本文介绍传统机器学习算法中的“支撑向量机（SVM）”，在深度学习变得流行之前，SVM是许多论文中经常会用到的算法。它有非常完备的数学推导，我在刚接触它的时候也被搞得云里雾里。现在打算系统的介绍一下它。

本文一共分成以下几个部分

1. 线性可分支撑向量机及其对偶问题
2. 线性不可分支撑向量机及其对偶问题
3. 非线性支撑向量机（核技巧）
4. SMO算法

文章重点参考了《统计学习方法》的支撑向量机一章。

声明：文章中用到的上下标，上标的标示是括号加数字，例如$\mathbf{x}^{(i)}$表示样本集中第i个样本点，加粗的$\mathbf{x}$表示这是一个向量。下标x_i表示向量$\mathbf{x}$的第i个维度。这种标记方式是follow ng的课程的。

4. SMO序列最小化优化算法

回顾第1或第2节求解分离超平面时，在求解以α向量为变量凸二次规划问题时，我们都是假设通过某种方法找到了$\alpha^\star = \{(\alpha^{(1)})^\star, (\alpha^{(2)})^\star, \cdots, (\alpha^{(N)})^\star\}^T \in \mathbb{R}^N$。可以看到，$\alpha \in \mathbb{R}^N$是一个$N$维的向量，$N$表示训练样本点的个数。训练样本点的个数很大时，一般求解二次规划问题的方法的性能会比较低。SMO算法是一种快速求解$\alpha^\star$的算法。

4.1 算法思路

SMO算法的基本思路：判断一个α向量是否为找到的对偶问题的最优解时，要看它是否满足KKT条件，这是一个充分必要条件。KKT条件是要求α向量中的每一个元素$\alpha^{(i)}$都满足条件。如果不满足KKT条件，就说明当前的α还不是最优解。此时，选择α向量中的两个维度，固定其他维度。针对这两个维度来构建二次规划问题，称为子问题。因为子问题仍然是是二次规划问题，找到的解为全局最优解，所以即使只允许两个维度的变量进行调整，找到的解也要比当前的解要更优、更加接近全局最优解。重要的是，子问题可以通过解析方法来求解，这可以大大提高整个问题的求解速度。选择两个维度的变量的方法是，选择违反KKT条件最严重的一个，另外一个由约束条件确定。在KKT条件中，α向量需要满足

$$\sum_{i=1}^{N} \alpha^{(i)} y^{(i)} = 0$$

因此我们确定的两个维度，实际上只有一个维度是自由变量。假设$\alpha^{(1)}, \alpha^{(2)}$是选定的两个变量，当通过求解二次规划问题得到了$\alpha^{(2)}$后，根据$\alpha^{(1)} = -y^{(1)} \sum_{i=2}^{N} \alpha^{(i)} y^{(i)}$同时可以确定$\alpha^{(1)}$，这样就同时把$\alpha^{(1)}, \alpha^{(2)}$更新了。循环这个过程。

注意，这个算法是一个启发式算法，即最后找到的解可能不是$\alpha^\star$，而是一个近似解$\hat{\alpha}$，需要在算法开始前设定一个精度ϵ。

下面会解释一下怎样求解两个维度的变量的二次规划问题，以及具体怎样选出两个维度的变量。

4.2 $\alpha^{(1)}$和$\alpha^{(2)}$的二次规划求解方法

不失一般性，假设选择$\alpha^{(1)}$和$\alpha^{(2)}$，其余变量固定，标记函数名为$W(\alpha^{(1)}, \alpha^{(2)})$，则对偶问题

$$
\begin{aligned}
\min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha^{(i)}\alpha^{(j)}y^{(i)}y^{(j)}(\mathbf{x}^{(i)}\cdot\mathbf{x}^{(j)})-\sum_{i=1}^{N}\alpha^{(i)} \\
\text{s. t.} \quad & \sum_{i=1}^{N}\alpha^{(i)}y^{(i)}=0 \\
& C\ge\alpha^{(i)}\ge 0, i=1,2,\cdots,N
\end{aligned}
$$

