第7章 支持向量机

7.1 线性可分支持向量机 & 硬间隔最大化

假设空间:

$$w \cdot x + b = 0$$

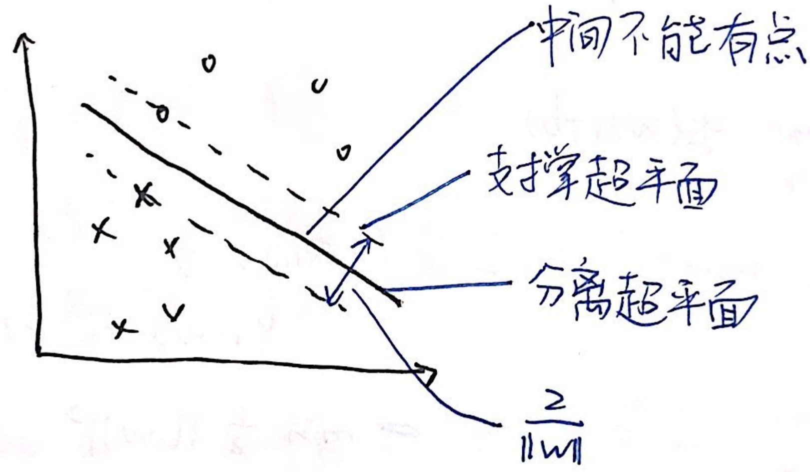
决策函数

$$f(x) = \mathrm{sgn}(w^* \cdot x + b^*) \begin{cases} +1 \\ -1 \end{cases}$$

模型选择依据 (SVM 超平面唯一)

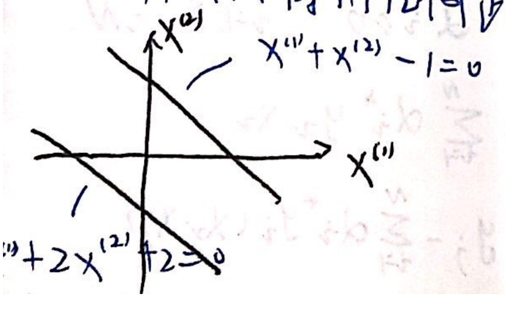
(感知机～不唯一)

硬间隔最大化



$$\left.\begin{array}{ll} \text{函数距离} & |w \cdot x^* + b| \\ \text{几何距离} & \dfrac{|w \cdot x^* + b|}{\|w\|} \end{array}\right\} \text{求点 } x^* \text{ 到分离超平面的距离}$$

为什么得用几何距离,而不能用函数距离呢?



若是函数距离 则为 $|0+0-1|=1$

会因为系数产生改变. $|0+0+2|=2$

$\therefore$ ✗

总结：衡量一个点到不同的超平面用几何距离

不能用函数距离

衡量不同的点到一个给定的超平面

可用函数距离

模型选择依据

硬间隔最大化

去掉绝对值方法

① $\frac{|w\cdot x+b|}{\|w\|} \Rightarrow \frac{y_i(w\cdot x_i+b)}{\|w\|}$

② $\max\limits_{w,b} \min\limits_{i} \frac{y_i(w x_i+b)}{\|w\|}$ ~~先找到到超平面最小的点~~ 距离

先找 最小的点，再让这个点到超平面的距离最大

③ $\max\limits_{w,b} \frac{1}{\|w\|} \min\limits_{i} y_i(w x_i+b)$

为了使求解 $\min\limits_{i} y_i(w x_i+b)$ 简单，我们认为

$y_i(w x_i+b) \geq 1$

接着，找 $\max\limits_{w,b} \frac{1}{\|w\|}$ 等价于求 $\min \frac{1}{2}\|w\|^2$

④ 最小化问题

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

s.t. $y_i(w\cdot x_i+b)-1 \geq 0$

$i=1,2,\cdots N$

对偶问题

$\min\limits_{\alpha} \frac{1}{2}\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)-\sum\limits_{i=1}^{N}\alpha_i$

s.t. $\sum\limits_{i=1}^{N}\alpha_i y_i=0$

$\alpha_i \geq 0$, $i=1,2,\cdots,N$

KKT $\begin{cases} w^* = \sum\limits_{i=1}^{N}\alpha_i^* y_i x_i \\ b^* = y_j - \sum\limits_{i=1}^{N}\alpha_i^* y_i (x_i\cdot x_j) \end{cases}$

可以让原问题转为对偶问题，求得α.

