Support Vector Machine 推导

Soul Walker

2020 年 7 月 22 日

1 Support Vector Machine

SVM 有三种情况及其处理手段:

1. 线性可分问题: Hard-Margin SVM

2. 线性可分但是有一点错误: Soft-Margin SVM

3. 非线性问题: Kernel SVM

NOTE: SMO 算法是支持向量机学习的一种快速算法

1.1 模型定义

首先我们定义数据如下

输入：训练数据集

$$T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N))\} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$$

其中，$\mathbf{x}_i \in \chi \in \mathbb{R}^N$ 为实例的特征向量，$\mathbf{x}_i = (\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)}, ..., \mathbf{x}_i^{(n)})^T$ ，$y_i \in \{+1, -1\}$ 为实例的类别，$i = 1, 2, \cdots, N$ ，实例特征向量 $\mathbf{x}$，$\mathbf{X}$ 为训练样本集，形状为 (N, p)，$\mathbf{Y}$ 是训练标签集，形状为 $(N, 1)$。这里我们使用 α 来作为拉格朗日乘子。

输出：实例 $\mathbf{x}$ 所属的类别 y。

SVM（最大间隔分类器）的分类决策函数为 $f(\mathbf{x}) = sign(w^* \cdot \mathbf{x} + b^*)$，其中 w^*, b^* 是 w 和 b 的最优解。

由名称(最大间隔分类器)拆分解: 1. 最大: max 2. 间隔 (数据和直线的距离的最小值): margin(w, b) 3. 分类器: $y_i(w^T\mathbf{x}_i + b) > 0 \quad \forall i = 1, ..., N$

对于分类器的推导:

$$\begin{cases} w^T\mathbf{x}_i + b > 0, y_i = +1 \\ w^T\mathbf{x}_i + b < 0, y_i = -1 \end{cases}$$

那么分类正确的条件是 $y_i(w^T\mathbf{x}_i + b) > 0$。

对于间隔的推导:

定义点 $(\mathbf{x}_i, y_i)$ 到直线 $f(\mathbf{x}) = w^T\mathbf{x} + b$ 的距离 $distance(w, b, \mathbf{x}_i) = \frac{1}{||w||}|w^T\mathbf{x}_i + b|$

$$\begin{aligned}
&margin(w, b) \\
&= \min_{w,b,\mathbf{x}_i} distance(w, b, \mathbf{x}_i) \\
&= \min_{w,b,\mathbf{x}_i} \frac{1}{||w||}|w^T\mathbf{x}_i + b|
\end{aligned}$$

综合以上推导：

$$\begin{aligned}
&\max_{w,b} \min_{\mathbf{x}_i} \frac{1}{||w||}|w^T\mathbf{x}_i + b| \\
\Rightarrow &\max_{w,b} \min_{\mathbf{x}_i} \frac{1}{||w||} y_i(w^T\mathbf{x}_i + b) \\
\Rightarrow &\max_{w,b} \frac{1}{||w||} \min_{\mathbf{x}_i} y_i(w^T\mathbf{x}_i + b)
\end{aligned}$$

令

$$\gamma = \min_{\mathbf{x}_i} y_i(w^T\mathbf{x}_i + b) = 1$$

那么

$$y_i(w^T\mathbf{x}_i + b) > 0 \Rightarrow \gamma > 0 \quad s.t. \quad \min_{\mathbf{x}_i y_i}(w^T\mathbf{x}_i + b) = \gamma$$

$$\Rightarrow \begin{cases} \max\limits_{w,b} \dfrac{1}{||w||} \\ s.t. \quad y_i(w^T\mathbf{x}_i + b) = 1 \end{cases}$$

可以转换为最小化问题（优化问题且是 Convex Optimization（凸优化）问题）：

$$\begin{cases} \min\limits_{w,b} ||w|| = \min\limits_{w,b} \dfrac{1}{2} w^T w \\ s.t. \quad y_i(w^T x_i + b) \geq 1 \quad \forall i = 1, 2, ..., N \end{cases}$$