可以写成

$$
\begin{aligned}
\min_{\alpha^{(1)},\alpha^{(2)}} \quad W(\alpha^{(1)},\alpha^{(2)}) &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha^{(i)}\alpha^{(j)}y^{(i)}y^{(j)}(\mathbf{x}^{(i)}\cdot\mathbf{x}^{(j)})-\sum_{i=1}^{N}\alpha^{(i)} \\
&= \frac{1}{2}K(\alpha^{(1)},\alpha^{(1)})(y^{(1)})^2(\alpha^{(1)})^2+\frac{1}{2}K(\alpha^{(2)},\alpha^{(2)})(y^{(2)})^2(\alpha^{(2)})^2+y^{(1)}y^{(2)}K(\alpha^{(1)},\alpha^{(2)})\alpha^{(1)}\alpha^{(2)} \\
&\quad -\alpha^{(1)}-\alpha^{(2)}+y^{(1)}\alpha^{(1)}\sum_{i=3}^{N}y^{(i)}\alpha^{(i)}K(\alpha^{(i)},\alpha^{(1)})+y^{(2)}\alpha^{(2)}\sum_{i=3}^{N}y^{(i)}\alpha^{(i)}K(\alpha^{(i)},\alpha^{(2)}) \\
\text{s. t.}\ \alpha^{(1)}y^{(1)}+\alpha^{(2)}y^{(2)} &= -\sum_{i=3}^{N}\alpha^{(i)}y^{(i)}=\zeta \\
C\ge\alpha^{(i)}\ge 0, i&=1,2,\cdots,N
\end{aligned}
$$

因为在这个问题中的变量仅有$\alpha^{(1)}$和$\alpha^{(2)}$，其余不可调，因此把和$\alpha^{(1)}$和$\alpha^{(2)}$无关的项去掉了，它们是常数项。

先看约束条件。$C\ge\alpha^{(i)}\ge 0, i=1,2,\cdots,N$是一个box constraint，将$\alpha$限制在了一个盒子中，包括$\alpha^{(1)}$和$\alpha^{(2)}$，如下图。

等式约束$\alpha^{(1)}y^{(1)}+\alpha^{(2)}y^{(2)}=\zeta$将$\alpha^{(1)}$和$\alpha^{(2)}$只能在平行于盒子对角线的线段上移动

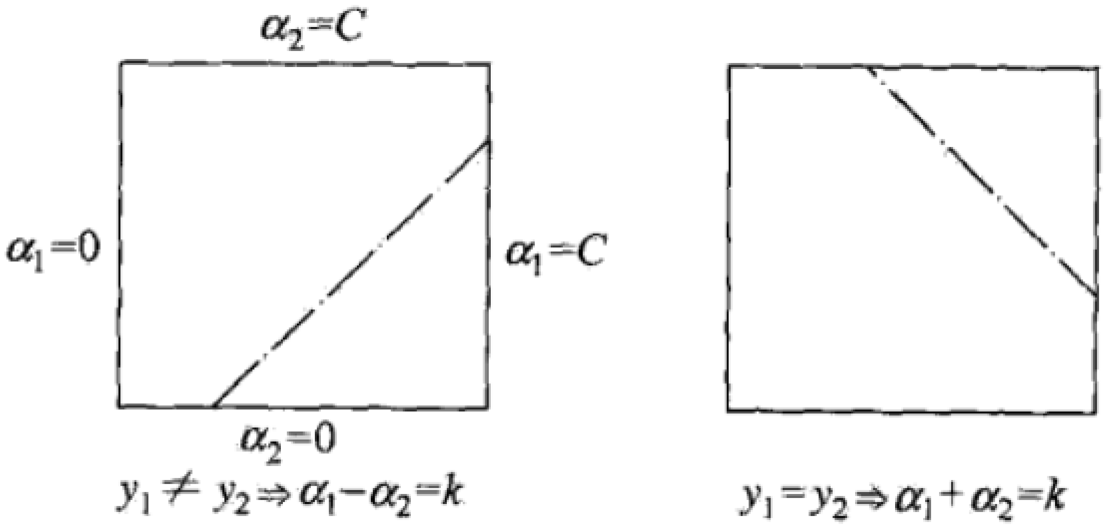

图 7.8　二变量优化问题图示

图中是以$\alpha^{(1)}$为横轴，$\alpha^{(2)}$为纵轴。

左图和右图，区别在于选出的变量$\alpha^{(1)}$和$\alpha^{(2)}$，对应的样本点的标签$y^{(1)}$和$y^{(2)}$是否一致来决定。

