support

vector

machine

SVM

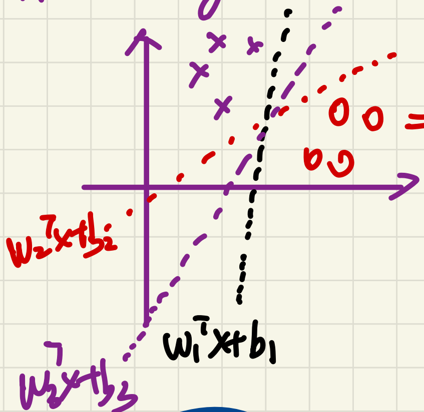
Support vector machine.

支持向量机

(硬间隔SVM)

Hard-margin SVM:

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



⇒ 存在许多超平面可以对数据集进行分类。

而我们的目标是为了寻找最好的那个超平面。

寻找泛化误差最小的超平面.(优化问题)

⇓

使所有数据点中到面距离最小点的最大的超平面.

点到面的distance $= \frac{|w^T x + b|}{\|w\|}$

即: max margin(w, b)

这里
$$\begin{cases} \text{margin} = \min\limits_{\substack{w,b \\ x_i,\, i=1\cdots N}} \dfrac{|w^T x_i + b|}{\|w\|} \\ \text{s.t. } y_i(w^T x_i + b) > 0 \quad \text{for } i=1\cdots N \end{cases}$$

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

$$\Rightarrow = \max_{w,b} \frac{1}{\|w\|} \min_{\substack{x_i \\ i=1\cdots N}} y_i(w^T x_i + b)$$

⇓

$$\exists\, r > 0, \quad \min_{\substack{x_i \\ y_i \\ i=1\cdots N}} y_i(w^T x_i + b) = r$$

令 $r = 1$

$$\therefore \max_{w,b} \frac{1}{\|w\|} = \max_{w,b} \frac{1}{\frac{1}{2} w^T w} = \min_{w,b} \frac{1}{2} w^T w$$

∴ goal function: $\begin{cases} \min\limits_{w,b} \frac{1}{2} w^T w \\ s.t.\ y_i(w^T x_i + b) \ge 1 \Leftrightarrow 1 - y_i(w^T x - b) \le 0 \\ \text{for } i = 1 \dots N \end{cases}$

带约束

由拉格朗日来解

$$L(w,b,\lambda) = \frac{1}{2} w^T w + \sum_{i=1}^{N} \underbrace{\lambda_i}_{\ge 0} \big(\underbrace{1 - y_i (w^T x - b)}_{\le 0} \big)$$

不约束 ∴

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

把 $1 - y_i(w^T x - b) \le 0$ 隐含了进来

(拉格朗日的条件)

P.S: Reason for $\min\limits_{w,b} \frac{1}{2} w^T w = \min\limits_{w,b} \max\limits_{\lambda} L(w,b,\lambda)$

if $(1 - y_i(w^T x - b)) > 0$, $\max\limits_{\lambda} L(w,b,\lambda) = \infty$

if $(1 - y_i(w^T x - b)) \le 0$, $\max\limits_{\lambda} L(w,b,\lambda) = \frac{1}{2} w^T w$. $(\lambda_i = 0)$.

∴ $\min\limits_{w,b} \frac{1}{2} w^T w = \min\limits_{w,b} \max\limits_{\lambda} L(w,b,\lambda)$

这条件被丢弃.

无约束问题：$\begin{cases}\min\limits_{w,b}\max\limits_{\lambda} L(w,b,\lambda)\\ s.t.\ \lambda_i \ge 0 \\ \quad\quad i=1\cdots N\end{cases}$

对偶

$\min\max L \ge \max\min L$（弱对偶）

（宁为凤尾，不为鸡头）　"="时 强对偶

凸优化二次问题满足强对偶

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

∴ max min L 与 min max L 同解

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

$$\min_{w,b} L(w,b,\lambda).$$

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

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

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

$$\frac{\partial L}{\partial w}=\frac{1}{2}\times 2w-\sum_{i=1}^{N}\lambda_i y_i x_i \triangleq 0 \Rightarrow w=\sum_{i=1}^{N}\lambda_i y_i x_i$$ 代入 L

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

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

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

已将最优解代入，故此时 $L(w,b,\lambda)$ 为最小值。

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

等价代换

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

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

KKT条件:

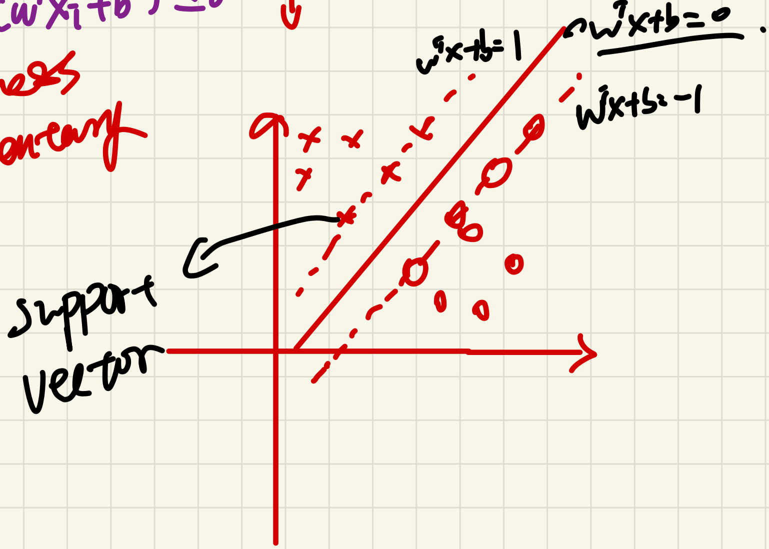
原问题,对偶问题具有强对偶关系

$\Longleftrightarrow$ 满足 KKT条件

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

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

slackness complementary



Soft margin - SVM

model: $\begin{cases} \min\limits_{w,b} \frac{1}{2}w^Tw \\ s.t.\ y_i(w^Tx_i+b) \geq 1 \\ \quad \text{for } i=1,2,\dots,N \end{cases}$

soft = 允许一点点错误

$$\min_{w,b} \frac{1}{2}w^Tw + loss$$

数学性质不好

① $loss = \sum_{i=1}^{N} I\{y_i(w^Tx_i+b)<1\}$

误分类的个数，关于w不连续

故不可导。

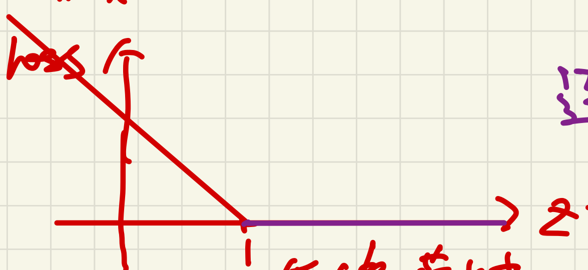
② loss = 距离 → hinge loss

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

如果 $y_i(w^Tx_i+b) < 1$, $loss = 1-y_i(w^Tx_i+b)$

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

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



连续可导。

个像合叶。

∴ soft-margin ~ 允许存在噪声 noise。

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

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

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

弱对偶性的证明:

Primal problem

$$\begin{cases} \min\limits_{x \in \mathbb{R}^p} f(x) \\ s.t. \quad m_i(x) \le 0, \ for \ i=1,\dots,M \\ \qquad n_j(x) = 0, \ for \ j=1,\dots,N \end{cases}$$

拉格朗日函数: $L(x,\lambda,\eta) = f(x) + \sum_{i=1}^{M} \lambda_i m_i + \sum_{j=1}^{N} \eta_j n_j$

无约束

$$\begin{cases} \min\limits_{x} \max\limits_{\lambda,\eta} L(x,\lambda,\eta) \\ s.t. \quad \lambda_i \ge 0 \end{cases}$$

证: $\min\limits_{x} \max\limits_{\lambda,\eta} L(x,\lambda,\eta) \ge \max\limits_{\lambda,\eta} \min\limits_{x} L(x,\lambda,\eta)$

$$\underbrace{\min_{x} L(x,\lambda,\eta)}_{F(\lambda,\eta)} \le L(x,\lambda,\eta) \le \underbrace{\max_{\lambda,\eta} L(x,\lambda,\eta)}_{G(x)}$$

$F(\lambda,\eta) \le G(x)$

$\therefore F(\lambda,\eta) \le \min B(x)$

$\therefore \max\limits_{\lambda,\eta} F(\lambda,\eta) \le \min\limits_{x} B(x)$

$\therefore \max\limits_{\lambda,\eta} \min\limits_{x} L \le \min\limits_{x} \max\limits_{\lambda,\eta} L$ 弱对偶性。

得证 //

对偶性的几何解释

为了方便表示，我们将不等式约束定为一个，并且没有等式约束。

∴ primal problem:

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} f(x) \\ m_1(x) \le 0 \end{cases}, \quad x \in D.$$

D为定义域

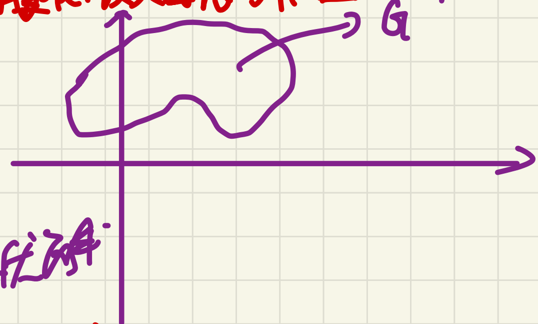
则 Dual problem 为

$$\begin{cases} \max\limits_{\lambda} \min\limits_{x} L(\lambda, x) \\ \lambda \ge 0 \end{cases}$$