1.2 Hard-Margin SVM 求解

带约束问题:

$$\begin{cases} \min\limits_{w,b} \dfrac{1}{2} w^T w \\ s.t. \quad y_i(w^T\mathbf{x}_i + b) \geq 1 \Leftrightarrow 1 - y(w^T\mathbf{x}_i + b) \leq 0 \quad \forall i = 1, 2, ..., N \end{cases} \tag{1}$$

可以得到拉格朗日函数 (Lagrange function)

$$\mathfrak{L}(w, b, \alpha) = \frac{1}{2} w^T w + \sum_{i=1}^{N} \alpha_i (1 - y_i(w^T x_i + b)) \tag{2}$$

于是得到

$$\begin{cases} \min\limits_{w,b} \max\limits_{\alpha} \mathfrak{L}(w, b, \alpha) \\ s.t. \quad \alpha_i \geq 0 \end{cases} \tag{3}$$

此处我们展开

$$\begin{cases} when \quad 1-y_i(w^T\mathbf{x}_i+b)>0, \quad \max\limits_{\alpha}\mathfrak{L}(w,b,\alpha)=\frac{1}{2}w^Tw+\infty=+\infty \\ when \quad 1-y_i(w^T\mathbf{x}_i+b)\leq, \quad \max\limits_{\alpha}\mathfrak{L}(w,b,\alpha)=\frac{1}{2}w^Tw+0=\frac{1}{2}w^Tw \end{cases} \tag{4}$$

对于第一个公式取到 ∞ 所以无意义，舍弃；第二个公式可以得到最优解出处

$$\min_{w,b}\max_{\alpha}\mathfrak{L}(w,b,\alpha)=\min_{w,b}\frac{1}{2}w^Tw \tag{5}$$

因此综合上述三个公式可以推导出

$$\min_{w,b}\max_{\alpha}\mathfrak{L}(w,b,\alpha)=\min_{w,b}(\infty,\frac{1}{2}w^Tw)=\min_{w,b}\frac{1}{2}w^Tw \tag{6}$$

对于式 (3) 的对偶问题

$$\begin{cases} \max\limits_{\alpha}\min\limits_{w,b}\mathfrak{L}(w,b,\alpha) \\ s.t. \quad \alpha_i\geq 0 \end{cases} \tag{7}$$

对于式 (7) 的求解如下

Step1

$$\begin{aligned} let \quad 0=\frac{\partial\mathfrak{L}}{\partial b}&=\frac{\partial}{\partial b}[\sum_{i=1}^{N}\alpha_i-\sum_{i=1}^{N}\alpha_iy_i(w^T\mathbf{x}_i+b)] \\ &=\frac{\partial}{\partial b}[-\sum_{i=1}^{N}\alpha_iy_ib] \\ &=-\sum_{i=1}^{N}\alpha_iy_i \end{aligned} \tag{8}$$

所以可得 $\sum_{i=1}^{N}\alpha_iy_i=0$，将其代入 $\mathfrak{L}(w,b,\alpha)$ 式 (2) 可得

$$\begin{aligned} \mathfrak{L}(w,b,\alpha)&=\frac{1}{2}w^Tw+\sum_{i=1}^{N}\alpha_i-\sum_{i=1}^{N}\alpha_iy_i(w^T\mathbf{x}_i+b) \\ &=\frac{1}{2}w^Tw+\sum_{i=1}^{N}\alpha_i-\sum_{i=1}^{N}\alpha_iy_iw^T\mathbf{x}_i+\sum_{i=1}^{N}\alpha_iy_ib \\ &=\frac{1}{2}w^Tw+\sum_{i=1}^{N}\alpha_i-\sum_{i=1}^{N}\alpha_iy_iw^T\mathbf{x}_i \end{aligned} \tag{9}$$

Step2

$$\begin{aligned} &let \quad 0=\frac{\partial\mathfrak{L}}{\partial w}=\frac{1}{2}2w-\sum_{i=1}^{N}\alpha_iy_i\mathbf{x}_i \\ &w^*=\sum_{i=1}^{N}\alpha_iy_i\mathbf{x}_i \end{aligned} \tag{10}$$

