序列最小最优化算法

Author： 中山大学 17数据科学与计算机学院 YSY

https://github.com/ysyisyourbrother

SMO算法要解如下凸二次规划的对偶问题

$$
\begin{aligned}
\min_{\alpha} \ & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_k K(x_i,x_j)-\sum_{i=1}^{N}\alpha_i \\
s.t. \ & \sum_{i=1}^{N}\alpha_i y_i = 0 \\
& 0\le \alpha_i \le C, i=1,2,\dots,N
\end{aligned}
$$

基本思路：由于KKT条件是该最优化问题的充要条件，如果所有变量α都满足最优化问题的KKT条件，那么最优化的解就得到了。否则，利用启发式算法选择两个变量， 固定其他变量，构建一个二次规划问题，使得关于这两个变量的解更接近SMO问题的解。

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。构建一个二次规划问题

$$
\begin{aligned}
\min_{\alpha_1,\alpha_2} W(\alpha_1,\alpha_2) = & \frac{1}{2}K_{11}\alpha_1^2+\frac{1}{2}K_{22}\alpha_2^2+y_1y_2K_{12}\alpha_1\alpha_2 \\
& -(\alpha_1+\alpha_2)+y_1\alpha_1\sum_{i=3}^{N}y_i\alpha_iK_{i1}+y_2\alpha_2\sum_{i=3}^{N}y_i\alpha_iK_{i2} \\
s.t. \quad & \alpha_iy_i+\alpha_2y_2=-\sum_{i=3}^{N}y_i\alpha_i=constant \\
& 0\le\alpha_i\le C, i=1,2
\end{aligned}
$$

根据约束条件，目标函数是一条平行于对角线的线段上最优值。换言之，该问题实质是一个单变量的优化问题。考虑对α_2进行优化，记沿着约束方向未经过剪辑时α_2的最优解为$\alpha_2^{new,unc}$。根据约束条件，它的取值范围满足

$$
L\le\alpha_2^{new}\le H
$$

其中，L,H是目标函数所在线段与边界共同构成的界。

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

记函数$g(x)$为预测函数

$$
g(x)=\sum_{i=1}^{N}\alpha_iy_iK(x_i,x)+b
$$

记E_i为对输入x_i的预测值与真实值y_i的差。

$$
E_i=g(x_i)-y_i
$$

根据定理7.6

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

其中，$\eta = K_{11} + K_{22} - 2K_{12}$

剪辑后，

$$\alpha_2^{new} = \begin{cases} H, & \alpha_2^{new,unc} > H \\ \alpha_2^{new,unc}, & L \le \alpha_2^{new,unc} \le H \\ L, & \alpha_2^{new,unc} < L \end{cases}$$

随之

$$\alpha_1^{new} = \alpha_1^{old} + y_1 y_2(\alpha_2^{old} - \alpha_2^{new})$$

$$\begin{aligned} b_1^{new} &= y_1 - \sum_{i=3}^{N} \alpha_i y_i K_{i1} - \alpha_1^{new} y_1 K_{11} - \alpha_2^{new} y_2 K_{21} \\ &= -E_1 - y_1 K_{11}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{21}(\alpha_2^{new} - \alpha_2^{old}) + b^{old} \\ b_2^{new} &= -E_2 - y_1 K_{12}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{22}(\alpha_2^{new} - \alpha_2^{old}) + b^{old} \\ E_i^{new} &= \sum_S y_j \alpha_j K_{ij} + b^{new} - y_i \end{aligned}$$

其中，S是所有支持向量x_j的集合

变量的选择方法

第一个变量的选择

选择违反KKT条件最严重的样本点。

$$\begin{aligned} \alpha_i &= 0 \Leftrightarrow y_i g(x_i) \ge 1 \\ 0 < \alpha_i &< C \Leftrightarrow y_i g(x_i) = 1 \\ \alpha_i &= C \Leftrightarrow y_i g(x_i) \le 1 \end{aligned}$$

具体来说，首先遍历虽有满足条件$0 \le \alpha_i < C$的样本点，即在间隔边界上的支持向量点。如果这些样本点都满足KKT条件，那么遍历整个训练集，检验它们是否满足KKT条件。

第二个变量的选择

由于α_2^{new}依赖于$|E_1 - E_2|$，需要寻找使得差值最大的α，为了节省计算时间，可以把所有E_i的值保存在一个列表中。