当两个标签的符号不同时，对应左图。假设$y^{(1)}=1$，$y^{(2)}=-1$，则等式约束变成$\alpha^{(1)}-\alpha^{(2)}=\zeta$，移项后$\alpha^{(2)}=\alpha^{(1)}-\zeta$，这是一条斜率为正45度，截距为$-\zeta$的直线。当$y^{(1)}=-1$，$y^{(2)}=1$时，只需要提出一个负号，可知斜率不变，截距变成$\zeta$。

当两个标签的符号相同时，对应右图。假设$y^{(1)}=1$，$y^{(2)}=1$，则等式约束变成$\alpha^{(1)}+\alpha^{(2)}=\zeta$，移项后$\alpha^{(2)}=-\alpha^{(1)}+\zeta$，这是一条斜率为负45度，截距为$\zeta$的直线。当$y^{(1)}=-1$，$y^{(2)}=-1$时，只需要提出一个负号，可知斜率不变，截距变成$-\zeta$。

可以看到实际上只需要求解一个变量，另外一个变量就可以根据约束条件得出下面以求解$\alpha^{(2)}$为例。假设原来的变量是$(\alpha^{(1)})^{\mathrm{old}}$和$(\alpha^{(2)})^{\mathrm{old}}$。首先判断$\alpha^{(2)}$。求解出的变量是$(\alpha^{(1)})^{\mathrm{new}}$和$(\alpha^{(2)})^{\mathrm{new}}$。其中要得到$(\alpha^{(2)})^{\mathrm{new}}$，必须要使得经过解析方法求解出的变量满足约束条件，即$L\leq(\alpha^{(2)})^{\mathrm{new}}\leq H$，其中$L$和$H$分别是$(\alpha^{(2)})^{\mathrm{new}}$的上下限。这里再定义，没有经过约束条件剪辑的$\alpha^{(2)}$为$(\alpha^{(2)})^{\mathrm{new,unc}}$。下面讨论$L$和$H$的取值。

当$y^{(1)}\neq y^{(2)}$时，则$L=\max(0,(\alpha^{(2)})^{\mathrm{old}}-(\alpha^{(1)})^{\mathrm{old}}),H=\min(C,C+(\alpha^{(2)})^{\mathrm{old}}-(\alpha^{(1)})^{\mathrm{old}})$。

当$y^{(1)}=y^{(2)}$时，则$L=\max(0,(\alpha^{(2)})^{\mathrm{old}}+(\alpha^{(1)})^{\mathrm{old}}-C),H=\min(C,(\alpha^{(2)})^{\mathrm{old}}+(\alpha^{(1)})^{\mathrm{old}})$。

下面推导求解$(\alpha^{(2)})^{\mathrm{new,unc}}$的解析过程

定义

$$g(\mathbf{x})=\sum_{i=1}^{N}\alpha^{(i)}y^{(i)}K(\mathbf{x}^{(i)},\mathbf{x})+b$$

这是分离超平面用α表达的函数。 定义误差

$$E^{(j)}=g(\mathbf{x}^{(j)})-y^{(j)}=(\sum_{i=1}^{N}\alpha^{(i)}y^{(i)}K(\mathbf{x}^{(i)},\mathbf{x}^{(j)})+b)-y^{(j)},\quad j=1,2$$

这是分离函数对$\mathbf{x}^{(j)}$的预测值与真实值$y^{(j)}$之间的误差。

定义

$$\begin{aligned}\nu^{(j)}&=\sum_{i=3}^{N}\alpha^{(i)}y^{(i)}K(\mathbf{x}^{(i)},\mathbf{x}^{(j)})\\&=g(\mathbf{x}^{(j)})-\sum_{i=1}^{2}\alpha^{(i)}y^{(i)}K(\mathbf{x}^{(i)},\mathbf{x}^{(j)})-b\quad j=1,2\end{aligned}$$

带入$W(\alpha^{(1)},\alpha^{(2)})$，把其中带有从i=3到N的求和项替换，得到

$$\begin{aligned}W(\alpha^{(1)},\alpha^{(2)})=&\frac{1}{2}K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})(\alpha^{(1)})^2+\frac{1}{2}K(\mathbf{x}^{(2)},\mathbf{x}^{(2)})(\alpha^{(2)})^2+y^{(1)}y^{(2)}K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})\alpha^{(1)}\alpha^{(2)}\\&-\alpha^{(1)}-\alpha^{(2)}+y^{(1)}\alpha^{(1)}\nu^{(1)}+y^{(2)}\alpha^{(2)}\nu^{(2)}\end{aligned}$$