设：$G = \{(m_1(x), f(x)) \mid x \in D\}$，其中 $t = m_1(x)$，$u = f(x)$

为了一般化，设其图像分布为：

∴ G



assuming p^* 为 primal p 的最优解

∴ $p^* = \inf\{u \mid t \le 0, (t,u) \in G\}$

→ 下确界

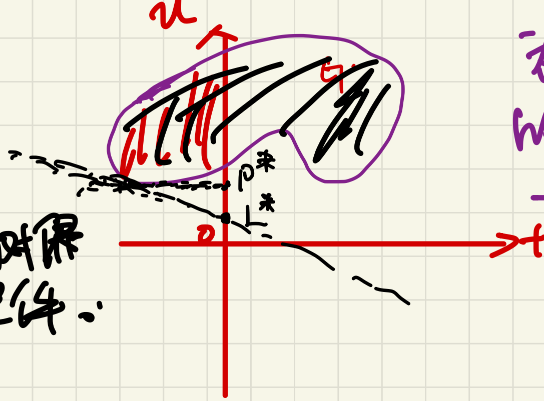
$$\min_x L(\lambda, x) = \min_x f(x) + \lambda m_1(x) = \min_x \underline{\boxed{u + \lambda t}}$$

为斜率为 $-\lambda$ 的直线在u上的值。

∴ $\min\limits_x L(\lambda, x) = \inf\{u + \lambda t \mid (t,u) \in G\} = L^*$

∴ $d^* = \max\limits_{\lambda} L^*$

∴ 可做图：



$L^* \le p^*$

$d^* \le p^*$ → 弱对偶条件。

⇓

如果图形G为凸的，
那么就有 $\alpha^* = p^*$ => 强对偶条件。

---

slater condition:
slater con + 凸 => 强对偶

$\exists \hat{x}$，满足 $m_i(\hat{x}) \le 0$

relint of D（D的内点）。

可以理解为
仿射函数 ≤ 0
一阶线性函数

放松的 slater condition：

如果 N个不等式条件中，存在K个仿射函数。

则只需存在 $\hat{x}$ 满足剩下 N-K个不等式条件即可

即 for $i = N+1 \cdots K$，$m_i(\hat{x}) \le 0$。

由于凸二次规划问题的不等式约束

都为仿射函数

所以 凸二次规划问题一定是满足放松的

slater - condition 的，=> 一定满足强对偶

关系。

KKT条件。

Primal problem:

$$\begin{cases} \min\limits_{x} f(x) \\ \text{s.t.} \quad m_i(x) \le 0,\ i=1,\dots,M \\ \qquad n_j(x) = 0,\ j=1,\dots,N \end{cases}$$

$$L(x,\lambda,\eta) = f(x) + \sum_{i=1}^{M} \lambda_i m_i(x) + \sum_{j=1}^{N} \eta_j n_j(x)$$

$$g(\lambda,\eta) = \min_{x} L(x,\lambda,\eta)$$

Dual prob:

$$\begin{cases} \max\limits_{\lambda,\eta} g(\lambda,\eta) \\ \text{s.t.} \ \lambda_i \ge 0,\ \text{for } i=1,\dots,N \end{cases}$$

$(d^* = p^*)$

convex + slater $\Rightarrow$ strong Duality

$\Updownarrow$

KKT

KKT condition:

可行条件：$\begin{cases} m_i(x^*) \le 0 \\ n_j(x^*) = 0 \\ \lambda^* \ge 0 \end{cases}$

互补松弛：$\sum_{i=1}^{m} \lambda_i^* m_i(x^*) = 0$

梯度为0：$\frac{dL(x,\lambda^*,\eta^*)}{dx} = 0$

强对偶关系

$$d^* = \max_{\lambda,\eta} g(\lambda,\eta) = g(\lambda^*,\eta^*)$$

$$= \min_x L(x,\lambda^*,\eta^*)$$

$$② \le L(x^*,\lambda^*,\eta^*)$$

$$= f(x^*) + \sum_{i=1}^{m} \lambda_i^* m_i + \sum_{j=1}^{N} \underbrace{\eta_j^* n_j}_{0}$$

$$= f(x^*) + \underline{\sum_{i=1}^{m} \lambda_i^* m_i}$$

$\sum_{i=1}^{m} \lambda_i^* m_i$, $m_i(x^*) \le 0$, $\lambda^* \ge 0$

$\therefore m_i(x^*)\lambda_i^* \le 0$

①

$\therefore \le f(x^*) = p^*$

$\because$ 要满足强对偶关系，$\therefore$ 不等式中等号必须成立。

$\therefore$ ① $f(x^*) = f(x^*) + \sum_{i=1}^{m} \lambda_i^* m_i(x^*)$

$\Rightarrow \sum_{i=1}^{m} \lambda_i^* m_i(x^*) = 0$ ～ 互补松弛条件。

② $\because \min_x L(x,\lambda^*,\eta^*) = L(x^*,\lambda^*,\eta^*)$

$\therefore \left.\frac{dL(x,\lambda^*,\eta^*)}{dx}\right|_{x=x^*} = 0$ —— 梯度为0