再用 KKT 求出原问题的最优解 w^*, b^*

$$\underbrace{w^* \cdot x}_{x_i \cdot x} + \underbrace{b^*}_{x_i \cdot x_j} = 0$$

7.2 线性支持向量机与软间隔最大化

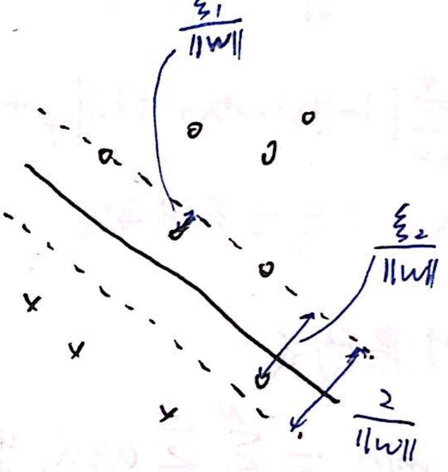
提出意义：
① 若原数据集是线性不可分的.
② 支持向量对硬间隔中的分离超平面起着决定性因素，要削弱这个影响.

最优化问题.

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

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

$$\xi_i \geq 0, \quad i=1,2,\cdots,N$$



理解：间隔之间也可以出现点，但是要对他们进行惩罚.

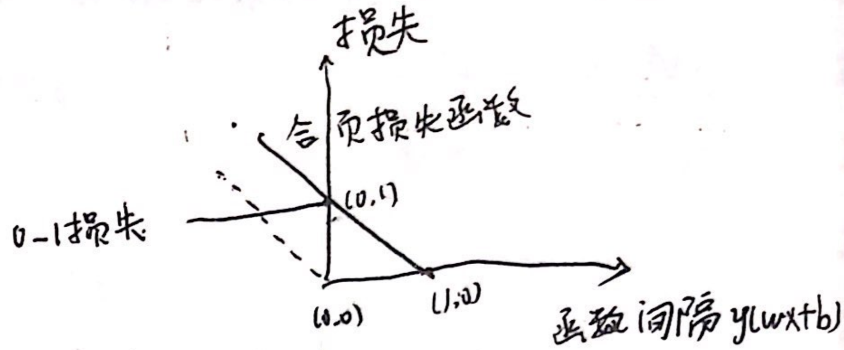
$$\min_{w,b,\xi} \underbrace{\frac{1}{2}\|w\|^2}_{\text{表示最大化超平面间隔}} + C\sum_{i=1}^{N}\xi_i$$

C → 在这两个之间做出权衡

$\sum_{i=1}^{N}\xi_i$ → 表示最小化越界的点.

那我们如何表示这个惩罚呢？

用 合页损失函数



$\geq 1-\underbrace{\xi_i}_{\geq 0}=0$

$\xi_i=1-y(w\cdot x+b)$ 代入

原问题

$$\frac{1}{2}\|w\|^2+C\sum\xi_i$$

$$\Downarrow$$

$$\sum_{i=1}^{N}[1-y_i(w\cdot x_i+b)]_+ +\lambda\|w\|^2$$

λ 由 C 与 $\frac{1}{2}$ 共同决定

对偶形式

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

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

$$0\leq\alpha_i\leq C,\quad i=1,2,\cdots N$$

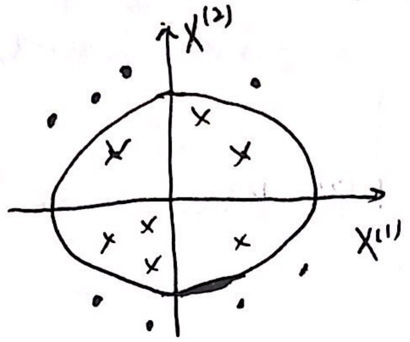
KKT

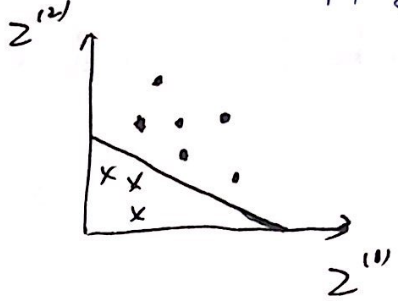
$$w^*=\sum_{i=1}^{N}\alpha_i^*\cdot y_i\cdot x_i$$