因为$\alpha^{(1)}y^{(1)}+\alpha^{(2)}y^{(2)}=\zeta$，两边同时乘以$y^{(1)}$，得到$\alpha^{(1)}(y^{(1)})^2+\alpha^{(2)}y^{(2)}y^{(1)}=\zeta y^{(1)}$。其中$(y^{(1)})^2=1$。整理后得到

$$\alpha^{(1)}=(\zeta-\alpha^{(2)}y^{(2)})y^{(1)}$$

将$\alpha^{(1)}$用$\alpha^{(2)}$来表示，代回$W(\alpha^{(1)},\alpha^{(2)})$中，再根据$(y^{(2)})^2=1$，整理得

$$\begin{aligned}&\frac{1}{2}K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})(\zeta-\alpha^{(2)}y^{(2)})^2+\frac{1}{2}K(\mathbf{x}^{(2)},\mathbf{x}^{(2)})(\alpha^{(2)})^2+y^{(2)}K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})\alpha^{(2)}(\zeta-\alpha^{(2)}y^{(2)})\\&-(\zeta-\alpha^{(2)}y^{(2)})y^{(1)}-\alpha^{(2)}+(\zeta-\alpha^{(2)}y^{(2)})\nu^{(1)}+y^{(2)}\alpha^{(2)}\nu^{(2)}\end{aligned}$$

核函数看成常数，可以看到只有一个变量$\alpha^{(2)}$，可以标记目标函数用$W(\alpha^{(2)})$表示。

对$\alpha^{(2)}$求导，得到

$$
\begin{aligned}
\frac{\partial W}{\partial \alpha^{(2)}} = & -K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})\zeta y^{(2)} + K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})\alpha^{(2)} + K(\mathbf{x}^{(2)},\mathbf{x}^{(2)})\alpha^{(2)} + y^{(2)}K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})\zeta \\
& -2K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})\alpha^{(2)} + y^{(1)}y^{(2)} - 1 - \nu^{(1)}y^{(2)} + y^{(2)}\nu^{(2)}
\end{aligned}
$$

令$\frac{\partial W}{\partial \alpha^{(2)}}=0$，整理两边，把$\alpha^{(2)}$整理出来，因为这里的自变量是$\alpha^{(2)}$，整理得

$$
\begin{aligned}
(K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})+K(\mathbf{x}^{(2)},\mathbf{x}^{(2)})-2K(\mathbf{x}^{(1)},\mathbf{x}^{(2)}))\alpha^{(2)} &= K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})\zeta y^{(2)} - y^{(2)}K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})\zeta - y^{(1)}y^{(2)} + 1 + \nu^{(1)}y^{(2)} - y^{(2)}\nu^{(2)} \\
&= y^{(2)}(K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})\zeta - K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})\zeta - y^{(1)} + y^{(2)} + \nu^{(1)} - \nu^{(2)})
\end{aligned}
$$

将$\nu^{(j)}=g(\mathbf{x}^{(j)})-\sum_{i=1}^{2}\alpha^{(i)}y^{(i)}K(\mathbf{x}^{(i)},\mathbf{x}^{(j)})-b$带入整理得到

$$
\begin{aligned}
(K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})+K(\mathbf{x}^{(2)},\mathbf{x}^{(2)})-2K(\mathbf{x}^{(1)},\mathbf{x}^{(2)}))\alpha^{(2)} &= K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})\zeta y^{(2)} - y^{(2)}K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})\zeta - y^{(1)}y^{(2)} + 1 + \nu^{(1)}y^{(2)} - y^{(2)}\nu^{(2)} \\
&= y^{(2)}[K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})\zeta - K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})\zeta - y^{(1)} + y^{(2)} + \\
&\quad (g(\mathbf{x}^{(1)}) - \sum_{i=1}^{2}\alpha^{(i)}y^{(i)}K(\mathbf{x}^{(i)},\mathbf{x}^{(1)}) - b) - (g(\mathbf{x}^{(2)}) - \sum_{i=1}^{2}\alpha^{(i)}y^{(i)}K(\mathbf{x}^{(i)},\mathbf{x}^{(2)}) - b)]
\end{aligned}
$$

