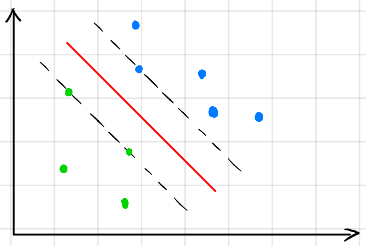
线性可分



分类面： $W^T x + b = 0$

支持面： $W^T x^+ + b = +1$

$W^T x^- + b = -1$

空间中点到面的距离： $r = \frac{|W^T x + b|}{\|W\|}$ $\longrightarrow$ 人为将这个约束成 1，得支持面到分类面距离

x $W^T x+b=0$

$W^T x + b = 1$

$\downarrow$

两支持面间隔 $margin = 2r = \frac{2}{\|W\|}$ （约束：

$W^T x^+ + b = 1$ $y_i = +1$

$W^T x^- + b = -1$ $y_i = -1$

$\Downarrow$ 合并

$y_i(W^T x_i + b) = 1$ $i \in$ support vector

$\downarrow$

间隔最大 $\max\ margin = \max \frac{2}{\|W\|}$

s.t. $y_i(W^T x_i + b) = 1$ $i \in$ support vector

$\downarrow$ 扩展到所有点

其它点分类正确的约束：

$$\begin{cases} W^T x^+ + b > 1 & y_i = +1 \\ W^T x^- + b < 1 & y_i = -1 \end{cases}$$

$\Downarrow$ 合并

$y_i(W^T x_i + b) > 1$ $i \notin$ support

$\downarrow$

$\max\ margin = \max \frac{2}{\|W\|}$

s.t. $y_i(W^T x_i + b) = 1$ $i \in$ support vector

$y_i(W^T x_i + b) > 1$ $i \notin$ support vector

$\Rightarrow$

$\max\ margin = \max \frac{2}{\|W\|}$

s.t. $y_i(W^T x_i + b) \geqslant 1$ $\forall i$

$\downarrow$

原问题：

$$\boxed{\begin{aligned} &\max \frac{2}{\|W\|} \\ &s.t.\ y_i(W^T x_i + b) \geqslant 1 \end{aligned}} \longrightarrow \boxed{\begin{aligned} &\min \frac{W^T W}{2} \\ &s.t.\ y_i(W^T x_i + b) \geqslant 1 \end{aligned}}$$

引入拉格朗日乘子

$$\mathcal{L}(W, b; \lambda) = \frac{W^T W}{2} - \sum_{i=1}^{n} \lambda_i [y_i(W^T x_i + b) - 1]$$

$$\boxed{\begin{aligned} &\min_{W,b} \frac{W^T W}{2} \\ &s.t.\ y_i(W^T x_i + b) \geqslant 1 \end{aligned}}$$

原问题

$$\min_{W,b} \max_{\lambda_i \geqslant 0} \mathcal{L}(W, b; \lambda)$$

$\Downarrow$ 转对偶问题

$$\max_{\lambda_i \geqslant 0} \min_{W,b} \mathcal{L}(W, b; \lambda)$$

用 λ 表达 W, b

$$\frac{\partial \mathcal{L}(W, b; \lambda)}{\partial W} = W - \sum_{i=1}^{n} \lambda_i y_i x_i = 0 \longrightarrow W = \underbrace{\sum_{i \in \text{support vector}}^{n}}_{} \lambda_i y_i x_i$$

$\downarrow$

只有支持向量 $\lambda_i > 0$ 约束起作用

$$\frac{\partial \mathcal{L}(W, b; \lambda)}{\partial b} = \sum_{i=1}^{n} \lambda_i y_i = 0$$

展开 $\mathcal{L}(w, b; \lambda)$

$$\mathcal{L}(w, b; \lambda) = \frac{w^T w}{2} - \sum_{i=1}^{n} \lambda_i \left[y_i (w^T x_i + b) - 1 \right]$$

$$= \frac{w^T w}{2} - w^T \underbrace{\boxed{\sum_{i=1}^{n} \lambda_i y_i x_i}}_{w} - b \underbrace{\boxed{\sum_{i=1}^{n} \lambda_i y_i}}_{0} + \sum_{i=1}^{n} \lambda_i$$

$$= \sum_{i=1}^{n} \lambda_i - \frac{w^T w}{2} = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\lambda_i \lambda_j y_i y_j x_i^T x_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \lambda_i y_i = 0$$

$$\min_{w, b} \frac{w^T w}{2}$$

$$\text{s.t.} \quad y_i (w^T x_i + b) \geqslant 1$$

原问题

$$\max_{\lambda_i \geqslant 0} \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\lambda_i \lambda_j y_i y_j x_i^T x_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \lambda_i y_i = 0$$