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

7.3 非线性支持向量与核函数

问题
不知道用什么曲面





$x^{(1)^2}+x^{(2)^2}$ 让 $z=\varphi(x)$，$z^{(1)}=x^{(1)^2}$ $z^{(2)}=x^{(2)^2}$

$$x_i\cdot x_j \Rightarrow \underbrace{K(x_i,x_j)}_{\text{核函数}}=\varphi(x_i)\cdot\varphi(x_j)$$

好处：

不用写出 φ 具体的形式，只用写出 K 的形式即可

原
~~对偶~~问题 将 $x\to\varphi(x)$

$$\min \frac{1}{2}\|w\|^2+C\sum \xi_i$$

$$\text{s.t.}\quad y_i[w\cdot\varphi(x)]\geq 1-\xi_i$$

$$\xi_i\geq 0$$

分离超平面：$w\cdot\varphi(x)+b=0$

对偶问题

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

$$\text{s.t.}\quad \sum_{i=1}^{N}\alpha_i y_i=0\quad,\ 0\leq\alpha_i\leq C\ ,\ i=1,2,\cdots N$$

决策函数

$$f(x)=\text{sign}\left(\sum_{i=1}^{N}\alpha_i^* y_i K(x\cdot x_i)+b^*\right)$$

7.4 序列最小优化算法

背景：7.3是最普通的形式，如果 $\begin{cases} ① & \overset{\text{为}7.1}{C \to \infty, K(x_i, x_j) = x_i \cdot x_j} \\ ② & \overset{\text{为}7.2}{K(x_i \cdot x_j) = x_i \cdot x_j} \end{cases}$

对偶问题

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

$$s.t. \quad \sum_{i=1}^{N} \alpha_i \cdot y_i = 0 \quad , 0 \le \alpha_i \le C. \quad i=1,2,\cdots,N.$$

理解：

我们总共要优 N 个 α，但是同时优化 N 个 α 难度太大，

那么如果一个优化一个 α 的话，可以吗？

不行。$\because \sum_{i=1}^{N} \alpha_i y_i = 0 \qquad \alpha_1 = \dfrac{-\sum_{i=2} \alpha_i y_i}{y_1}$ α_1 由其它 $N-1$ 个 α 决定

$\therefore$ 不能一次优化一个

因此出现了给定 α 初值，一次优化 (α_1, α_2) 2个 α，直到收敛。具体看书

软间隔最大化对偶问题

推导过程

这是个凸优化~~函数~~问题

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

$$s.t.\quad y_i(w\cdot x_i+b) \geq 1-\xi_i,\ i=1,\cdots,N$$

$$\xi_i \geq 0,\ i=1,\cdots,N$$

w是唯一的.(即分离超平面斜率唯一)

b是不唯一的 ($\because$ ξ_i的解也不唯一, 但是$\sum\xi_i$是唯一的

② 写出拉格朗日函数

根据拉格朗日概念,
凸优化问题的对偶形式
的最优解与原问题最优
解一致,且满足KKT条件

$$L(w,b,\xi,\alpha,\mu) = \frac{1}{2}\|w\|^2 + C\sum\xi_i +$$
$$\sum\alpha_i[1-\xi_i-y_i(w\cdot x_i+b)] +$$
$$\sum\mu_i(-\xi_i)$$
$$= \frac{1}{2}\|w\|^2 + C\sum\xi_i - \sum\alpha_i[y_i(w\cdot x_i+b)-1+\xi_i] - \sum\mu_i\xi_i$$

如何写出拉格朗日函数

① L(关于哪些变量,+ 有n个约束条件就有n个拉格朗日乘子) =

② 最小化函数照抄 $+\sum_i$ 拉格朗日乘子 $\times$ 约束条件

约束条件一定要写成 小于等于0形式

如: $1-\xi_i-y_i(w\cdot x_i+b) \leq 0$

③ 整理

③ 写出对偶问题

1) 用拉格朗日函数，分别对参数求导，再让其 $=0$

$$L(w,b,\xi,\alpha,u)=\frac{1}{2}\|w\|^2+C\sum\xi_i-\sum\alpha_i[y_i(w\cdot x_i+b)-1+\xi_i]-\sum u_i\xi_i$$

$$\nabla_w L=w-\sum\alpha_i\cdot y_i\cdot x_i=0 \Rightarrow w=\sum\alpha_i\cdot y_i\cdot x_i$$

$$\nabla_b L=-\sum\alpha_i\cdot y_i=0$$

$$\nabla_{\xi_i} L=C-\alpha_i-u_i=0$$

2) ~~代入原问题的最优化函数~~ 代入之前写的拉格朗日函数

$$L=\frac{1}{2}(\sum\alpha_i\cdot y_i\cdot x_i)\cdot(\sum\alpha_i\cdot y_i\cdot x_i)+C\sum\xi_i$$
$$-(\sum\alpha_i y_i x_i)(\sum\alpha_i y_i x_i)-b\underbrace{\sum\alpha_i y_i}_{=0}+\sum\alpha_i-\sum\alpha_i\xi_i-\sum u_i\xi_i$$