将其带入 $\mathfrak{L}(w,b,\alpha)$ 式 (9) 可得

$$\begin{aligned} \mathfrak{L}(w,b,\alpha)&=\frac{1}{2}(\sum_{i=1}^{N}\alpha_iy_i\mathbf{x}_i)^T(\sum_{j=1}^{N}\alpha_jy_j\mathbf{x}_j)-\sum_{i=1}^{N}\alpha_iy_i(\sum_{j=1}^{N}\alpha_jy_j\mathbf{x}_j)^T\mathbf{x}_i+\sum_{i=1}^{N}\alpha_i \\ &=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jy_iy_j\mathbf{x}_i^T\mathbf{x}_j+\sum_{i=1}^{N}\alpha_i \end{aligned} \tag{11}$$

最后，结合式 (11) 可得

$$\begin{cases} \max\limits_{\alpha} \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j + \sum_{i=1}^{N}\alpha_i \\ s.t. \quad \alpha_i \geq 0 \end{cases} \tag{12}$$

求解的步骤为式 (1)-> 式 (3)-> 式 (7)-> 式 (12)

而我们通过以上求解可以得出一点结论：非支持向量点满足 $1 - y_i(w^T\mathbf{x}_i + b) > 0$，则为了满足 KKT 条件的话，$\alpha_i = 0$ 必须成立，也就是说非支持向量点在对 w^*, b^* 的求解问题上无贡献，求解时，他们没有任何用处。

KKT 条件

$$\begin{cases} \frac{\partial \mathfrak{L}}{\partial w} = 0 \quad \frac{\partial \mathfrak{L}}{\partial b} = 0 \\ \alpha_i(1 - y_i(w^T\mathbf{x}_i + b)) = 0 \\ \alpha_i \geq 0 \\ 1 - y_i(w^T\mathbf{x}_i + b) \leq 0 \end{cases} \tag{13}$$

⇒w*, b*

$$w^* = \sum_{i=1}^{N}\alpha_i y_i \mathbf{x}_i \tag{14}$$

$$\begin{aligned} &\exists\left(\mathbf{x}_k, y_k\right), s.t. \quad 1 - y_k\left(w^\top\mathbf{x}_k + b\right) = 0 \\ \Rightarrow & y_k\left(w^\top\mathbf{x}_k + b\right) = 1 \\ & y_k^2\left(w^\top\mathbf{x}_k + b\right) = y_k \\ b^* &= y_k - w^\top\mathbf{x}_k \\ &= y_k - \frac{v}{\mathbf{x}_i}\lambda_i y_i \mathbf{x}_i^\top\mathbf{x}_k \end{aligned} \tag{15}$$

那么联立以上公式可得 w^*, b^* 的求解

$$\begin{cases} w^* = \sum_{i=1}^{N}\alpha_i y_i \mathbf{x}_i \\ b^* = y_k - \sum_{i=1}^{N}\alpha_i y_i \mathbf{x}_i^T\mathbf{x}_k \end{cases} \tag{16}$$

那么通过分类器 $f(\mathbf{x}) = sign(w^*\mathbf{x} + b^*)$ 可以进行分类，此时划分超平面就是 $w^{*T}\mathbf{x} + b^* = 0$

1.3 Soft-Margin SVM 求解

此处 Soft-Margin SVM 指的是处理**线性不可分或可分但存在噪声**的数据。Soft：允许一点点错误所以此处由硬间隔约束问题 (1) 可得此处可以转换为 $\min_{w,b}\frac{1}{2}w^Tw + loss$，此处的讨论为 $loss$。我们有以下：

1.loss 使用犯错个数：

$$loss = \sum_{i=1}^{N} I\left\{y_i\left(w^\top\mathbf{x}_i + b\right) < 1\right\} \tag{17}$$

