NO. ①

支持向量机

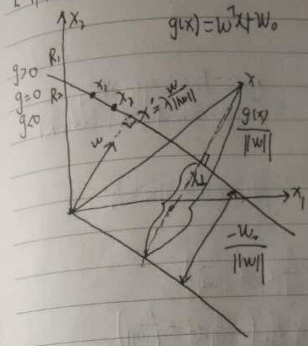
SVM原始形式推导

大纲:

1. 常见的几何性质(欧氏空间)
   - 式6.1 w和b到底有什么几何含义?
   - 划分超平面为什么可以用式(6.1)表述?
   - 式(6.2)中,高维空间里,点到平面的距离公式是怎么推出来的!
2. SVM原始公式的导出
   - 式(6.3)中,怎么理解侧栏的提示?
   - 如何从式(6.2)推到式(6.3)?
3. SVM的性质
   - 图6.2中有什么值得注意的?
   - "正中间"要怎么理解?

NO. ②

1. 常见的几何性质(欧氏空间)



$$g(x) = w^Tx + w_0$$

$$\begin{cases} w^Tx_1 + b = 0 \\ w^Tx_2 + b = 0 \Rightarrow w是法向量 \\ w^T(x_1 - x_2) = 0 \end{cases}$$

$$\begin{cases} w^Tx_1 + b = 0 \\ \lambda w^T\frac{w}{\|w\|} + b = 0 \\ \text{offset} = \frac{-b}{\|w\|} \end{cases}$$

$\frac{-b}{\|w\|}$是原点到平面的"距离"

$$\gamma = \frac{|w^Tx + b|}{\|w\|}$$

$$\gamma = \left\| \frac{w^Tx}{\|w\|^2} w - \frac{-b}{\|w\|^2} w \right\|$$

$$= \left\| \frac{w^Tx}{\|w\|^2} + \frac{b}{\|w\|^2} \right\| \|w\|$$

$\frac{w}{\|w\|}$ 单位法向量

$$= \frac{\|w^Tx + b\|}{\|w\|}$$

$$= \frac{|w^Tx + b|}{\|w\|}$$

NO. ③

2. SVM原始公式的导出(核心最大化间隔)

$$\gamma = \frac{|w^Tx_i + b|}{\|w\|}$$

$$\gamma_i = \frac{y_i(w^Tx_i + b)}{\|w\|}$$

(i) 任一数据点到平面距离
(假设w可以分开数据集)

$$d = \min_i y_i\left(\frac{w^T}{\|w\|}x_i + \frac{b}{\|w\|}\right)$$

(ii) 最接近平面的距离

$$\begin{cases} d^* = \max\limits_{w,b} \min\limits_i y_i\left(\frac{w^T}{\|w\|}x_i + \frac{b}{\|w\|}\right) \\ \gamma = 2d^* = \max\limits_{w,b} \min\limits_i 2 \times y_i\left(\frac{w^T}{\|w\|}x_i + \frac{b}{\|w\|}\right) \end{cases}$$

NO. ④

$$\gamma = 2d^* = \max_{w,b} \min_i 2 \times y_i\left(\frac{w^T}{\|w\|}x_i + \frac{b}{\|w\|}\right)$$

等价

$$\begin{cases} \max\limits_{w,b} 2d \\ \text{s.t. } y_i\left(\frac{w^T}{\|w\|}x_i + \frac{b}{\|w\|}\right) \geq d \quad \text{for } i=1,\dots,m \end{cases}$$

$\Downarrow$

引入新的定义:函数间隔 $\hat{d} = \|w\| d$

$$\max_{w,b} \frac{2\hat{d}}{\|w\|}$$

$$\text{s.t. } y_i(w^Tx_i + b) \geq \hat{d} \quad \text{for } i=1,\dots N \quad \hat{d}>0$$

实际上,$\hat{d}$取值并不影响最优化问题的解!
我们可以始终取$\hat{d}=1$

$$\max_{w,b} \frac{2}{\|w\|}$$

$$\text{s.t. } y_i(w^Tx_i + b) \geq 1, \quad i=1,2,\dots,m$$

式(6.5)

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

$$\text{s.t. } y_i(w^Tx_i + b) \geq 1$$

式(6.6)