整理得（根据3个对参数求导的式子来整理）

$$L=-\frac{1}{2}(\sum\alpha_i\cdot y_i\cdot x_i)(\sum\alpha_i\cdot y_i\cdot x_i)+\sum\alpha_i$$

（ξ没了，所以最大化的问题是对α_i求解）

↳ 对偶问题 $\max\limits_{w,u}\min L$

$\because (\sum\limits_i\alpha_i y_i x_i)(\sum\limits_j\alpha_j\cdot y_j\cdot x_j)$

$$=\sum_i\sum_j\alpha_i\alpha_j\cdot y_i y_j\cdot x_i\cdot x_j$$

$\therefore$ 关于α的最大化问题即变为（拉格朗日形式）

$$\max_\alpha -\frac{1}{2}\sum_i\sum_j\alpha_i\alpha_j y_i y_j\cdot x_i\cdot x_j+\sum\alpha_i$$

s.t. $\sum\alpha_i\cdot y_i=0 \quad \alpha_i\geq 0$

$C-\alpha_i-u_i=0 \quad u_i\geq 0$

④写出KKT形式.

$$\begin{cases} \nabla_w L = w - \sum \alpha_i y_i x_i = 0 \\ \nabla_b L = -\sum \alpha_i y_i = 0 \\ \nabla_{\xi_i} L = C - \alpha_i - u_i = 0 \end{cases}$$

$$\begin{cases} \alpha_i [y_i (w \cdot x_i + b) - 1 + \xi_i] = 0 \\ u_i \xi_i = 0 \end{cases}$$

$$\begin{cases} y_i (w \cdot x_i + b) - 1 + \xi_i \geq 0 \\ \xi_i \geq 0 \end{cases}$$

$$\begin{cases} \alpha_i \geq 0 \\ u_i \geq 0 \end{cases}$$

最大间隔分离超平面存在唯一性

前提：训练数据集T线性可分.

① 最大间隔分离超平面是由优化问题得到的.

$$\min \ \frac{1}{2}\|w\|^2$$

$$s.t. \quad y_i(w\cdot x_i+b)-1\geq 0$$

存在唯一性 即上面优化问题的解是存在且唯一的

② 存在性

1) 线性可分（训练数据集）
2) 目标函数有下界
3) 数据集有正类点又有负类点 ∴ $(w,b)=(0,b)$ 不是最优解 ∴ $w^*\neq 0$

③ 唯一性（反证法）

1) 假设有2个最优解

$$(w_1^*, b_1^*) \quad (w_2^*, b_2^*)$$

证 $w_1^*=w_2^*$ $\quad b_1^*=b_2^*$

2) 由1)可得

step1 $\|w_1^*\|=\|w_2^*\|=c$

$$w=\frac{w_1^*+w_2^*}{2} \qquad b=\frac{b_1^*+b_2^*}{2}$$

step2 ∵ w, b 确定了. ∴ 现在要验证 w 与 b 满不满足约束条件

$$y_i\left(\frac{w_1^*+w_2^*}{2}\cdot x_i+\frac{b_1^*+b_2^*}{2}\right)-1$$

$$=\frac{1}{2}\left[y_i(w_1^*\cdot x_i+b_1^*)-1+y_i(w_2^*\cdot x_i+b_2^*)-1\right]\geq 0$$

∵ w^* 与 b^* 在可行域里（满足 N 个约束条件）

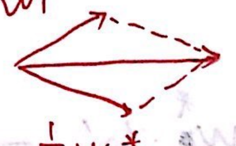
Step3 （∵ w_1^* 之前被假设为最优解，若 $\|w_1^*\| > \|w\|$ 那么 w 才是最优解

即 $\|w_1^*\| \le \|w\|$ 　　假设 $w_1^* = c$

即 $c \le \|w\| = \|\frac{1}{2}w_1^* + \frac{1}{2}w_2^*\| \le \frac{1}{2}\|w_1^*\| + \frac{1}{2}\|w_2^*\| = c$

