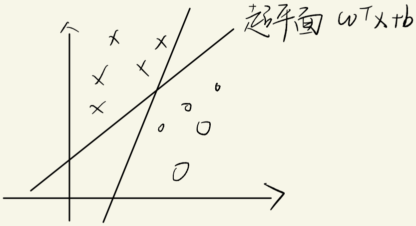
SVM 白板推导 ch1-ch4

SVM 三宝：间隔，对偶，核技巧

$$\begin{cases} \text{hard-margin SVM} \\ \text{soft-margin SVM} \\ \text{kernel SVM} \end{cases}$$



$f(w) = \text{sign}(w^Tx+b) \Longrightarrow$ 判别模型

最大间隔分类器

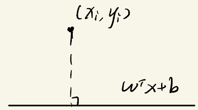
$\hookrightarrow \max \ \text{margin}(w,b)$

$$\text{s.t.} \begin{cases} w^Tx_i+b>0, & y_i=+1 \\ w^Tx_i+b<0, & y_i=-1 \end{cases}$$

$\Longleftrightarrow \quad y_i(w^Tx_i+b)>0, \ \text{for} \ \forall \ i=1,\cdots,N$

$$\text{margin}(w,b) = \min_{\substack{w,b,x_i \\ i=1,\cdots,N}} \text{distance}(w,b,x_i)$$

$$= \min_{w,b,x_i} \frac{1}{\|w\|}|w^Tx_i+b|$$



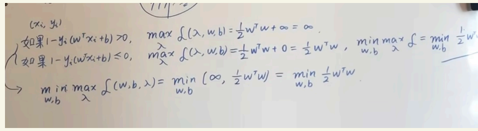
$\text{distance} = \frac{1}{\|w\|}|w^Tx+b|$

primal problem $\Longleftarrow$

$$\Rightarrow \begin{cases} \max\limits_{w,b} \min\limits_{x_i} \frac{1}{\|w\|} y_i(w^Tx_i+b) = \max\limits_{w,b} \frac{1}{\|w\|} \underbrace{\min_{\substack{x_i \\ i=1\cdots N}} y_i(w^Tx_i+b)}_{=r} \\ \text{s.t.} \ y_i(w^Tx_i+b)>0 \Rightarrow \exists \ r>0, \ \text{st.} \min\limits_{\substack{x_i\cdots \\ y_i\cdots \\ i=1,\cdots N}} y_i(w^Tx_i+b)=r \end{cases}$$

$$\Rightarrow \begin{cases} \max\limits_{w,b} \frac{1}{\|w\|} \\ \text{s.t.} \ \min y_i(w^Tx_i+b)=1 \\ \qquad \Updownarrow \\ \qquad y_i(w^Tx_i+b) \geq 1 \end{cases} \Rightarrow \boxed{\begin{cases} \min \frac{1}{2}w^Tw \\ \text{s.t.} \ y_i(w^Tx_i+b) \geq 1, \ \text{for} \ \forall i=1,\cdots,N \end{cases}}$$

带约束 $\Rightarrow \begin{cases} \min \frac{1}{2} w^T w \\ s.t.\ y_i(w^T x_i + b) \geq 1 \iff 1 - y_i(w^T x_i + b) \leq 0 \\ \quad \text{for } \forall i = 1, \cdots, N \end{cases}$



$$\mathcal{L}(w, b, \lambda) = \frac{1}{2} w^T w + \sum_{i=1}^{N} \lambda_i (1 - y_i(w^T x_i + b))$$

无约束 $\begin{cases} \min\limits_{w,b} \max\limits_{\lambda} \mathcal{L}(w, b, \lambda) \\ s.t.\ \lambda_i \geq 0 \end{cases}$

$\min\max \mathcal{L}$ "$\geq$" $\max\min \mathcal{L}$

弱对偶关系(不等)

强对偶关系(等于) "="

强对偶问题 $\begin{cases} \max\limits_{\lambda} \underline{\min\limits_{w,b} \mathcal{L}(w, b, \lambda)} \\ s.t.\ \lambda_i \geq 0 \end{cases} \Rightarrow \boxed{\sum_{i=1}^{N} \lambda_i y_i = 0}, \quad \boxed{w = \sum_{i=1}^{N} \lambda_i y_i x_i}$

w, b 最小化 $\mathcal{L}(w,b,\lambda)$

$$\begin{cases} \max\limits_{\lambda} -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j + \sum_{i=1}^{N} \lambda_i \\ s.t.\ \lambda_i \geq 0 \\ \quad \sum_{i=1}^{N} \lambda_i y_i = 0 \end{cases}$$

硬间隔模型求解推导:

$\min_{w,b} \mathcal{L}(w,b,\lambda)$ 对 w, b 分别求偏导　　强对偶关系时，$\min_{w,b} \mathcal{L}$ 可以得到下述式

$\Longleftrightarrow$

$$\text{Let } \frac{\partial \mathcal{L}}{\partial b} = \frac{\partial}{\partial b}\left(\sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{N}\lambda_i y_i (w^T x_i + b)\right)$$