NO. ⑤
DATE

SVM的性质：① 最大间隔超平面的存在唯一性
② 支持向量和间隔边界

存在性易得，现证唯一性
假设有两个最优解 (w_1^*, b_1^*), (w_2^*, b_2^*)
1) 证 $w_1^* = w_2^*$
根据假设，有 $\|w_1^*\| = \|w_2^*\| = c \neq 0$，设 $w = \frac{w_1^* + w_2^*}{2}$, $b = \frac{b_1^* + b_2^*}{2}$

$$c \le \|w\| = \frac{\|w_1^* + w_2^*\|}{2} \le \frac{\|w_1^*\| + \|w_2^*\|}{2} = c$$

$$\Rightarrow \|w\| = \frac{\|w_1^* + w_2^*\|}{2} = \frac{\|w_1^*\| + \|w_2^*\|}{2} \Rightarrow \underbrace{w_1^* = k w_2^*}_{\text{共线 colinear}}$$

$$\Rightarrow k = \begin{cases} 1 \\ -1 \end{cases} \Rightarrow w = \begin{cases} w_1^* = w_2^* & \text{if } k=1 \\ 0 & \text{if } k=-1 \end{cases} \Rightarrow w_1^* = w_2^*$$

2) 证 $b_1^* = b_2^*$ 设 $x_1', x_2' \in \{x_i \mid y_i = +1\}$，$x_1'', x_2'' \in \{x_i \mid y_i = -1\}$
且有

$$\begin{cases} w^T x_1' + b_1^* = 1 & ① \\ w^T x_2' + b_2^* = 1 & ② \\ w^T x_1'' + b_1^* = -1 & ③ \\ w^T x_2'' + b_2^* = -1 & ④ \end{cases} \qquad \begin{cases} b_1^* = -\frac{1}{2} w^T (x_1' + x_1'') & ①+③ \\ b_2^* = -\frac{1}{2} w^T (x_2' + x_2'') & ②+④ \end{cases}$$

$$\Rightarrow b_1^* - b_2^* = -\frac{1}{2}\left[w^T (x_1' - x_2') + w^T (x_1'' - x_2'')\right]$$

$$\begin{cases} w^T x_2' + b_1^* \ge 1 = w^T x_1' + b_1^* \\ w^T x_1' + b_2^* \ge 1 = w^T x_2' + b_2^* \\ w^T x_2'' + b_1^* \ge 1 = w^T x_1'' + b_1^* \\ w^T x_1'' + b_2^* \ge 1 = w^T x_2'' + b_2^* \end{cases} \Rightarrow \begin{cases} w^T (x_1' - x_2') = 0 \\ w^T (x_1'' - x_2'') = 0 \end{cases} \Rightarrow b_1^* = b_2^*$$

SVM对偶形式推导

拉格朗日乘子法在SVM问题上的应用

原始形式：

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{s.t. } y_i(w^T x_i+b)\geq 1,\ i=1,2,\cdots,m \qquad (\text{式 }6.6)$$

$$\Downarrow$$

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{s.t. } 1-y_i(w^T x_i+b)\leq 0 \quad i=1,2,\cdots,m$$

拉格朗日函数

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

$$\text{s.t. } \alpha_i\geq 0 \text{ for } i=1,\cdots,m \qquad (\text{式 }6.8)$$

等价问题

$$\min_{w,b}\max_{\alpha} L(w,b,\alpha) \quad \text{s.t. } \alpha_i\geq 0 \text{ for } i=1,\cdots,m$$

进一步求解SVM的拉格朗日对偶问题

$$\max_{\alpha}\min_{w,b} \frac{1}{2}\|w\|^2+\sum_{i=1}^{m}\alpha_i(1-y_i(w^T x_i+b))$$

$$\text{s.t. } a_i\geq 0,\ \text{for } i=1,\ldots,m$$

① 必要条件

$$\frac{\partial L}{\partial w}=0 \Rightarrow w=\sum_{i=1}^{m}a_i y_i x_i \qquad (\text{式 }6.9)$$

$$\frac{\partial L}{\partial b}=0 \Rightarrow \sum_{i=1}^{m}a_i y_i=0 \qquad (\text{式 }6.10)$$

② 回代得式(6.11)

$$\max_{\alpha} \sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$\text{s.t. } \sum_{i=1}^{m}a_i y_i=0,\ a_i\geq 0,\ i=1,\cdots,m \qquad (\text{式 }6.11)$$

