1 线性可分支持向量机 (硬间隔最大化)

对于约束最优化问题:

$$\min_{w,b} \frac{1}{2}||w||^2 \tag{1}$$

$$s.t \quad y_i(w \cdot x_i + b) - 1 \geq 0, \quad i = 1, 2, \ldots, N \tag{2}$$

易知式 (2) 满足等号"="的样本点实例为支持向量。

运用拉格朗日乘数法，定义拉格朗日函数:

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{N} \alpha_i \cdot (y_i(w \cdot x_i + b) - 1) \tag{3}$$

$$= \frac{1}{2}||w||^2 - \sum_{i=1}^{N} \alpha_i y_i(w \cdot x_i + b) + \sum_{i=1}^{N} \alpha_i \tag{4}$$

即该问题是来探究 $\max\limits_{\alpha} \min\limits_{w,b} L(w, b, \alpha)$，因此需要先求 $L(w, b, \alpha)$ 对 w, b 的极小，再求对 α 的极大.

(1) 先求 $L(w, b, \alpha)$ 对 w, b 的极小根据式 (4) 分别对 w,b 求偏导数并令其等于 0

对 w,

$$\bigtriangledown_w L(w, b, \alpha) = w - \sum_{i=1}^{N} \alpha_i y_i x_i = 0 \tag{5}$$

对 b,

$$\bigtriangledown_b L(w, b, \alpha) = -\sum_{i=1}^{N} \alpha_i y_i = 0 \tag{6}$$

因此有

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{7}$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \tag{8}$$

把式 (7) 代入 (4) 可得到如下公式:

$$L(w, b, \alpha) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N} \alpha_i y_i ((\sum_{j=1}^{N} \alpha_j y_j x_j) \cdot x_i + b) + \sum_{i=1}^{N} \alpha_i \tag{9}$$

又因为式 (8)，因此式 (9) 的第二式中 $\sum_{i=1}^{N}\alpha_i y_i b=0$，因此有：

$$
\begin{aligned}
L(w,b,\alpha) &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)-\sum_{i=1}^{N}\alpha_i y_i((\sum_{j=1}^{N}\alpha_j y_j x_j)\cdot x_i)+\sum_{i=1}^{N}\alpha_i \quad (10)\\
&= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)-\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i \quad (11)\\
&= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i \quad (12)
\end{aligned}
$$

即

$$
\min_{w,b} L(w,b,\alpha) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i \tag{13}
$$

(2) 再求 $\min\limits_{w,b} L(w,b,\alpha)$ 对 α 的极大，即：

$$
\begin{aligned}
\max_{\alpha} \quad & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i && (14)\\
s.t \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 && (15)\\
& \alpha_i \geq 0, \quad i=1,2,\ldots,N && (16)
\end{aligned}
$$

将式 (14) 的目标函数由求极大值转换为求极小值，就得到下面与之等价的对偶问题：

$$
\begin{aligned}
\min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)-\sum_{i=1}^{N}\alpha_i && (17)\\
s.t \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 && (18)\\
& \alpha_i \geq 0, \quad i=1,2,\ldots,N && (19)
\end{aligned}
$$

我们需要求解(利用SMO算法)) 得到最优解$\alpha^*=(\alpha_1^*,\alpha_2^*,\ldots,\alpha_N^*)^T$ 我们选择一个 $\alpha_j^*>0$，这样我们能够求得原始目标优化问题 (1) 的解为:

$$w^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i \tag{20}$$

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j) \tag{21}$$

于是可以求得分离超平面及分类决策函数

$$w^* \cdot x + b^* = 0 \tag{22}$$

$$f(x) = sign(w^* \cdot x + b^*) \tag{23}$$

2 线性支持向量机 (软间隔最大化)

其中凸二次优化的原始问题为:

$$\min_{w,b,\xi} \quad \frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \xi_i \tag{24}$$

$$s.t. \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, N \tag{25}$$

$$\xi_i \geq 0, \quad i = 1, 2, \ldots, N \tag{26}$$

我们需要求解该原始问题 (24) 的对偶问题, 先写出拉格朗日函数:

$$L(w, b, \xi, \alpha, \mu) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i (y_i(w \cdot x_i) - 1 + \xi_i) - \sum_{i=1}^{N} \mu_i \xi_i \tag{27}$$

其中，$\alpha_i \geq 0, \mu_i \geq 0.$

与前面同理，先求 $L(w, b, \xi, \alpha, \mu)$ 对 w, b, ξ 的极小:

$$\bigtriangledown_w L(w, b, \xi, \alpha, \mu) = w - \sum_{i=1}^{N} \alpha_i y_i x_i = 0 \tag{28}$$

$$\bigtriangledown_b L(w, b, \xi, \alpha, \mu) = -\sum_{i=1}^{N} \alpha_i y_i = 0 \tag{29}$$

$$\bigtriangledown_{\xi_i} L(w, b, \xi, \alpha, \mu) = C - \alpha_i - \mu_i = 0 \tag{30}$$