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

$$= -\sum_{i=1}^{N}\lambda_i y_i \overset{\text{令偏导为零}}{\triangleq} 0$$

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial b} \triangleq 0 \Rightarrow \boxed{\sum_{i=1}^{N}\lambda_i y_i = 0}$$ 将其代入 $\mathcal{L}(w,b,\lambda)$ 中

$$\mathcal{L}(w,b,\lambda) = \frac{1}{2}w^T w + \sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{N}\lambda_i y_i w^T x_i - \sum_{i=1}^{N}\lambda_i y_i b$$

$$= \frac{1}{2}w^T w + \sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{N}\lambda_i y_i w^T x_i$$

$$\text{Let } \frac{\partial \mathcal{L}(w,b,\lambda)}{\partial w} = w - \sum_{i=1}^{N}\lambda_i y_i x_i \triangleq 0 \Rightarrow \boxed{w = \sum_{i=1}^{N}\lambda_i y_i x_i}$$ 代入上中

$$\min \mathcal{L}(w,b,\lambda) = \frac{1}{2}\left(\sum_{i=1}^{N}\lambda_i y_i x_i\right)^T\left(\sum_{j=1}^{N}\lambda_j y_j x_j\right) - \sum_{i=1}^{N}\lambda_i y_i \left(\sum_{j=1}^{N}\lambda_j y_j x_j\right)^T x_i + \sum_{i=1}^{N}\lambda_i$$

$$= \frac{1}{2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j x_i^T x_j\right) - \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j x_j^T x_i + \sum_{i=1}^{N}\lambda_i$$

$$= -\frac{1}{2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j x_i^T x_j\right) + \sum_{i=1}^{N}\lambda_i$$

KKT条件

$$
\begin{cases}
\frac{\partial L}{\partial w}=0 \quad \frac{\partial L}{\partial b}=0 \quad \frac{\partial L}{\partial \lambda}=0 \\
\boxed{\lambda_i(1-y_i(w^T x_i+b_i))=0} \\
\lambda_i \geq 0 \\
1-y_i(w^T x_i+b_i) \leq 0
\end{cases}
$$

互补松弛定理

slackness complementary

(求 b^*)

原问题与对偶问题具有强对偶关系

$\Longleftrightarrow$ 满足KKT条件

上一节已求 $\boxed{w^* = \sum_{i=0}^{N} \lambda_i y_i x_i}$ → 均为(x_i, y_i)的线性组合

data

$\exists\ (x_k, y_k),\ s.t.\ 1-y_k(w^T x_k+b)=0$

$y_k \cdot y_k(w^T x_k+b)=1 \cdot y_k \quad (y_k^2=1)$

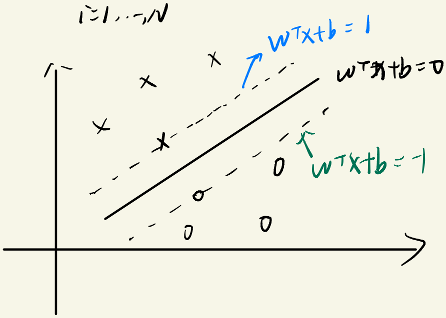
$$
\Rightarrow \boxed{\begin{aligned} b^* &= y_k - w^T \cdot x_k \\ &= y_k - \sum_{i=0}^{N} \lambda_i y_i x_i^T x_k \end{aligned}}
$$

不在虚线上，$\lambda_i=0$

Soft-Margin SVM 软间隔

Data $= \{(x_i, y_i)\}_{i=1}^{N}$, $x_i \in \mathbb{R}^p$, $y_i \in \{+1, -1\}$

$$\begin{cases} \min\limits_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{N} \max\{0, 1 - y_i(w^T x_i + b)\} \\ \text{s.t. } y_i(w^T x + b) \geq 1 - \xi_i \\ \quad i = 1, \dots, N \end{cases}$$

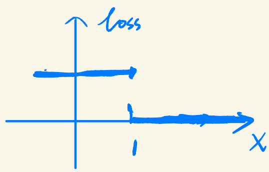


Soft: 允许一点点错误

$\hookrightarrow \min \frac{1}{2} w^T w + loss$

① $loss = \sum_{i=1}^{N} I\{\underbrace{y_i(w^T x + b) < 1}_{\text{不连续}}\}$

令 $z = y(w^T x + b)$ $\quad$ loss fun $= \begin{cases} 1, & z < 1 \\ 0, & ow \end{cases}$



⊗

引入 $\xi_i = 1 - y_i(w^T x_i + b)$, $\xi_i \geq 0$

$$\boxed{\begin{cases} \min\limits_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \\ \text{s.t. } y_i(w^T x + b) \geq 1 - \xi_i \\ \quad \xi_i \geq 0 \end{cases}}$$

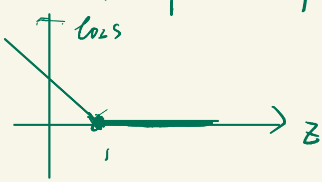
② loss 距离

如果 $y_i(w^T x_i + b) \geq 1$, $loss = 0$

如果 $y_i(w^T x_i + b) < 1$, $loss = 1 - y_i(w^T x + b)$

$\Rightarrow loss = \max\{0, 1 - \underbrace{y_i(w^T x + b)}_{z}\}$

$\Rightarrow loss_{max} = \{0, 1 - z\}$



$\boxed{\text{hinge loss}}$