但此处指示函数 (I) 关于 w 不连续令 $z = y(w^T\mathbf{x}_i + b)$ 那么此处取 **0-1 损失函数**

$$loss_{0-1} = \begin{cases} 1, z \leq 1 \\ 0, otherwise \end{cases} \tag{18}$$

2.loss 使用距离：Hinge-Loss

$$\begin{cases} when \quad y_i(w^T\mathbf{x}_i + b) \geq 1, let \quad loss = 0 \\ when \quad y_i(w^T\mathbf{x}_i + b) < 1, let \quad loss = 1 - y_i(w^T\mathbf{x}_i + b) \end{cases} \tag{19}$$

那么综合式 (19) 可得:

$$loss = max\{0, 1 - y_i(w^T\mathbf{x}_i + b)\} \tag{20}$$

那么将这个错误 (loss) 加入到 Hard-Margin SVM 中，可得：

$$\begin{cases} \min\limits_{w,b} \frac{1}{2} w^T w + C \sum\limits_{i=1}^{N} max\{0, 1 - y_i(w^T\mathbf{x}_i + b)\} \\ s.t. \quad y_i(w^T\mathbf{x}_i + b) \geq 1 - \xi_i, i = 1, 2, ..., N \end{cases} \tag{21}$$

引入 $\xi_i = 1 - y_i(w^T\mathbf{x}_i + b), \xi_i \geq 0$

$$\begin{cases} \min\limits_{w,b,\xi} \frac{1}{2} w^T w + C \sum\limits_{i=1}^{N} \xi_i \\ s.t. \quad y_i(w^T\mathbf{x}_i + b) \geq 1 - \xi_i, i = 1, 2, ..., N \end{cases} \tag{21}$$

2 Kernel Method

这里对核方法从两个角度有两种解释:

1. 非线性带来的高维转换（从模型角度）：由非线性可分的源数据经过非线性转换 $\phi(\mathbf{x})$ 得到线性可分的数据。

2. 对偶表示带来内积（从优化角度）

对于 1. 的解释：由简单的异或问题 (二维空间上线性不可分)：$f(1,1) = 1, \quad f(0,0) = 1, \quad f(1,0) = -1, \quad f(0,1) = -1$ 通过非线性转换即 $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2) \quad \overset{\phi(\mathbf{x})}{\rightarrow} \quad \mathbf{z} = (\mathbf{x}_1, \mathbf{x}_2, (\mathbf{x}_1 - \mathbf{x}_2)^2)$，此时在三维空间中四个点转换为 $f(1,1,0) = 1, \quad f(0,0,0) = 1, \quad f(1,0,1) = -1, \quad f(0,1,1) = -1$ 那么这个时候就是线性可分的，也就是说**高维比低维更易线性可分**。

对于 2. 的解释：在 Hard-Margin SVM 中，我们通过 Primal Problem(原始问题)(公式 (1)) 得到 Dual Problem(对偶问题)(公式 (11)) 其中 $x_i^T x_j$ 就是对 $\phi(x_i)^T\phi(x_j)$ 的解释。

那么将核方法应用在 SVM 中可以得到上面公式的对偶问题

$$\begin{cases} \max\limits_{\alpha} -\frac{1}{2} \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum\limits_{i=1}^{N} \alpha_i \\ s.t. \quad \alpha_i \geq 0 \end{cases} \tag{22}$$

在求解的时候需要求得内积，于是不可分数据在通过特征变换后，需要求得变换后的内积。我们常常很难求得变换函数的内积。于是直接引入内积的变换函数：

$$\begin{aligned}&\forall \mathbf{x}, \mathbf{z} \in \chi, \exists \phi \in \mathcal{H}: \mathbf{x} \rightarrow \mathbf{z} \\ &s.t. \quad k(\mathbf{x}, \mathbf{z})=\phi(\mathbf{x})^{T} \phi(\mathbf{z})\end{aligned} \tag{23}$$

称 $k(\mathbf{x}, \mathbf{z})$ 为一个正定核函数，其中 $\mathcal{H}$ 是 Hilbert 空间（完备的线性内积空间），如果去掉内积这个条件我们简单地称为核函数。

