SMO算法

SMO算法要解如下凸二次规划的对偶问题：

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^t x_j - \sum_{i=1}^{N}\alpha_i$$
$$s.t.\ \ \sum_{i=1}^{N}\alpha_i y_i = 0,$$
$$0 \le \alpha_i \le C\ (i \in [N]) \tag{5.1}$$

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$\alpha_i (i = 3, 4, ..., N)$是固定的。于是SMO的最优化问题5.1的子问题可以写成：

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

$$s.t.\qquad \alpha_1 y_1 + \alpha_2 y_2 = -\sum_{i=3}^{N} y_i\alpha_i = \zeta \tag{5.3}$$

$$0 \le \alpha_i \le C,\ \ i = 1, 2 \tag{5.4}$$

其中，$K_{ij} = K(x_i, x_j), i, j = 1, 2, ..., N, \zeta$是常数，目标函数5.2省略了不含$\alpha_1, \alpha_2$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：

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$
$$\text{其中，}\quad E_i = (\sum_{j=1}^{N}\alpha_j y_j K(x_j, x_i) + b) - y_i, i = 1, 2$$

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

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：

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

所以经剪辑后α_2的解是：

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

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
(x_i,y_i)$是否满足KKT条件，即：

$$\alpha_i = 0 \Leftrightarrow y_i g(x_i) \geqslant 1 \tag{7.111}$$

$$0 < \alpha_i < C \Leftrightarrow y_i g(x_i) = 1 \tag{7.112}$$

$$\alpha_i = C \Leftrightarrow y_i g(x_i) \leqslant 1 \tag{7.113}$$

其中，$g(x_i) = \sum_{j=1}^{N} \alpha_j y_j K(x_i, x_j) + b$。

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α2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：

算法 7.5（SMO 算法）

输入：训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$，其中，$x_i\in\mathcal{X}=\mathbf{R}^n$，$y_i\in\mathcal{Y}=\{-1,+1\}$，$i=1,2,\cdots,N$，精度 ε；

输出：近似解 $\hat{\alpha}$。

（1）取初值 $\alpha^{(0)}=0$，令 $k=0$；

（2）选取优化变量 $\alpha_1^{(k)},\alpha_2^{(k)}$，解析求解两个变量的最优化问题 (7.101)~(7.103)，求得最优解 $\alpha_1^{(k+1)},\alpha_2^{(k+1)}$，更新 α 为 $\alpha^{(k+1)}$；

（3）若在精度 ε 范围内满足停机条件

$$\sum_{i=1}^{N}\alpha_i y_i=0,\quad 0\leqslant\alpha_i\leqslant C,\quad i=1,2,\cdots,N$$

$$y_i\cdot g(x_i)\begin{cases}\geqslant 1, & \{x_i|\alpha_i=0\}\\ =1, & \{x_i|0<\alpha_i<C\}\\ \leqslant 1, & \{x_i|\alpha_i=C\}\end{cases}$$

其中，

$$g(x_i)=\sum_{j=1}^{N}\alpha_j y_j K(x_j,x_i)+b$$

则转 (4)；否则令 $k=k+1$，转 (2)；

（4）取 $\hat{\alpha}=\alpha^{(k+1)}$。 ■