根据三角不等式

∴有 $\|\frac{1}{2}w_1^* + \frac{1}{2}w_2^*\| = \frac{1}{2}\|w_1^*\| + \frac{1}{2}\|w_2^*\|$



∴ w_1^* 与 w_2^* 一定是同方向的

Step4

即 $w_1^* = \lambda w_2^*$ 　$\lambda \ge 0$ 　（第1部分证明完毕）

$\Rightarrow$ $w_1^* = w_2^*$

又有 $\|w_1^*\| = \|w_2^*\|$

（∵ 根据之前的假设 w_1^* 与 w_2^* 都是最优解

现在要证明 b 的最优解是唯一的

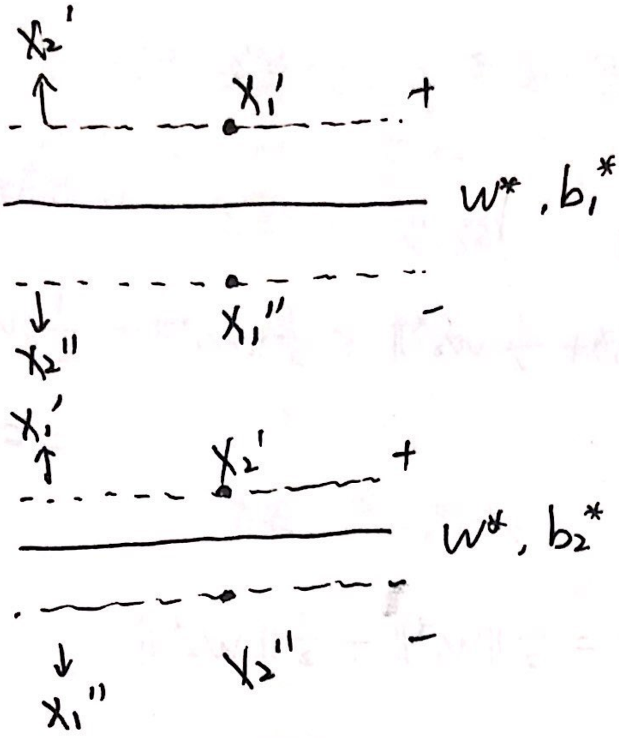
分析：因为 b 出现在不等式约束中，且证明的是唯一性问题

因此我们要从这个不等式什么时候相等入手

即位于支撑超平面上的点，代入不等式中相等.

之前假设了2个最优解. 并证明了 $w_1^* = w_2^*$

∴有2个假设空间，如下图



(1)

$$\left.\begin{aligned} w^*\cdot x_1' + b_1^* - 1 &= 0 \\ w^*\cdot x_2' + b_1^* - 1 &\geq 0 \end{aligned}\right\} \Rightarrow w^*\cdot x_1' \leq w^*\cdot x_2' \quad \text{— ①}$$

$$\left.\begin{aligned} w^*\cdot x_2' + b_2^* - 1 &= 0 \\ w^*\cdot x_1' + b_2^* - 1 &\geq 0 \end{aligned}\right\} \Rightarrow w^*\cdot x_2' \leq w^*\cdot x_1' \quad \text{— ②}$$

由①②得

$$w^*\cdot(x_1' - x_2') = 0$$

同理

$$w^*(x_1'' - x_2'') = 0$$

(2)

$$\left\{\begin{aligned} w^*\cdot x_1' + b_1^* - 1 &= 0 \\ -[w^*\cdot x_1'' + b_1^*] - 1 &= 0 \end{aligned}\right. \Rightarrow \begin{aligned} b_1^* &= 1 - w^*\cdot x_1' \quad \text{— ③} \\ b_1^* &= -1 - w^*\cdot x_1'' \quad \text{— ④} \end{aligned}$$

$\frac{1}{2}\times③ + \frac{1}{2}\times④$

$$\Rightarrow b_1^* = -\frac{1}{2} w^*(x_1' + x_1'') \quad ⑤$$

同理

$$b_2^* = -\frac{1}{2} w^*(x_2' + x_2'') \quad ⑥$$

③现要证 b_1^* 与 b_2^* 相等

∴现在将 ⑤ − ⑥

即 $b_1^* - b_2^* = -\frac{1}{2}[w^* \cdot (x_1' - x_2')]$

$+ (-\frac{1}{2}[w^* \cdot (x_1'' - x_2'')]$

$= 0$

即 $b_1^* = b_2^*$

$\therefore (w_1^*, b_1^*) = (w_2^*, b_2^*)$

得证.