

SMO算法

SMO算法要解如下凸二次规划的对偶问题：

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i^t x_j - \sum_{i=1}^N \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^N \alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C \quad (i \in [N]) \end{aligned} \quad (5.1)$$

在这个问题中，变量是拉格朗日乘子，一个变量 α_i 对应于一个样本点 (x_i, y_i) ；变量的总数等于训练样本容量 N 。

SMO算法是一种启发式算法，其基本思路是：如果所有变量的解都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。因为KKT条件是该最优化问题的充分必要条件。SMO算法包括两个部分：求解两个变量二次规划的解析方法和选择变量的启发式方法。

两个变量二次规划的求解方法

不失一般性，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。于是SMO的最优化问题5.1的子问题可以写成：

$$\begin{aligned} \min_{\alpha_1, \alpha_2} W(\alpha_1, \alpha_2) = & \frac{1}{2} K_{11} \alpha_1^2 + \frac{1}{2} K_{22} \alpha_2^2 + y_1 y_2 K_{12} \alpha_1 \alpha_2 - (\alpha_1 + \alpha_2) \\ & + y_1 \alpha_1 \sum_{i=3}^N y_i \alpha_i K_{i1} + y_2 \alpha_2 \sum_{i=3}^N y_i \alpha_i K_{i2} \end{aligned} \quad (5.2)$$

$$\text{s.t.} \quad \alpha_1 y_1 + \alpha_2 y_2 = - \sum_{i=3}^N y_i \alpha_i = \zeta \quad (5.3)$$

$$0 \leq \alpha_i \leq C, \quad i = 1, 2 \quad (5.4)$$

其中， $K_{ij} = K(x_i, x_j), i, j = 1, 2, \dots, N, \zeta$ 是常数，目标函数5.2省略了不含 α_1, α_2 的常数项。

根据式5.3，将 α_1 用 α_2 表示，最终得到只有 α_2 的目标函数，然后对 α_2 求偏导，并置为零，最终得沿着约束方向未经剪辑时的解是：

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2 (E_1 - E_2)}{\eta}$$

$$\text{其中, } E_i = \left(\sum_{j=1}^N \alpha_j y_j K(x_j, x_i) + b \right) - y_i, i = 1, 2$$

$$\eta = K_{11} + K_{22} - 2K_{12} = \|\Phi(x_1) - \Phi(x_2)\|^2 \quad (5.5)$$

由于不等式约束5.4的存在，再加上等式5.3的约束，最终可得 α_2^{new} 的取值范围可以表示为：

$$\begin{cases} L = \max(0, \alpha_2^{old} - \alpha_1^{old}), H = \min(C, C + \alpha_2^{old} - \alpha_1^{old}), & \text{if } y_1 \neq y_2 \\ L = \max(0, \alpha_2^{old} + \alpha_1^{old} - C), H = \min(C, \alpha_2^{old} + \alpha_1^{old}), & \text{if } y_1 = y_2 \end{cases} \quad (5.6)$$

所以经剪辑后 α_2 的解是：

$$\alpha_2^{new} = \begin{cases} H, & \alpha_2^{new,unc} > H \\ \alpha_2^{new,unc}, & L \leq \alpha_2^{new,unc} \leq H \\ L, & \alpha_2^{new,unc} < L \end{cases} \quad (5.7)$$

由 α_2^{new} 求得 α_1^{new} 是： $\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_1^{old})$

变量的选择方法

SMO算法在每个子问题中选择两个变量优化，其中至少一个变量是违反KKT条件的。

第一个变量的选择

SMO称选第一个变

量的过程为外层循环。外层循环在训练样本中选择违反KKT条件最严重的样本 (x_i, y_i) 是否满足KKT条件，即：

![5-1 KKT条件](images/5-1 KKT条件.png)

第2个变量的选择

SMO称选择第二个变量的过程为内层循环。假设在外层循环中已经找到第一个变量 α_1 ，现在要在内层循环中找到第二个变量 α_2 。第二个变量选择的标准是希望能使 α_2 有足够大的变化。由式5.5, 5.7可知， α_2^{new} 依赖于 $|E_1 - E_2|$ 的，为了加快速度，选择最大的 $|E_1 - E_2|$ 对应的 α_2 即可。如果能使目标函数有足够下降的 α_2 ，则更换 α_2 。

SMO算法：

![5-2 SMO算法](images/5-2 SMO算法.png)