*) \leq 0, n_i(x^*) = 0, \lambda^* \geq 0$ 。

2. 互补松弛条件： $\lambda_i m_i = 0$ 。

3. 梯度为零： $\frac{\partial \mathcal{L}(x, \lambda^*, \eta^*)}{\partial x} \Big|_{x=x^*} = 0$ 。

我们可以对比之前学习的 SVM 的 KKT 条件。