这样，可以得到：

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{31}$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \tag{32}$$

$$C - \alpha_i - \mu_i = 0 \tag{33}$$

将式（31），（32），（33）代入式（27），可得到

$$\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N} \tag{34}$$

再全对 $\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu)$ 求 α 的极大，即得到对偶问题：

$$\max_{\alpha} \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i \tag{35}$$

$$s.t. \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \tag{36}$$

$$C - \alpha_i - \mu_i = 0 \tag{37}$$

$$\alpha_i \geq 0 \tag{38}$$

$$\mu_i \geq 0, \quad i = 1, 2, \ldots, N \tag{39}$$

利用式 (37) 消去 μ_i，可得到如下对偶问题：

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i \tag{40}$$

$$s.t. \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \tag{41}$$

$$0 \leq \alpha_i \leq C, \quad i = 1, 2, \ldots, N \tag{42}$$

这里，C>0 成为惩罚参数，一般由应用问题决定，C 值大时对误分类的惩罚增大，C 值小时对误分类的惩罚减小

目标是：使间隔尽量大同时使误分类点的个数尽量小

分类超平面及其决策函数与线性可分线性向量机的相似，只是分量选择为 $0 \leq \alpha_j^* \leq C$。

3 序列最小最优化算法 (Sequential minimal optimization)SMO 算法来求解 α

SMO 算法是一种启发式算法，如果所有变量的解都满足此最优问题的 KKT 条件 (Karush-Kuhn-Tucker conditions)，那么这个最优化问题的解就得到了。

在这里用非线性支持向量机学习算法：即 SMO 算法要求解如下凸二次优化问题：

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j)-\sum_{i=1}^{N}\alpha_i \tag{43}$$

$$s.t. \quad \sum_{i=1}^{N}\alpha_i y_i = 0 \tag{44}$$

$$0 \le \alpha_i \le C, \quad i=1,2,\ldots,N \tag{45}$$

整个 SMO 算法包括两部分：

1.求解两个变量的二次规划的解析方法

2.选择变量的启发式方法

3.1 两个变量二次规划的求解方法

不失一般性，将假设选择的 α_1 与 α_2 作为自变量，而将其它 $\alpha_i(i=3,4,...,N)$ 进行固定，这样的 (43) 优化问题可以转写为

$$\min_{\alpha_1,\alpha_2} \quad W(\alpha_1,\alpha_2)=\frac{1}{2}\alpha_1^2y_1^2K_{11}+\frac{1}{2}\alpha_2^2y_2^2K_{22}+\frac{1}{2}\alpha_1\alpha_2y_1y_2K_{12}+\frac{1}{2}\alpha_2\alpha_1y_2y_1K_{21}-\sum_{i=1}^{N}\alpha_i \tag{46}$$

$$+\frac{1}{2}\alpha_1y_1\sum_{i=3}^{N}y_i\alpha_i(K_{i1}+K_{1i})+\frac{1}{2}\alpha_2y_2\sum_{i=3}^{N}y_i\alpha_i(K_{i2}+K_{2i})+\frac{1}{2}\sum_{i=3}^{N}\sum_{j=3}^{N}\alpha_i\alpha_jy_iy_jK_{ij} \tag{47}$$

问题可以转换为

$$\min_{\alpha_1,\alpha_2} \quad W(\alpha_1,\alpha_2)=\frac{1}{2}K_{11}\alpha_1^2+\frac{1}{2}K_{22}\alpha_2^2+y_1y_2K_{12}\alpha_1\alpha_2$$

$$-(\alpha_1+\alpha_2)+y_1\alpha_1\sum_{i=3}^{N}y_i\alpha_iK_{i1}+y_2\alpha_2\sum_{i=3}^{N}y_i\alpha_iK_{i2} \tag{48}$$

$$s.t. \quad \alpha_iy_i+\alpha_2y_2=-\sum_{i=3}^{N}y_i\alpha_i=\varsigma \tag{49}$$

$$0\le\alpha_i\le C, \quad i=1,2 \tag{50}$$

式 (49) 同时乘上 y_1

$$\alpha_1y_1^2+\alpha_2y_1y_2=y_1\varsigma \tag{51}$$

$$\alpha_1=y_1(\varsigma-\alpha_2y_2) \tag{52}$$

代入 (48),

$$W(\alpha_2)=\frac{1}{2}K_{11}(\varsigma-\alpha_2y_2)^2+\frac{1}{2}K_{22}\alpha_2^2+K_{12}y_2(\varsigma-\alpha_2y_2)\alpha_2-y_1\varsigma+(y_1y_2-1)\alpha_2$$

$$+(\varsigma-\alpha_2y_2)\sum_{i=1}^{N}y_i\alpha_iK_{i1}+y_2\alpha_2\sum_{i=3}^{N}y_i\alpha_iK_{i2} \tag{53}$$