利用$\alpha^{(1)}y^{(1)}+\alpha^{(2)}y^{(2)}=\zeta$替换等式右边的$\zeta$。注意，这里因为我们要求$\alpha^{(2)}$，是解析求法，得到的是$(\alpha^{(2)})^{\text{new,unc}}$，所以等式右边替换了$\zeta$是用old的$\alpha$，即$(\alpha^{(1)})^{\text{old}}y^{(1)}+(\alpha^{(2)})^{\text{old}}y^{(2)}=\zeta$，无论是old还是new的$\alpha^{(1)}$或$\alpha^{(2)}$都会满足这个约束条件。带入后整理得到

$$
\begin{aligned}
& (K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})+K(\mathbf{x}^{(2)},\mathbf{x}^{(2)})-2K(\mathbf{x}^{(1)},\mathbf{x}^{(2)}))(\alpha^{(2)})^{\text{new,unc}} \\
=& y^{(2)}[K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})((\alpha^{(1)})^{\text{old}}y^{(1)}+(\alpha^{(2)})^{\text{old}}y^{(2)}) - K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})((\alpha^{(1)})^{\text{old}}y^{(1)}+(\alpha^{(2)})^{\text{old}}y^{(2)}) \\
& - y^{(1)} + y^{(2)} + (g(\mathbf{x}^{(1)}) - \sum_{i=1}^{2}\alpha^{(i)}y^{(i)}K(\mathbf{x}^{(i)},\mathbf{x}^{(1)}) - b) - (g(\mathbf{x}^{(2)}) - \sum_{i=1}^{2}\alpha^{(i)}y^{(i)}K(\mathbf{x}^{(i)},\mathbf{x}^{(2)}) - b)] \\
=& y^{(2)}[K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})(\alpha^{(1)})^{\text{old}}y^{(1)} + K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})(\alpha^{(2)})^{\text{old}}y^{(2)} \\
& - K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})(\alpha^{(1)})^{\text{old}}y^{(1)} - K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})(\alpha^{(2)})^{\text{old}}y^{(2)} \\
& + g(\mathbf{x}^{(1)}) - y^{(1)} - (g(\mathbf{x}^{(2)}) - y^{(2)}) \\
& - (K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})(\alpha^{(1)})^{\text{old}}y^{(1)} + K(\mathbf{x}^{(2)},\mathbf{x}^{(1)})(\alpha^{(2)})^{\text{old}}y^{(2)}) \\
& + (K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})(\alpha^{(1)})^{\text{old}}y^{(1)} + K(\mathbf{x}^{(2)},\mathbf{x}^{(2)})(\alpha^{(2)})^{\text{old}}y^{(2)})] \\
=& y^{(2)}[(K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})+K(\mathbf{x}^{(2)},\mathbf{x}^{(2)})-2K(\mathbf{x}^{(1)},\mathbf{x}^{(2)}))(\alpha^{(2)})^{\text{old}}y^{(2)} + E^{(1)} - E^{(2)}] \\
=& (K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})+K(\mathbf{x}^{(2)},\mathbf{x}^{(2)})-2K(\mathbf{x}^{(1)},\mathbf{x}^{(2)}))(\alpha^{(2)})^{\text{old}} + y^{(2)}(E^{(1)} - E^{(2)})
\end{aligned}
$$

定义$\eta=K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})+K(\mathbf{x}^{(2)},\mathbf{x}^{(2)})-2K(\mathbf{x}^{(1)},\mathbf{x}^{(2)})$，则上面的方程可以变为

$$
\eta(\alpha^{(2)})^{\text{new,unc}} = \eta(\alpha^{(2)})^{\text{old}} + y^{(2)}(E^{(1)} - E^{(2)})
$$

所以

$$
(\alpha^{(2)})^{\text{new,unc}} = (\alpha^{(2)})^{\text{old}} + \frac{y^{(2)}(E^{(1)} - E^{(2)})}{\eta}
$$

这部分推导看起来很是繁琐，但是无非就是一些等式的带入以及整理，式子中的每一项都是标量的运算，因此仔细推导后自己也可以得出结论。

计算完了$(\alpha^{(2)})^{\text{new,unc}}$，则要把它带入$L$和$H$的范围内进行剪辑。则

$$
(\alpha^{(2)})^{\text{new}} = \begin{cases} H, & (\alpha^{(2)})^{\text{new,unc}} > H \\ (\alpha^{(2)})^{\text{new,unc}}, & L \le (\alpha^{(2)})^{\text{new,unc}} \le H \\ L, & (\alpha^{(2)})^{\text{new,unc}} < L \end{cases}
$$