$\frac{\partial w^T}{\partial w}=I$ 单位向量

$$\frac{\partial L}{\partial w}=0=\frac{\partial}{\partial w}\left(\frac{1}{2}w^T w+\sum_{i=1}^{m}\alpha_i-\sum_{i=1}^{m}a_i y_i(w^T x_i+b)\right)$$

$$=\frac{\partial}{\partial w}\left(\frac{1}{2}w^T w+\sum_{i=1}^{m}\alpha_i-\sum_{i=1}^{m}a_i y_i w^T x_i-\sum_{i=1}^{m}a_i y_i x_i b\right)$$

$$=\frac{\partial}{\partial w}\left(\frac{1}{2}w^T w-\sum_{i=1}^{m}a_i y_i w^T x_i\right)$$

$$=\frac{\partial}{\partial w}\left(\frac{1}{2}w^T w-w^T\sum_{i=1}^{m}a_i y_i x_i\right)$$

$$=w-\sum_{i=1}^{m}\alpha_i y_i x_i=0$$

$$\frac{\partial L}{\partial b}=0=\frac{\partial}{\partial b}\left(\frac{1}{2}w^T w+\sum_{i=1}^{m}a_i-\sum_{i=1}^{m}\alpha_i y_i(w^T x_i+b)\right)$$

$$=\frac{\partial}{\partial b}\left(\frac{1}{2}w^T w+\sum_{i=1}^{m}a_i-\sum_{i=1}^{m}\alpha_i y_i w^T x_i-\sum_{i=1}^{m}a_i y_i b\right)$$

$$=\frac{\partial}{\partial b}\left(-\sum_{i=1}^{m}a_i y_i b\right)$$

$$=\frac{\partial}{\partial b}\left(-b\sum_{i=1}^{m}a_i y_i\right)$$

$$=-\sum_{i=1}^{m}a_i y_i=0$$

$$\begin{cases} w=\sum_{i=1}^{m}a_i y_i x_i \\ 0=\sum_{i=1}^{m}a_i y_i \end{cases}$$

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

$$= \frac{1}{2} w^T w - w^T \left(\sum_{i=1}^{m} \alpha_i y_i x_i\right) - b\left(\sum_{i=1}^{m} \alpha_i y_i\right) + \sum_{i=1}^{m} \alpha_i$$

$$= \frac{1}{2} w^T \left(\sum_{i=1}^{m} \alpha_i y_i x_i\right) - w^T \left(\sum_{i=1}^{m} \alpha_i y_i x_i\right) - b \cdot 0 + \sum_{i=1}^{m} \alpha_i$$

$$= \sum_{i=1}^{m} \alpha_i + \left(\frac{1}{2} w^T - w^T\right)\left(\sum_{i=1}^{m} \alpha_i y_i x_i\right)$$

$$= \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\left(\sum_{i=1}^{m} \alpha_i y_i x_i\right)^T \left(\sum_{i=1}^{m} \alpha_i y_i x_i\right)$$

$$= \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\left(\sum_{i=1}^{m} \alpha_i y_i x_i^T\right)\left(\sum_{j=1}^{m} \alpha_j y_j x_j\right)$$

$$= \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\left(\sum_{i=1}^{m} \alpha_i y_i x_i^T \left(\sum_{j=1}^{m} \alpha_j y_j x_j\right)\right)$$

$$= \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (式6.11)$$

$$= L(\alpha)$$

补充：

$$w^* = \sum_{i=1}^{m} \alpha_i y_i x_i$$

$$b^* = y_j - \sum_{i=1}^{m} \alpha_i y_i (x_i^T x_j)$$

理论上，b^*只有一个值。故，在求得α以后，可挑选出任一支持向量，其满足：$y_j(w^{*T} x_j + b^*) - 1 = 0 \Rightarrow y_j y_j (w^{*T} x_j + b^*) - y_j = 0$

$$\Rightarrow b^* = y_j - w^{*T} x_j = y_j - \sum_{i=1}^{m} \alpha_i y_i (x_i^T x_j)$$

NO. ⑨

SMO算法

1. SMO算法的思路

重点：①每次迭代过程仅优化两个参数，有闭式解

②启发式寻找每次优化的两个参数，有效减少迭代次数

思路：①设置α列表，并设其初值为0（每个数据点对应一个α_i）

②选取两个待优化变量（为了方便，记为α_1, α_2）

③解释地求解两个变量的最优解，α_1^*, α_2^*，并更新至α的列表中

④检查更新后的α列表是否在某个精度范围内满足KKT条件，若不满足，返回②

2. SMO算法的简单实现

解释地求解两个变量的最优解，其它变量固定不变，那么

$$\max_{\alpha} \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j \qquad K_{ij} \Rightarrow x_i^T x_j$$