3 SMO 算法

对于如下凸二次规划问题：

$$\begin{cases}\max\limits_{\alpha} -\dfrac{1}{2} \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} \alpha_{i} \alpha_{j} y_{i} y_{j} \mathbf{x}_{i}^{T} \mathbf{x}_{j}+\sum\limits_{i=1}^{N} \alpha_{i} \\ \text{s.t.} \ \sum\limits_{i=1}^{N} \alpha_{i} y_{i}=0 \\ 0 \leqslant \alpha_{i} \leqslant C, \quad i=1,2, \cdots, N\end{cases} \tag{24}$$

我们令

$$\mathfrak{L}=-\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_{i} \alpha_{j} y_{i} y_{j} \mathbf{x}_{i}^{T} \mathbf{x}_{j}+\sum_{i=1}^{N} \alpha_{i} \tag{25}$$

那么

$$\begin{aligned}\mathfrak{L}(\alpha_{1}, \alpha_{2})=&\alpha_{1}+\alpha_{2}+\sum_{i=3}^{N} \alpha_{i}-\frac{1}{2}[\alpha_{1} y_{1} \alpha_{2} y_{2} \mathbf{x}_{1}^{T} \mathbf{x}_{1}+2 \alpha_{1} y_{1} \alpha_{2} y_{2} \mathbf{x}_{1}^{T} \mathbf{x}_{2}+2 \sum_{j=3}^{N} \alpha_{1} y_{1} \alpha_{j} y_{j} \mathbf{x}_{1}^{T} \mathbf{x}_{j} \\ &+\alpha_{2} y_{2} \alpha_{2} y_{2} \mathbf{x}_{2}^{T} \mathbf{x}_{2}+2 \sum_{i=3}^{N} \alpha_{2} y_{2} \alpha_{i} y_{i} \mathbf{x}_{2}^{T} \mathbf{x}_{i}+\sum_{i=3}^{N} \sum_{j=3}^{N} \alpha_{i} y_{i} \alpha_{j} y_{j} \mathbf{x}_{i}^{T} \mathbf{x}_{j}]\end{aligned} \tag{26}$$

此处令 $\mathbf{x}_{i}^{T} \mathbf{x}_{j}=k_{ij}$ 那么

$$\mathfrak{L}^{\prime}(\alpha_{1}, \alpha_{2})=\alpha_{1}+\alpha_{2}-\frac{1}{2}[\alpha_{1}^{2} k_{11}+2 \alpha_{1} y_{1} \alpha_{2} y_{2} k_{12}+\alpha_{2}^{2} k_{22}+2 \sum_{j=3}^{N} \alpha_{1} y_{1} \alpha_{j} y_{j} k_{1j}+2 \sum_{i=3}^{N} \alpha_{2} y_{2} \alpha_{i} y_{i} k_{2i}] \tag{27}$$

根据 $\sum_{i=1}^{N} \alpha_{i} y_{i}=0$ 可以得出 $\alpha_{1} y_{1}+\alpha_{2} y_{2}+\sum_{i=3}^{N} \alpha_{i} y_{i}=0$，这里我们令 $\sum_{i=3}^{N} \alpha_{i} y_{i}=-C$（C 为常数），可以得到 $\alpha_{1} y_{1}+\alpha_{2} y_{2}=C$(或者可以写作 $\alpha_{1}^{\text{old}} y+\alpha_{2}^{\text{old}} y_{2}=C$)，所以 $\alpha_{1}=(C-\alpha_{2} y_{2}) y_{1}$ 然后代入式 (27) 可得

$$\begin{aligned}\mathfrak{L}(\alpha_{2})=&y_{1}(C-\alpha_{2} y_{2})+\alpha_{2}-\frac{1}{2}[(C-\alpha_{2} y_{2})^{2} k_{11}+2(C-\alpha_{2} y_{2}) \alpha_{2} y_{2} k_{12}+\alpha_{2}^{2} k_{22} \\ &+2 \sum_{i=3}^{N}(C-\alpha_{2} y_{2}) \alpha_{i} y_{i} k_{1i}+2 \sum_{i=3}^{N} \alpha_{2} y_{2} \alpha_{i} y_{i} k_{2i}]\end{aligned} \tag{28}$$

