参考：① bilibili 《快速理解SMO算法》
② 《统计学习方法》 李航.
笔记记录者：Little H

序列最小优化算法 SMO (Sequential Minimal Optimization)

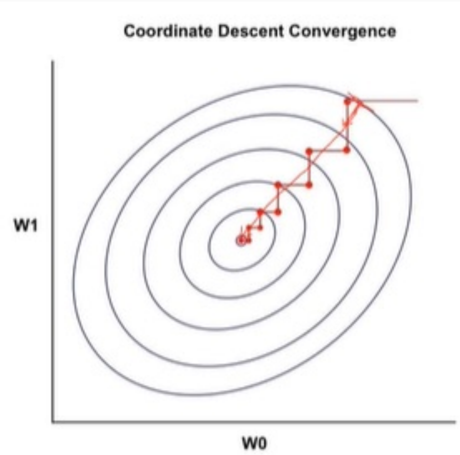
解如下凸二次规划的对偶问题

$$\begin{cases}\min\limits_{\alpha} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j)-\sum_{i=1}^{N}\alpha_i \\ s.t. \quad \sum_{i=1}^{N}\alpha_i y_i = 0 \\ 0\le\alpha_i\le C, \quad i=1,2,\cdots,N\end{cases}$$

整体更新，复杂度会非常高

Note：变量是拉格朗日乘子 α_i，一个对应一个样本.

坐标下降



For example, on iteration k we select a variable j_k and set

$$w_{j_k}^{k+1} = w_{j_k}^{k} - \alpha_k \nabla_{j_k} f(w^k),$$

a gradient descent step on coordinate j_k (other w_j stay the same).

在传统的梯度下降中，我们寻找每个等高线的切线位置（梯度）。
对每个参数求偏，得到整体导数，会比较复杂，考虑使用坐标下降.
坐标下降：每一次只在一个坐标轴上修正.

序列最小最优化算法

- SVM的对偶问题是有约束二次凸优化问题
- 每次最少要调整两个变量

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j)-\sum_{i=1}^{N}\alpha_i \downarrow \downarrow \alpha_1 + \nabla$$

$$s.t. \quad \boxed{\sum_{i=1}^{N}\alpha_i y_i = 0} \qquad \alpha_1 \qquad \alpha_1 + \underbrace{\alpha_2 + \cdots + \alpha_N}=0$$

$$0 \le \alpha_i \le C, \quad i=1,2,...,N$$

$\alpha_1+\alpha_2+\cdots+\alpha_N=0$

若只对α_1做修正，其余不变，则导致 $\alpha_1+\alpha_2+\cdots+\alpha_n\neq 0$

因此，不能一次只更新一个α，每次至少要调整2个变量。

$\alpha_k\ (k\neq i,j)$: fixed　　　　α_i, α_j

以上，凸优化问题被转化为子问题。

两个变量二次规划的求解过程

- 选择两个变量，其它变量固定
- SMO将对偶问题转化成一系列子问题：

$$\min_{\alpha_1,\alpha_2} \quad W(\alpha_1,\alpha_2)=\frac{1}{2}K_{11}\alpha_1^2+\frac{1}{2}K_{22}\alpha_2^2+y_1y_2K_{12}\alpha_1\alpha_2$$
$$-(\alpha_1+\alpha_2)+y_1\alpha_1\sum_{i=3}^{N}y_i\alpha_iK_{i1}+y_2\alpha_2\sum_{i=3}^{N}y_i\alpha_iK_{i2}$$
$$s.t.\quad \alpha_1y_1+\alpha_2y_2=-\sum_{i=3}^{N}y_i\alpha_i=\zeta$$
$$0\le\alpha_i\le C,\ i=1,2$$

$\alpha_1 y_1 = -\alpha_2 y_2 + \zeta$

$\alpha = \alpha_2 y_1 y_2 + \zeta y_1$

因此，优化问题可以有解析解！

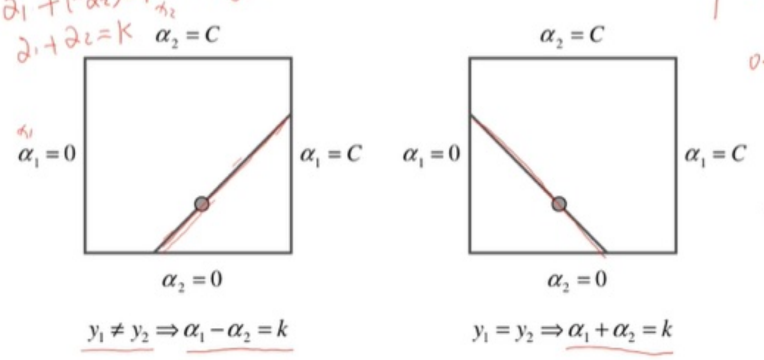
两个变量二次规划的求解过程

- 两个变量，约束条件用二维空间中的图形表示

$a_1y_1+a_2y_2=K$

$a_1+(-a_2)=K=K'\cdot y_1$

$a_1+a_2=K$



$0<\alpha_2<C$

$x=a$

$x\in[b,c]$

$x=c$ 最小

$x=e$

$\alpha_1y_1+\alpha_2y_2=k$

① $y_1\neq y_2$:

两边同乘以y_1，则

$\alpha_1\cdot y_1^2+\alpha_2\cdot y_2\cdot y_1=k\cdot y_1$

$\alpha_1-\alpha_2=k'=ky_1$

② $y_1 = y_2$:

两边同乘以 y_1，则

$\alpha_1 \cdot y_1^2 + \alpha_2 \cdot y_1 y_2 = k y_1$

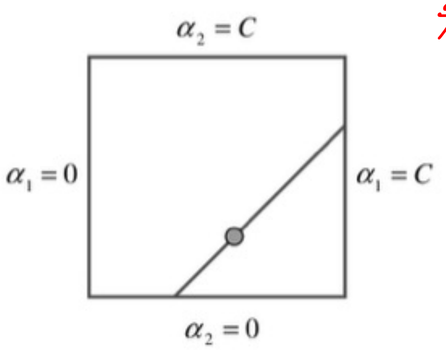
$\alpha_1 + \alpha_2 = k' = k y_1$

两个变量二次规划的求解过程

- 根据不等式条件 α_2^{new} 的取值范围：

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

$$L = \max(0, \alpha_2^{old} - \alpha_1^{old}) \quad H = \min(C, C + \alpha_2^{old} - \alpha_1^{old})$$



$y_1 \ne y_2 \Rightarrow \boxed{\alpha_1 - \alpha_2 = k}$

考虑 Case 1：

$\alpha_1 - \alpha_2 = k$，

$\Rightarrow \alpha_2 = \alpha_1 - k$，$\alpha_1 \in [0, C]$

则 $\alpha_1 - k \in [-k, C-k]$

又 $k = \alpha_1 - \alpha_2$，即 $\alpha_2 = \alpha_1 - 1$

$\therefore \begin{cases} \alpha_2 \in [\alpha_2 - \alpha_1, C + \alpha_2 - \alpha_1] \\ \alpha_2 \in [0, C] \end{cases}$

下界 L　　上界 H

$L = \max(0, \alpha_2^{old} - \alpha_1^{old})$

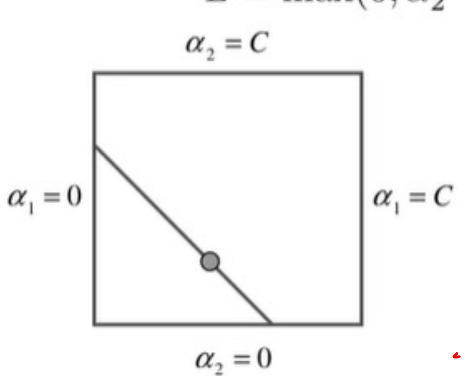
$H = \min(C, C + \alpha_2^{old} - \alpha_1^{old})$

两个变量二次规划的求解过程

- 根据不等式条件 α_2^{new} 的取值范围：

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

$$L = \max(0, \alpha_2^{old} + \alpha_1^{old} - C) \quad H = \min(C, \alpha_2^{old} + \alpha_1^{old})$$



$y_1 = y_2 \Rightarrow \boxed{\alpha_1 + \alpha_2 = k}$

Case 2： $\alpha_1 + \alpha_2 = k$

$\Rightarrow \alpha_2 = k - \alpha_1$

$\alpha_1 \in [0, C]$

$k - \alpha_1 \in [k, k - C]$

$\therefore \begin{cases} \alpha_2 \in [k - C, k]，即 \alpha_2 \in [\alpha_1 + \alpha_2 - C, \alpha_1 + \alpha_2] \\ 又\ \alpha_2 \in [0, C] \end{cases}$

$L = \max(0, \alpha_1^{old} + \alpha_2^{old} - C)$，$H = \min(C, \alpha_1^{old} + \alpha_2^{old})$