$$s.t. \sum_{i=1}^{m}\alpha_i y_i = 0$$

$$\alpha_i \geq 0, i=1,\cdots,m$$

$$\max_{a_1,a_2} W(\alpha_1,\alpha_2) = (\alpha_1+\alpha_2+\sum_{i=3}^{m}\alpha_i) - \frac{1}{2}K_{11}\alpha_1^2 - \frac{1}{2}K_{22}\alpha_2^2 - y_1y_2K_{12}\alpha_1\alpha_2 - y_1\alpha_1\sum_{i=3}^{m}y_i\alpha_iK_{i1} - y_2\alpha_2\sum_{i=3}^{m}y_i\alpha_iK_{i2}$$

$$= (\alpha_1+\alpha_2) - \frac{1}{2}K_{11}\alpha_1^2 - \frac{1}{2}K_{22}\alpha_2^2 - y_1y_2K_{12}\alpha_1\alpha_2 - y_1\alpha_1\sum_{i=3}^{m}y_i\alpha_iK_{i1} - y_2\alpha_2\sum_{i=3}^{m}y_i\alpha_iK_{i2}$$

$$s.t. \ a_1y_1 + a_2y_2 = -\sum_{i=3}^{m}y_i\alpha_i = \varepsilon,$$

$$a_i \geq 0, i=1,2, \text{其中} K_{ij} = x_i^T x_j$$

NO. ⑩

可以将问题再次转换成带约束的一元二次问题，

记 $v_i = \sum_{j=3}^{N}\alpha_j y_j K_{ij} = f(x_i) - \sum_{j=1}^{2}\alpha_j y_j K_{ij} - b, \quad i=1,2$

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

$- (\alpha_1+\alpha_2) + y_1v_1\alpha_1 + y_2v_2\alpha_2$

回代 $a_1y_1 + a_2y_2 = -\sum_{i=3}^{m}y_i\alpha_i = \varepsilon \Leftrightarrow \alpha_1 = (\varepsilon - y_2\alpha_2)y_1$

得 $W(\alpha_2) = \frac{1}{2}K_{11}(\varepsilon-\alpha_2y_2)^2 + \frac{1}{2}K_{22}\alpha_2^2 + y_2K_{12}(\varepsilon-a_2y_2)\alpha_2$

$- (\varepsilon - a_2y_2)y - \alpha_2 + v_1(\varepsilon - a_2y_2) + y_2v_2\alpha_2$

对α_2求偏导

$$\frac{\partial W}{\partial \alpha_2} = K_{11}\alpha_2 + K_{22}\alpha_2 - 2K_{12}\alpha_2 - K_{11}\varepsilon y_2 + K_{12}\varepsilon y_2 + y_1y_2 - 1 - v_1y_2 + y_2v_2 = 0$$

$\Downarrow$

$$(K_{11}+K_{22}-2K_{12})\alpha_2 = y_2(y_2 - y_1 + \varepsilon K_{11} - \varepsilon K_{12} + v_1 - v_2)$$

$$= y_2\Big[y_2 - y_1 + \varepsilon K_{11} - \varepsilon K_{12} + \big(f(x_1) - \sum_{j=1}^{2}y_j a_j K_{1j} - b\big) - \big(f(x_2) - \sum_{j=1}^{2}y_j a_j K_{2j} - b\big)\Big]$$

将 $\varepsilon = a_1^{old}y_1 + a_2^{old}y_2$ 代入，并设 $E_i = f(x_i) - y_i$

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

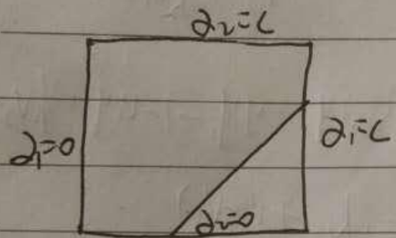
设 $\eta=K_{11}+K_{22}-2K_{12}$

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

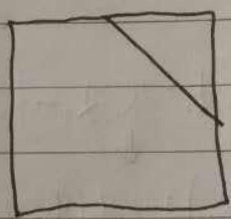
由于最优问题的定义域不是全体实数，所以在定义域内的最优解应该需要通过边界条件去判定，以获得 α_2^{new}

假设我们已获得 α_2^* 根据 $y_1\alpha_1^{new}+y_2\alpha_2^{new}=y_1\alpha_1^{old}+y_2\alpha_2^{old}$

有 $\alpha_1^{new}=\alpha_1^{old}+y_1y_2(\alpha_2^{old}-\alpha_2^{new})$



$y_1\neq y_2 \Rightarrow \alpha_1-\alpha_2=k$



$y_1=y_2 \Rightarrow \alpha_1+\alpha_2=k$

α的可行域

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

$$\alpha_1 y_1+\alpha_2 y_2=\xi$$

左图(异号)

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

$H=+\infty$

右图(同号)

$L=0$

$H=+\infty$

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