在这里，令 $v_1=\sum_{i=1}^{N}y_i\alpha_iK_{i1}$,$v_2=\sum_{i=3}^{N}y_i\alpha_iK_{i2}$ 上式可化简为

$$W(\alpha_2)=\frac{1}{2}K_{11}(\varsigma-\alpha_2y_2)^2+\frac{1}{2}K_{22}\alpha_2^2+K_{12}y_2(\varsigma-\alpha_2y_2)\alpha_2$$

$$-y_1(\varsigma-\alpha_2y_2)-\alpha_2+v_1(\varsigma-\alpha_2y_2)+y_2v_2\alpha_2 \tag{54}$$

$W(\alpha_2)$ 对 α_2 求导数得到：

$$\frac{\partial W}{\partial\alpha_2}=K_{11}y_2(\alpha_2y_2-\varepsilon)+K_{22}\alpha_2+y_2K_{12}\varsigma-2K_{12}\alpha_2+y_1y_2-1-v_1y_2+y_2v_2$$

$$=K_{11}\alpha_2+K_{22}\alpha_2-2K_{12}\alpha_2-K_{11}\varsigma y_2+K_{12}\varsigma y_2+y_1y_2-1-v_1y_2+y_2v_2 \tag{55}$$

令其为 0，又考虑到 SVM 对数据点的预测值为

$$g(x)=\sum_{i=1}^{N}\alpha_iy_iK(x_i,x)+b \tag{56}$$

在此，令

$$E_i=g(x_i)-y_i=(\sum_{j=1}^{N}\alpha_jy_jK(x_j,x_i)+b)-y_i, \quad i=1,2 \tag{57}$$

令 (55) 为 0, 并将 $\varsigma = \alpha_1^{old} y_1 + \alpha_2^{old} y_2$ 代入，得到

$$(K_{11} + K_{22} - 2K_{12})\alpha_2^{new,unc} = (K_{11} + K_{22} - 2K_{12})\alpha_2^{old} + y_2(E_1 - E_2) \quad (58)$$

将 $\eta = K_{11} + K_{22} - 2K_{12}$ 代入，有

$$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta} \quad (59)$$

还需要对原始解进行修剪

当 $y_1 \neq y_2$ 时，上下界表示为:

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

当 $y_1 \neq y_2$ 时，上下界表示为:

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

经过修剪后的 α_2：

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

又因为公式

$$\alpha_1^{old} y_1 + \alpha_2^{old} y_2 = \alpha_1^{new} y_1 + \alpha_2^{new} y_2 = \varsigma \quad (60)$$

我们可以计算出 α_1^{new}

3.2 变量选择的方法

3.2.1 阈值 b 的更新

当 $0 < \alpha_1^{new} < C$ 时,

$$b_1^{new} = y_1 - \sum_{i=3}^{N} \alpha_i y_i K_{i1} - \alpha_1^{new} y_1 K_{11} - \alpha_2^{new} y_2 K_{21} \quad (61)$$

由式 (57) 式我们可以得到

$$E_1 = (\sum_{j=1}^{N} \alpha_j y_j K(x_j, x_1) + b^{old}) - y_1 \quad (62)$$

$$= \sum_{i=3}^{N} \alpha_i y_i K_{i1} + \alpha_1^{old} y_1 K_{11} + \alpha_2^{old} y_2 K_{21} + b^{old} - y_1 \quad (63)$$

由此可得

$$y_1 - \sum_{i=3}^{N} \alpha_i y_i K_{i1} = -E_1 + \alpha_1^{old} y_1 K_{11} + \alpha_2^{old} y_2 K_{21} + b^{old} \tag{64}$$

代入式 (61)，得

$$b_1^{new} = -E_1 - y_1 K_{11}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{21}(\alpha_2^{new} - \alpha_2^{old}) + b^{old} \tag{65}$$

同理，当 $0 < \alpha_2^{new} < C$ 时, 有:

$$b_2^{new} = -E_2 - y_1 K_{12}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{22}(\alpha_2^{new} - \alpha_2^{old}) + b^{old} \tag{66}$$

并且，每次对两个变量进行优化之后，还需要更新对应的 E_i 值，并将它们保存在列表中:

$$E_i^{new} = \sum_S y_j \alpha_j K(x_i, x_j) + b^{new} - y_i \tag{67}$$

其中，S 是所有支持向量的集合