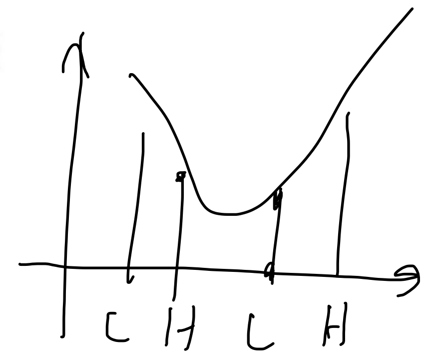
通过函数优化获得

$\alpha_1^{new,un}$，$\alpha_2^{new,un}$

un：未经剪辑

两个变量二次规划的求解过程

- 求解过程：
- 先求沿着约束方向未经剪辑时的 $\alpha_2^{new,unc}$
- 再求剪辑后的 α_2^{new}



- 记： $g(x) = \sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b$
- 令： $E_i = g(x_i) - y_i = \left(\sum_{j=1}^{N} \alpha_j y_j K(x_j, x_i) + b\right) - y_i, \quad i = 1, 2$
- E 为输入x的预测值和真实输出y的差，$i = 1,2$

两个变量二次规划的求解过程

引进记号

$$v_i = \sum_{j=3}^{N} \alpha_j y_j K(x_i, x_j) = g(x_i) - \sum_{j=1}^{2} \alpha_j y_j K(x_i, x_j) - b, \; i = 1, 2$$

目标函数写成：

$$W(\alpha_1, \alpha_2) = \frac{1}{2} K_{11} \alpha_1^2 + \frac{1}{2} K_{22} \alpha_2^2 + y_1 y_2 K_{12} \alpha_1 \alpha_2$$

前面：$\alpha_1 y_1 + \alpha_2 y_2 = \zeta$

$$-(\alpha_1 + \alpha_2) + y_1 v_1 \alpha_1 + y_2 v_2 \alpha_2$$

由 $\alpha_1 y_1 = \zeta - \alpha_2 y_2$ 及 $y_i^2 = 1$

$$\alpha_1 = (\zeta - y_2 \alpha_2) y_1$$

两个变量二次规划的求解过程

- 得到只是α_2 的函数的目标函数

$$W(\alpha_2)=\frac{1}{2}K_{11}(\zeta-\alpha_2y_2)^2+\frac{1}{2}K_{22}\alpha_2^2+y_2K_{12}(\zeta-\alpha_2y_2)\alpha_2$$
$$-(\zeta-\alpha_2y_2)y_1-\alpha_2+v_1(\zeta-\alpha_2y_2)+y_2v_2\alpha_2$$

- 对α_2求导：$\frac{\delta W}{\delta\alpha_2}=K_{11}\alpha_2+K_{22}\alpha_2-2K_{12}\alpha_2-K_{11}\zeta y_2+K_{12}\zeta y_2+y_1y_2-1-v_1y_2+y_2v_2$

- 令其为0：

$$(K_{11}+K_{22}-2K_{12})\alpha_2=y_2(y_2-y_1+\zeta K_{11}-\zeta K_{12}+v_1-v_2)$$
$$=y_2[y_2-y_1+\zeta K_{11}-\zeta K_{12}+(g(x_1)-\sum_{j=1}^{2}y_j\alpha_jK_{1j}-b)-(g(x_2)-\sum_{j=1}^{2}y_j\alpha_jK_{2j}-b)]$$

两个变量二次规划的求解过程

- 将$\zeta=\alpha_1^{old}y_1+\alpha_2^{old}y_2$代入：

$$(K_{11}+K_{22}-2K12)\alpha_2^{new,unc}=y_2((K_{11}+K_{22}-2K_{12})\alpha_2^{old}y_2+y_2-y_1+g(x_1)-g(x_2)))$$
$$=(K_{11}+K_{22}-2K_{12})\alpha_2^{old}+y_2(E_1-E_2)$$

- 将$\eta=K_{11}+K_{22}-2K_{12}$代入：

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

两个变量二次规划的求解过程

- 得到：
- 最优化子问题沿约束方向未经剪辑的解：

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

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

- 剪辑后的解

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

- 得到α_1的解 $\alpha_1^{new} = \alpha_1^{old} + y_1 y_2(\alpha_2^{old} - \alpha_2^{new})$

计算阈值b和E_i

完成两个变量的优化后，重新计算b，E_i

由KKT条件，如果 $0 < \alpha_1^{new} < C$

所对应的样例是一个支持向量，落在分割边界上。$y_1^2 \cdot g(x_1) = 1 \cdot y_1$

$g(x_1) =$

$$\sum_{i=1}^{N} \alpha_i y_i K_{i1} + b = y_1$$

$g(x_1) = y_1$

$$b_1^{new} = y_1 - \sum_{i=3}^{N} \alpha_i y_i K_{i1} - \alpha_1^{new} y_1 K_{11} - \alpha_2^{new} y_2 K_{21}$$