→ 通过 SMO 算法求解出 λ_i

- 求解 w : $w = \sum_{i=1}^{n} \lambda_i y_i x_i = \sum_{i \in support} \lambda_i y_i x_i$

- 求解 b : $y_i (w^T x_i + b) = 1$ （ $i \in$ 支持向量 ）

理论上，所有支持向量都能求出一个 b ，

实际中，通常使用平均值。

支持向量机 回归

SVR 模型：

这一项并非是推导得到的
是规范化惩罚，‖w‖过大会
造成方差过大 （过拟合）

为什么引入两个 ξ_i ，是为
了台鱼的统一

$$\min_{(w,b,\xi_i,\hat{\xi}_i)} \boxed{\frac{1}{2}\|w\|^2} + C\sum_{i=1}^{N}(\xi_i + \hat{\xi}_i)$$

$$\text{s.t.}\quad f(x_i) - y_i \leqslant \varepsilon + \xi_i$$

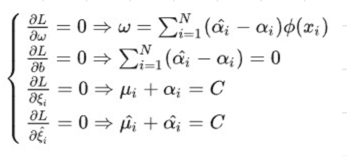
$$y_i - f(x_i) \leqslant \varepsilon + \hat{\xi}_i$$

$$\xi_i\ \hat{\xi}_i \geqslant 0$$

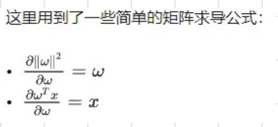
$$\downarrow$$

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

$$+\sum_{i=1}^{N}\alpha_i(w^T\phi(x_i)+b-y_i-1-\xi_i)$$

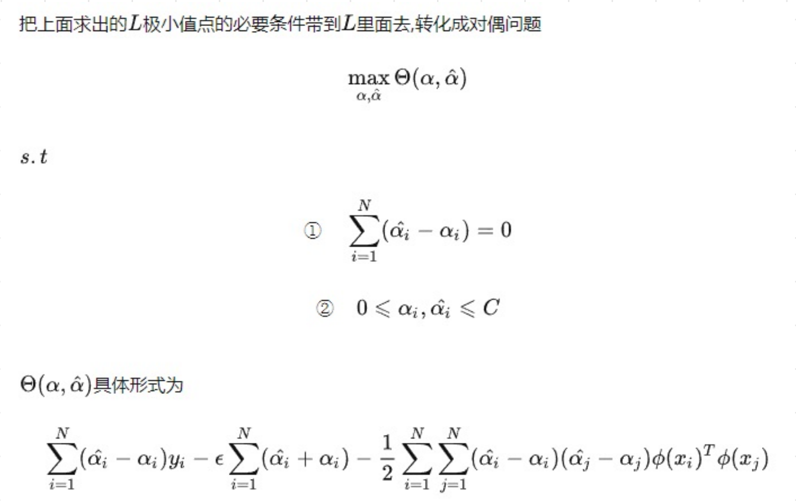


$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \Rightarrow \omega = \sum_{i=1}^{N}(\hat{\alpha}_i - \alpha_i)\phi(x_i) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N}(\hat{\alpha}_i - \alpha_i) = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \Rightarrow \mu_i + \alpha_i = C \\ \frac{\partial L}{\partial \hat{\xi}_i} = 0 \Rightarrow \hat{\mu}_i + \hat{\alpha}_i = C \end{cases}$$



这里用到了一些简单的矩阵求导公式：

- $\frac{\partial \|\omega\|^2}{\partial \omega} = \omega$
- $\frac{\partial \omega^T x}{\partial \omega} = x$



把上面求出的L极小值点的必要条件带到L里面去,转化成对偶问题

$$\max_{\alpha,\hat{\alpha}} \Theta(\alpha,\hat{\alpha})$$

$s.t$

$$① \quad \sum_{i=1}^{N}(\hat{\alpha}_i - \alpha_i) = 0$$

$$② \quad 0 \leqslant \alpha_i, \hat{\alpha}_i \leqslant C$$

$\Theta(\alpha,\hat{\alpha})$具体形式为

$$\sum_{i=1}^{N}(\hat{\alpha}_i - \alpha_i)y_i - \epsilon\sum_{i=1}^{N}(\hat{\alpha}_i + \alpha_i) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j)\phi(x_i)^T\phi(x_j)$$

zhihu/p/109824716