对式 (28) 求偏导

$$
\begin{aligned}
\frac{\partial \mathfrak{L}}{\partial \alpha_2} &= -y_1y_2 + 1 - \frac{1}{2}[2(C-\alpha_2 y_2)(-y_2)k_{11} + 2Cy_2k_{12} - 4\alpha_2k_{12} + 2\alpha_2k_{22} \\
&\quad - 2\sum_{i=3}^{N} y_2\alpha_i y_i k_{1i} + 2\sum_{i=3}^{N} y_2\alpha_i y_i k_{2i}] \\
&= 1 - y_1y_2 - Cy_2k_{11} - \alpha_2k_{11} - cy_2k_{12} + 2\alpha_2k_{12} - \alpha_2k_{22} + \sum_{i=3}^{N} y_2\alpha_i y_i k_{1i} \\
&\quad - \sum_{i=3}^{N} y_2\alpha_i y_i k_{2i}
\end{aligned} \tag{29}
$$

$$
\frac{\partial \mathfrak{L}}{\partial w} \Rightarrow w = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i
$$

那么对于 $f(x_1)$ 可得以下推导

$$
\begin{aligned}
f(\mathbf{x}_1) &= w^T\mathbf{x}_1 + b \\
&= \alpha_1 y_1 \mathbf{x}_1^T\mathbf{x}_1 + \alpha_2 y_2 \mathbf{x}_2^T\mathbf{x}_1 + \sum_{i=3}^{N} \alpha_i y_i \mathbf{x}_i^T\mathbf{x}_1 + b
\end{aligned} \tag{30}
$$

那么扩展到 $i = 1, 2$

$$
\begin{aligned}
f(\mathbf{x}_1) - \alpha_1y_1k_{11} - \alpha_2y_2k_{12} - b &= \sum_{i=3}^{N} \alpha_i y_i k_{1i} \\
f(\mathbf{x}_2) - \alpha_1y_1k_{12} - \alpha_2y_2k_{22} - b &= \sum_{i=3}^{N} \alpha_i y_i k_{2i}
\end{aligned} \tag{31}
$$

因此式 (29) 中令对 α_2 的偏导为 0，则

$$
\begin{aligned}
\alpha_2^{new}(k_{11} + k_{22} - 2k_{12}) &= 1 - y_1y_2 + Cy_2k_{11} - Cy_2k_{12} + \sum_{i=3}^{N} y_2\alpha_i y_i k_{1i} - \sum_{i=3}^{N} y_2\alpha_i y_i k_{2i} \\
&= y_2(y_2 - y_1 + Ck_{11} - CK_{12} + \sum_{i=3}^{N} \alpha_i y_i k_{1i} - \sum_{i=3}^{N} \alpha_i y_i k_{2i}) \\
&= y_2(y_2 - y_1 + \alpha_1^{old}y_1k_{11} + \alpha_2^{old}y_2k_{11} - \alpha_1^{old}y_1k_{12} - \alpha_2^{old}y_1k_{12} \\
&\quad + f(x_1) - \alpha_1^{old}y_1k_{11} - \alpha_2^{old}y_2k_{12} - b - f(x_2) + \alpha_1^{old}y_1k_{12} + \alpha_2^{old}y_2k_{22} + b) \\
&= y_2[f(x_1) - y_1 - (f(x_2) - y_2) + \alpha_2^{old}y_2(k_{11} + k_{22} - 2k_{12})]
\end{aligned} \tag{32}
$$

定义 $E_1 = f(\mathbf{x}_1) - y_1$, $E_2 = f(\mathbf{x}_2) - y_2$, $\eta = (k_{11} + k_{22} - 2k_{12})$ 所以综上变换式 (32) 为

$$
\alpha_2\eta = y_2\left[(E_1 - E_2) + \alpha_2^{old}y_2\eta\right] \tag{33}
$$