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

$$E_1 = \sum_{i=3}^{N} \alpha_i y_i K_{i1} + \alpha_1^{old} y_1 K_{11} + \alpha_2^{old} y_2 K_{21} + b^{old} - y_1$$

计算阈值b和E_i

$$b_1^{new} = \boxed{y_1 - \sum_{i=3}^{N} \alpha_i y_i K_{i1}} - \alpha_1^{new} y_1 K_{11} - \alpha_2^{new} y_2 K_{21}$$

$$E_1 = \sum_{i=3}^{N} \alpha_i y_i K_{i1} + \alpha_1^{old} y_1 K_{11} + \alpha_2^{old} y_2 K_{21} + b^{old} - y_1$$

$$y_1 - \sum_{i=3}^{N} \alpha_i y_i K_{i1} = -E_1 + \alpha_1^{old} y_1 K_{11} + \alpha_2^{old} y_2 K_{21} + b^{old}$$

$$b_1^{new} = -E_1 - y_1 K_{11}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{21}(\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

计算阈值b和E_i

- 如果:

$$0 < \alpha_2^{new} < C$$

$$b_2^{new} = -E_2 - y_1 K_{12}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{22}(\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

$$E_i^{new} = \sum_{S} y_j \alpha_j K(x_i, x_j) + b^{new} - y_i$$

- S是所有支持向量x_j的集合

如果 $\alpha_1^{\text{new}}, \alpha_2^{\text{new}}$ 同时满足条件 $0 < \alpha_i^{\text{new}} < C$，$i = 1, 2$，那么 $b_1^{\text{new}} = b_2^{\text{new}}$。如果 $\alpha_1^{\text{new}}, \alpha_2^{\text{new}}$ 是 0 或者 C，那么 b_1^{new} 和 b_2^{new} 以及它们之间的数都是符合 KKT 条件的阈值，这时选择它们的中点作为 b^{new}。

变量的启发式选择

- 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

非常关键，怎么选择的

1、第一个变量的选择：外循环

- 违反KKT最严重的样本点，
- 检验样本点是否满足KKT条件：

$$\alpha_i = 0 \leftrightarrow y_i g(x_i) \geq 1$$
$$0 < \alpha_i < C \leftrightarrow y_i g(x_i) \leq 1$$
$$\alpha_i = C \leftrightarrow y_i g(x_i) \leq 1$$
$$g(x_i) = \sum_{j=1}^{N} \alpha_j y_j K(x_i, x_j) + b$$

变量的选择方法

$$\alpha_2^{new} = \alpha_1^{old} + \frac{y_1(E_1 - E_2)}{\eta}$$

2、第二个变量的检查：内循环，

- 选择的标准是希望能使目标函数有足够大的变化
  - 即对应 $|E_1 - E_2|$ 最大，即E_1，E_2 的符号相反，差异最大
- 如果内循环通过上述方法找到的点不能使目标函数有足够的下降
- 则：遍历间隔边界上的样本点，测试目标函数下降
  - 如果下降不大，则遍历所有样本点
  - 如果依然下降不大，则丢弃外循环点，重新选择

SMO算法

- 输入：训练数据集$T=\{(x_1,y_1),(x_2,y_2),...,(x_N,y_N)\}$

  $x_i \in \mathcal{X}=R^n, y_i \in \mathcal{Y}=\{+1,-1\}, i=1,2,...,N$，精度ε
- 输出：近似解 α

(1)取初值$\alpha^{(0)}=0$，令 $k=0$

(2)选取优化变量α_1^k,α_2^k，解析求解两个变量的最优化问题，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 即KKT条件

$$\sum_{i=1}^{N}\alpha_i y_i=0$$

$$0\le\alpha_i\le C,\quad i=1,2,...,N$$

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

则转(4);否则令k=k+1,转(2);

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

(4)取$\hat{\alpha}=\alpha^{(k+1)}$

SMO算法

启发式算法，基本思路：

- 确定如果所有变量的解都满足此最优化问题的KKT条件，那么得到解；
- 否则，选择两个变量，固定其它变量，针对这两个变量构建一个二次规划问题，称为子问题，可通过解析方法求解，提高了计算速度。
- 子问题的两个变量：一个是违反KKT条件最严重的那个，另一个由约束条件自动

$$\alpha_1=-y_1\sum_{i=2}^{N}\alpha_i y_i$$

SMO算法包括两个部分：

- 求解两个变量二次规划的解析方法
- 选择变量的启发式方法