

一、阅读范围

统计学习方法第七章 支持向量机

二、负责讲解部分主要知识梳理

7.4.1 节主要介绍的是 SMO 算法的一个子问题——两个变量二次规划的求解方法

假设选择的两个变量是 α_1, α_2 ，其他变量 $\alpha_i (i = 3, 4, \dots, N)$ 是固定的。

这个时候 SMO 算法的最优化问题也就是求解两变量的二次规划问题：

$$\begin{aligned} \min_{\alpha_1, \alpha_2} W(\alpha_1, \alpha_2) = & \frac{1}{2} K_{11} \alpha_1^2 + \frac{1}{2} K_{22} \alpha_2^2 + y_1 y_2 K_{12} \alpha_1 \alpha_2 - (\alpha_1 + \alpha_2) + y_1 \alpha_1 \sum_{i=3}^N y_i a_i K_{i1} \\ & + y_2 \alpha_2 \sum_{i=3}^N y_i a_i K_{i2} \end{aligned} \quad (7.101)$$

$$s.t. \alpha_1 y_1 + \alpha_2 y_2 = - \sum_{i=3}^N y_i \alpha_i = \zeta \quad (7.102)$$

$$0 \leq \alpha_i \leq C, i = 1, 2 \quad (7.103)$$

因为只有两个变量且有式 7.103，故有图 7.8 所示的图示。

这里解释一下为什么会有图示这两个约束，首先由式 7.102，等式右边为常数。 α_1, α_2 是两个变量，而 $y_i \in \{+1, -1\}$ ，所以 y_1, y_2 的关系只有两种：同号与异号。

由于有 7.102 的约束条件，所以只需先对其中一个变量（这里是 α_2 进行讨论），并结合 7.103，7.102 和初始解确定最优值的取值范围。

然后书本给出了定理 7.6 来描述由不考虑 7.103 约束时的最优解到满足 7.103 得到的最终最优解。

三、问题

1. 在软间隔最大化学习方法中 b 值有一个范围，这个范围如何求出？

书本中介绍到线性不可分的线性支持向量机的学习问题最终转化为一个凸二次规划问题（原始问题）：

$$\begin{aligned} \min_{\omega, b, \xi} & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N \xi_i \\ s.t. & y_i(\omega \cdot x + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, N \end{aligned}$$

可以证明 ω 的解唯一，但 b 的解不唯一，是一个区间。书上没有给出对应的求解方法，而是给出了相关文献，这个留在后续的研读过程中进行学习。

2. KKT 条件违反严重是如何判断的？

由 (7.111) ~ (7.113) 给出：

$$\alpha_i = 0 \Leftrightarrow y_i g(x_i) \geq 1 \quad (7.111)$$

$$0 < \alpha_i < C \Leftrightarrow y_i g(x_i) = 1 \quad (7.112)$$

$$\alpha_i = C \Leftrightarrow y_i g(x_i) \leq 1 \quad (7.113)$$

3. 最大化 $1/\|\omega\|$ 与 最小化 $0.5\|\omega\|^2$ 等价，为什么一定要取 $0.5\|\omega\|^2$ ，而不是 $\|\omega\|$ ？

这里是为了简化计算，为了在构造拉格朗日函数求导后去掉多余常数。并且为了保证一阶求导后可以求出 $\|\omega\|$ 的表达式子。

4. 7.4.2 中新的 b 值的求解。

书本上的推导过程的意图是利用 $E_1 b_1^{old}$ 、 $E_2 b_2^{old}$ 和定义式、KKT 条件分别更新 b_1^{new} 和 b_2^{new} 。并规定了当 α_1^{new} 、 α_2^{new} 是 0 或者 C，选择 b_1^{new} 和 b_2^{new} 的中点作为 b^{new} 。

四、下周预计读书进度

第八章 提升方法