根据

$$
(\alpha^{(1)})^{\text{old}}y^{(1)} + (\alpha^{(2)})^{\text{old}}y^{(2)} = (\alpha^{(1)})^{\text{new}}y^{(1)} + (\alpha^{(2)})^{\text{new}}y^{(2)}
$$

两边同时乘以$y^{(1)}$，再整理后得到

$$(\alpha^{(1)})^{\text{new}}=y^{(1)}y^{(2)}((\alpha^{(2)})^{\text{old}}-(\alpha^{(2)})^{\text{new}})+(\alpha^{(1)})^{\text{old}}$$

以上就是求解$(\alpha^{(1)})^{\text{new}}$和$(\alpha^{(2)})^{\text{new}}$的解析求解方法。

4.3 两个变量的选择方法

第一个变量的选择 选择第一个变量$\alpha^{(1)}$的过程称为外层循环。首先检查在间隔边界上的支撑向量点，即$0<\alpha^{(i)}<C$的样本点（不能取等于C的原因是这种情况数据集可能在间隔边界上，也有可能在间隔与分离超平面之间，甚至在错误的一侧）。检测它们是否满足KKT条件。如果都满足，则在整个数据集上着，看是否满足KKT条件。选择违反KKT条件最严重的样本点。

（《统计学习方法》中说选择$0<\alpha^{(i)}<C$的样本点是因为根据KKT条件可知$y^{(i)}g(\mathbf{x}^i)=1$，其中$g(\mathbf{x}^{(i)})=\sum_{j=1}^N\alpha^{(j)}y^{(j)}K(\mathbf{x}^{(i)},\mathbf{x}^{(j)})+b$。但是这里有一个疑问是KKT条件是一组解是问题最优解的充要条件，然而这里的$\alpha^{(i)}$并不是这个数据集的最优解，即目前初始化的这组α对应的超平面只是一个普通的超平面，并没有做到间隔最大，没有必要满足KKT条件，因此这里选择$0<\alpha^{(i)}<C$个人有点疑问。）

第二个变量的选择 选择第二个变量$\alpha^{(1)}$的过程称为内层循环。标准是希望能使$\alpha^{(2)}$的变化足够大。回顾计算$(\alpha^{(2)})^{\text{new,unc}}$的公式。

$$(\alpha^{(2)})^{\text{new,unc}}=(\alpha^{(2)})^{\text{old}}+\frac{y^{(2)}(E^{(1)}-E^{(2)})}{\eta}$$

从$(\alpha^{(2)})^{\text{old}}$和$(\alpha^{(2)})^{\text{new,unc}}$之间的变化是$\frac{y^{(2)}(E^{(1)}-E^{(2)})}{\eta}$。所以要使得$\alpha^{(2)}$的变化足够大，可以让$E^{(1)}-E^{(2)}$足够大。$E^{(1)}$由外层循环选择的$\alpha^{(1)}$确定。那$\alpha^{(2)}$选择了，就随即确定了$E^{(2)}$。因此在选择$\alpha^{(2)}$时，要根据$E^{(1)}$的符号作为依据。如果$E^{(1)}$为正，则选择最负或最小的$E^{(i)}$作为$E^{(2)}$。反之如果$E^{(1)}$为负，则选择最正或最大的$E^{(i)}$作为$E^{(2)}$。

这是一种理想的选取方法。具体的效果还是要看最小化目标函数这个任务是否正在执行，因为有特殊情况，即使采用上述的选取方法，目标函数得不到足够的下降。此时，选取$(\alpha^{(2)})$的范围扩大。先在支撑向量点的范围内依次寻找，找到一个可以使得目标函数有足够下降的那个变量作为$(\alpha^{(2)})$。如果找不到，则再扩大范围，在整个数据集范围所对应的$(\alpha^{(i)})$。如果还不能使目标函数有足够的下降，则表明此时外层循环选择的$(\alpha^{(1)})$不好，退回去重新选择$(\alpha^{(1)})$。