所以可得

$$
\alpha_2^{new} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta} \tag{34}
$$

根据式 $\alpha_1 y_1 + \alpha_2 y_2 = C$ 可得

$$\alpha_1^{new} = y_1(C - \alpha_2^{new} y_2) \tag{35}$$

那么根据式 (29)(30) 以及支持向量满足的表达式可得

$$\begin{aligned} w^T x_1 + b_1^{new} &= y_1 \\ b_1^{new} &= y_1 - w^T \mathbf{x}_1 \\ b_1^{new} &= y_1 - \alpha_1^{new} y_1 \mathbf{x}_1^T \mathbf{x}_1 - \alpha_2^{new} y_2 \mathbf{x}_2^T \mathbf{x}_1 - \sum_{i=3}^{N} \alpha_i y_i \mathbf{x}_i^T \mathbf{x}_1 \end{aligned} \tag{36}$$

根据 $E_i = f(x_i) - y_i$ 以及式 (30) 可得

$$\begin{aligned} E_1 &= w^T \mathbf{x}_1 + b - y_1 \\ &= \alpha_1^{old} y_1 \mathbf{x}_1^T \mathbf{x}_1 + \alpha_2^{old} y_2 \mathbf{x}_2^T \mathbf{x}_1 + \sum_{i=3}^{N} \alpha_i y_i \mathbf{x}_i^T \mathbf{x}_i + b^{old} - y_1 \end{aligned} \tag{37}$$

式 (37) 做如下变换

$$y_1 - \sum_{i=3}^{N} \alpha_i y_i \mathbf{x}_i^T \mathbf{x}_1 = \alpha_1^{old} y_1 \mathbf{x}_1^T \mathbf{x}_1 + \alpha_2^{old} y_2 \mathbf{x}_2^T \mathbf{x}_1 + b^{old} - E_1 \tag{38}$$

带入式 (36) 得

$$\begin{aligned} b_1^{new} &= \alpha_1^{old} y_1 \mathbf{x}_1^T \mathbf{x}_1 + \alpha_2^{old} y_2 \mathbf{x}_2^T \mathbf{x}_1 + b^{old} - E_1 - \alpha_1^{new} y_1 \mathbf{x}_1^T \mathbf{x}_1 - \alpha_2^{new} y_2 \mathbf{x}_2^T \mathbf{x}_1 \\ &= -E_1 + y_1 \mathbf{x}_1^T \mathbf{x}_1 (\alpha_1^{old} - \alpha_1^{new}) + y_2 \mathbf{x}_2^T \mathbf{x}_1 (\alpha_2^{old} - \alpha_2^{new}) + b^{old} \end{aligned} \tag{39}$$

同理可得

$$b_2^{new} = -E_2 + y_1 \mathbf{x}_1^T \mathbf{x}_2 (\alpha_1^{old} - \alpha_1^{new}) + y_2 \mathbf{x}_2^T \mathbf{x}_2 (\alpha_2^{old} - \alpha_2^{new}) + b^{old} \tag{40}$$

那么对于 E_i 的求解如下

$$E_i^{new} = \sum_{j=1}^{N} y_j \alpha_j \mathbf{x}_i^T \mathbf{x}_j + b^{new} - y_i \tag{41}$$

将 $x_i^T x_j = k_{ij}$ 带入上式 (39), (40), (41) 并重新做出变换可得

$$b_1^{new} = -E_1 - y_1 k_{11} (\alpha_1^{new} - \alpha_1^{old}) - y_2 k_{21} (\alpha_2^{new} - \alpha_2^{old}) + b^{old} \tag{42}$$

$$b_2^{new} = -E_2 - y_1 k_{12} (\alpha_1^{new} - \alpha_1^{old}) - y_2 k_{22} (\alpha_2^{new} - \alpha_2^{old}) + b^{old} \tag{43}$$

$$E_i^{new} = \sum_{j=1}^{N} y_j \alpha_j k_{ij} + b^{new} - y_i \tag{44}$$