支持向量机-硬间隔分类器

2018年11月18日 星期日　下午2:21

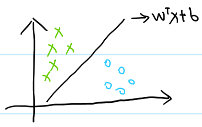
SVM有三宝：间隔，对偶，核技巧

SVM：hard-margin SVM
　　　soft-margin SVM
　　　kernel SVM

一、模型定义

SVM最初被提来解决二分类问题

几何意义：



能够做对的超平面有很多

而 SVM 就是找到其中一个最好的超平面

$f(w) = \text{sign}(w^Tx+b)$ → 判别模型

思想：

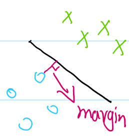
SVM 认为~~最好~~的超平面就是离两类最近的超平面

↓

最大间隔分类器

↓

$$\begin{cases} \max \ \text{margin}(w,b) \\ \text{s.t.} \begin{cases} w^Tx_i+b>0, & y_i=+1 \\ w^Tx_i+b<0, & y_i=-1 \end{cases} \\ \quad \Rightarrow y_i(w^Tx_i+b)>0 \quad \text{for } \forall i=1,\cdots,N \\ \text{margin}(w,b) = \min\limits_{\substack{w,b \\ x_i \\ i=1,\cdots,N}} \text{distance}(w,b,x_i) \end{cases}$$



margin：样本到超平面最短距离

点到直线距离公式

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

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

$|w^Tx_i+b| \xlongequal{\text{由限制条件}} y_i(w^Tx_i+b)$

$\Rightarrow \exists \gamma>0, \ \text{s.t.} \ \min\limits_{\substack{x_i,y_i \\ i=1,\cdots,N}} y_i(w^Tx_i+b)=\gamma$

$\gamma=1$　　γ等于什么无所谓

$$\Rightarrow \begin{cases} \max\limits_{w,b} \frac{1}{\|w\|} \min\limits_{\substack{x_i \\ i=1,\cdots,N}} y_i(w^Tx_i+b) \\ \text{s.t.} \quad \min\limits_{\substack{x_i,y_i \\ i=1,\cdots,N}} y_i(w^Tx_i+b)=\gamma \end{cases}$$

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

$$y_i(w^T x_i + b) \geq 1,\ i=1,\cdots,N$$

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

convex optimization

N个约束

二、模型求解

直接求解计算量很大

原问题 $\Longrightarrow$ 对偶问题

(primal problem) (dual problem)

带约束的优化问题，应该最先想到把拉格朗日函数写出来

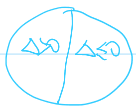
$$\mathcal{L}(w,b,\lambda) = \frac{1}{2} w^T w + \sum_{i=1}^{N} \lambda_i \underbrace{(1 - y_i(w^T x_i + b))}_{\Delta \leq 0}$$

$\lambda_i \geq 0$

则原问题

带约束原问题

无约束原问题



$$\begin{cases} \min\limits_{w,b} \max\limits_{\lambda} \mathcal{L}(w,b,\lambda) \\ s.t.\ \lambda_i \geq 0,\ i=1,\cdots,N \end{cases}$$

→ 其实这里是对有约束优化的理解

如果 $1-y_i(w^T x_i+b)>0$, $\max\limits_{\lambda} \mathcal{L}(\lambda,w,b) = \frac{1}{2} w^T w + \infty = \infty$

如果 $1-y_i(w^T x_i+b) \leq 0$, $\max\limits_{\lambda} \mathcal{L}(\lambda,w,b) = \frac{1}{2} w^T w + 0 = \frac{1}{2} w^T w$

$$\min_{w,b} \max_{\lambda} \mathcal{L} = \min_{w,b} \frac{1}{2} w^T w$$

$$\min_{w,b} \max_{\lambda} \mathcal{L}(w,b,\lambda) = \min_{w,b} (\infty, \frac{1}{2} w^T w) = \min_{w,b} \frac{1}{2} w^T w$$

强对偶关系

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

$\min \max \mathcal{L} \geq \max \min \mathcal{L}$ 弱对偶关系

而我们想要强对偶关系 "="，可证

$\min\limits_{w,b} \mathcal{L}(w,b,\lambda)$ 是无约束优化问题，求导即可

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

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

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

$$= -\sum_{i=1}^{N}\lambda_i y_i = 0$$ 代入 $\mathcal{L}(w,b,\lambda)$

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

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

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

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{1}{2}\cdot 2\cdot w - \sum_{i=1}^{N}\lambda_i y_i x_i = 0$$

$$\Rightarrow w = \sum_{i=1}^{N}\lambda_i y_i x_i$$ 代入 $\mathcal{L}(w,b,\lambda)$

$$\mathcal{L}(w,b,\lambda) = \frac{1}{2}\Big(\sum_{i=1}^{N}\underbrace{\lambda_i y_i}_{\text{实}}\underbrace{x_i}_{\text{向量}}\Big)^T\Big(\sum_{i=1}^{N}\lambda_i y_i x_i\Big) - \sum_{i=1}^{N}\lambda_i y_i \Big(\sum_{j=1}^{N}\lambda_j y_j x_j\Big)^T x_i + \sum_{i=1}^{N}\lambda_i$$

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

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

$$\Rightarrow \begin{cases} \min\limits_{\lambda} \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j x_i^T x_j - \sum_{i=1}^{N}\lambda_i \\ s.t.\ \ \lambda_i \geq 0 \\ \qquad \sum_{i=1}^{N}\lambda_i y_i = 0 \end{cases}$$

KKT条件　　　　原对偶问题具有强对偶关系

$\Leftrightarrow$ 满足 KKT 条件

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

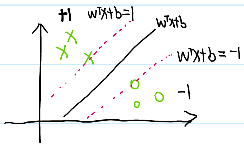
$\rightarrow$ 松弛互补条件

$y_i(w^T x_i+b)=1$，λ_i 起作用，$\lambda_i>0$

$y_i(w^T x_i+b)>1$，$\lambda_i=0$，λ_i 不起作用

满足了KKT条件，就可以求导得到 w^*，b^*

$$\boxed{w^* = \sum_{i=0}^{N}\lambda_i y_i x_i}$$



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

$$\Rightarrow y_k(w^T x_k+b)=1$$

$$\Rightarrow y_k^2(w^T x_k+b)=y_k$$ $y_k=\{1,-1\}$

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

决策函数：

$$f(x) = \mathrm{sign}(w^{*T}x + b^*)$$