更新阈值**b**和误差值$E^{(i)}$ 每次更新完两个变量后，要检查$(\alpha^1)^{(\text{new})}$和$(\alpha^2)^{(\text{new})}$是否还满足$0<\alpha^{(i)}<C$。根据KKT条件可知，$y^{(i)}g(\mathbf{x}^i)=1$，两边同时乘以$y^{(i)}$得到$g(\mathbf{x}^i)=y^{(i)}$，所以$g(\mathbf{x}^1)=\sum_{j=1}^N\alpha^{(j)}y^{(j)}K(\mathbf{x}^{(1)},\mathbf{x}^{(j)})+b=y^{(1)}$ 但是其中$\sum_{j=1}^N$中的第1和第2项要用新的α来代替，因此会破坏原来的KKT等式约束条件，因此必须对b做出调整，b^{new}整理得到

$$(b^{(1)})^{\text{new}}=y^{(1)}-\sum_{j=3}^N\alpha^{(j)}y^{(j)}K(\mathbf{x}^{(j)},\mathbf{x}^{(1)})-(\alpha^{(1)})^{\text{new}}y^{(1)}K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})-(\alpha^{(2)})^{\text{new}}y^{(2)}K(\mathbf{x}^{(2)},\mathbf{x}^{(1)})$$

又因为$E^{(i)}=g(\mathbf{x}^{(i)})-y^{(i)}$，老的误差项

$$(E^{(1)})^{\text{old}}=\sum_{i=3}^N\alpha^{(i)}y^{(i)}K(\mathbf{x}^{(i)},\mathbf{x}^{(1)})+(\alpha^{(1)})^{\text{old}}y^{(1)}K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})+(\alpha^{(2)})^{\text{old}}y^{(2)}K(\mathbf{x}^{(2)},\mathbf{x}^{(1)})+b^{\text{old}}-y^{(1)}$$

移项得

$$y^{(1)}-\sum_{i=3}^N\alpha^{(i)}y^{(i)}K(\mathbf{x}^{(i)},\mathbf{x}^{(1)})=-(E^{(1)})^{\text{old}}+(\alpha^{(1)})^{\text{old}}y^{(1)}K(\mathbf{x}^{(1)},\mathbf{x}^{(1)})+(\alpha^{(2)})^{\text{old}}y^{(2)}K(\mathbf{x}^{(2)},\mathbf{x}^{(1)})+b^{\text{old}}$$

把这两项抽出来是因为要把它们带入上面$(b^{(1)})^{\text{new}}$的表达式中去替换前两项，这样的好处是可以引入b^{new}和b^{old}的关系。整理得到

$$\begin{aligned}(b^{(1)})^{\text{new}} &= -(E^{(1)})^{\text{old}} + (\alpha^{(1)})^{\text{old}} y^{(1)} K(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) + (\alpha^{(2)})^{\text{old}} y^{(2)} K(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) + b^{\text{old}} \\ &- (\alpha^{(1)})^{\text{new}} y^{(1)} K(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) - (\alpha^{(2)})^{\text{new}} y^{(2)} K(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) \\ &= -(E^{(1)})^{\text{old}} - y^{(1)} K(\mathbf{x}^{(1)}, \mathbf{x}^{(1)})[(\alpha^{(1)})^{\text{new}} - (\alpha^{(1)})^{\text{old}}] - y^{(2)} K(\mathbf{x}^{(2)}, \mathbf{x}^{(1)})[(\alpha^{(2)})^{\text{new}} - (\alpha^{(2)})^{\text{old}}] + b^{\text{old}}\end{aligned}$$

同理，若$0 < (\alpha^{(2)})^{\text{new}} < C$，则

$$(b^{(2)})^{\text{new}} = -(E^{(2)})^{\text{old}} - y^{(1)} K(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})[(\alpha^{(1)})^{\text{new}} - (\alpha^{(1)})^{\text{old}}] - y^{(2)} K(\mathbf{x}^{(2)}, \mathbf{x}^{(2)})[(\alpha^{(2)})^{\text{new}} - (\alpha^{(2)})^{\text{old}}] + b^{\text{old}}$$

如果$(\alpha^1)^{(\text{new})}$和$(\alpha^2)^{(\text{new})}$同时满足$0 < \alpha^{(i)} < C$，则$(b^{(1)})^{\text{new}} = (b^{(2)})^{\text{new}} = b^{\text{new}}$。否则$b^{\text{new}}$取$(b^{(1)})^{\text{new}}$和$(b^{(2)})^{\text{new}}$的均值。

更新$(E^{(i)})^{\text{new}}$，根据误差项的定义，