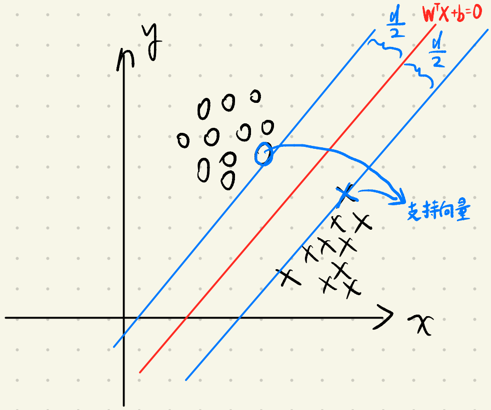


d: margin

支持向量机：最大化 margin 的线，找最好的超平面。

将平行线叉到的向量叫 支持向量

① 训练数据：$(X_1, y_1), \cdots, (X_n, y_n)$，其中 X_i 为向量，y_i 为标签，标签值仅为±1。

② 线性模型 (W, b)：$W^TX+b=0$ 超平面方程（类似线性回归的表达式）

eg. $X_1=\begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1m} \end{bmatrix}$，$W=\begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix}$，b: constant.

③ 线性可分：$\Rightarrow \{(x_i, y_i)\}_{i=1\sim N}$，$\exists (W, b)$，对 $\forall i=1\sim N$，有：
(defination)

$$\begin{cases} y_i=1, & W^TX_i+b\geq 0 \\ y_i=-1, & W^TX_i+b<0 \end{cases}$$

即：$y_i[W^TX_i+b]\geq 0$

优化问题：

最小化 $||W||^2$

限制条件：$y_i[W^TX_i+b]\geq 1$

在进行下一步前：

Case 1. 对 $W^TX+b=0$，若 $a>0$ 则 aW^TX+ab 与 W^TX+b 表示一个超平面。

Case 2. 点到平面的距离公式：

平面：$w_1x+w_2y+b=0$，点 (x_0, y_0)：

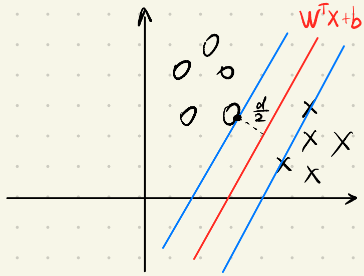
$$d=\frac{|w_1x_0+w_2y_0+b|}{\sqrt{w_1^2+w_2^2}}$$

推广：高维情况下，$W^TX+b=0$，X_0：

$$d=\frac{||W^TX_0+b||}{||W||}$$

① 欲使 margin d 最大:

$$\frac{d}{2}=\frac{|w^Tx+b|}{||w||}$$



为了使推导更加简单,我们使用cacel进行缩放,可以使 $|w'^Tx+b'|=1$,即:

设 $W=aw'$, $b=ab'$

$$\Rightarrow \frac{d}{2}=\frac{|aw'^Tx+ab'|}{||aw'||}=\frac{a|w'^Tx+b'|}{a||w'||}$$

$$=\frac{1}{||w'||}$$

$$\Rightarrow d=\frac{2}{||w'||}$$

即对于支持向量 X,欲求最大的 d,即需要最小化 w',为了方便起见:

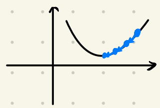
$$d=\frac{2}{||w||}$$

② 优化 —— 二次规化(凸优化)

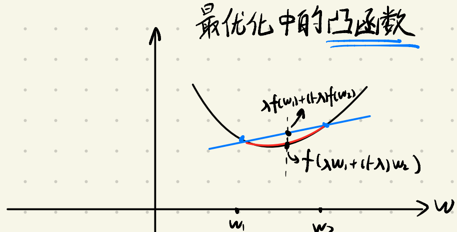
<目标函数>: 二次项 (要么无解,要么一个极值)

<限制条件>: 一次项

即:



最优化中的凸函数



定义: any w_1, w_2, $\forall \lambda\in[0,1]$

$$f(\lambda w_1+(1-\lambda)w_2)\le \lambda f(w_1)+(1-\lambda)f(w_2)$$

原问题: 最小化凸函数 $f(w)$

限制条件: $g_i(w)\le 0$, $i=1\sim k$

$h_i(w)=0$, $i=1\sim M$

对偶问题:

定义: $L(w,\alpha,\beta)=f(w)+\alpha^Tg(w)+\beta^Th(w)$ 其中 $\alpha=\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_k\end{bmatrix}$, $\beta=\begin{bmatrix}\beta_1\\ \vdots\\ \beta_m\end{bmatrix}$

最优化: $\Theta(\alpha,\beta)=\inf_{\text{所有}w}\{L(w,\alpha,\beta)\}$ 求最小值

限制条件: $\alpha_i\ge 0$, $i=1\sim K$ (即 $\alpha\ge 0$)

定理: 若 w^* 是原问题的解,而 α^*, β^* 是对偶问题的解,

那么: $f(w^*)\ge\Theta(\alpha^*,\beta^*)$

$$\Theta(\alpha^*,\beta^*)=\inf_{(\text{所有}w)}\{L(w,\alpha^*,\beta^*)\}$$

$$\le L(w^*,\alpha^*,\beta^*)$$

$$=f(w^*)+\underbrace{\alpha^{*T}g(w^*)}_{\le 0}+\underbrace{\beta^{*T}h(w^*)}_{=0}$$

$$\le f(w^*)$$

定义: $G = f(w^*) - \Theta(\alpha^*, \beta^*)$，G叫原问题与对偶问题的间距

(某些特定优化问题可以证明 $G=0$)

强对偶定理: 若 $f(w)$ 为凸函数, 且 $g(w) = Aw + b$

(暂不证明) $h(w) = Cw + d$

则当对偶间距 $G=0$ 时取得最优条件.

即: $f(w^*) = \Theta(\alpha^*, \beta^*)$

前文中 $f(w^*) = f(w^*) + \sum_{i=1}^{K} \alpha_i^* g_i(w^*)$，那么必然 $\begin{cases} \alpha_i^* = 0 \text{ 或} \\ g_i(w^*) = 0 \end{cases}$ ← KKT条件.

在支持向量机中，

最小化：$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$

限制条件：① $y_i[W^T\varphi(x_i)+b] \geq 1-\xi_i$

② $\xi_i \geq 0$

原问题 $f(w)$ 为凸函数

$g_i(w) \leq 0$

$h_i(w) = 0$

⇒ 改造：对 ξ_i 取相反数。

最小化：$f(w) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$

限制条件：① $1+\xi_i - y_iW^T\varphi(x_i) - y_ib \leq 0$

② $\xi_i \leq 0$

对偶问题：$\Theta(\alpha,\beta) = \inf\limits_{所有(w,\xi_i,b)} \left\{\frac{1}{2}\|w\|^2 - C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\beta_i\xi_i + \sum_{i=1}^{N}\alpha_i[1+\xi_i - y_iw^T\varphi(x_i) - y_ib_i]\right\}$

限制条件：$\begin{cases}\alpha_i \geq 0 \\ \beta_i \geq 0\end{cases}$

找极值：
$$\begin{cases}\frac{\partial L}{\partial w} = w - \sum_{i=1}^{N}\alpha_iy_i\varphi(x_i) = 0 \\ \frac{\partial L}{\partial \xi_i} = -C + \beta_i + \alpha_i = 0 \\ \frac{\partial L}{\partial b} = -\sum_{i=1}^{N}\alpha_iy_i = 0\end{cases}$$

⇐ 注意是对 ξ_i 求偏导，$\sum_{k=1}^{N}\xi_k$ 中，仅 $k=i$ 时为 ξ_i.

⇒ $w = \sum_{i=1}^{N}\alpha_iy_i\varphi(x_i)$ ①

$\alpha_i + \beta_i = C$ ②

$\sum_{i=1}^{N}\alpha_iy_i = 0$ ③

$C = \alpha_i + \beta_i$

回代入 L：

消除

代入③

$L = \frac{1}{2}\|w\|^2 - C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\beta_i\xi_i + \sum_{i=1}^{N}\alpha_i\xi_i + \sum_{i=1}^{N}\alpha_i \boxed{-\sum_{i=1}^{N}\alpha_iy_iW^T\varphi(x_i)} \cancel{-\sum_{i=1}^{N}\alpha_iy_i\cdot b}$

这儿不能代③，因为每组 $W^T\varphi(x_i)$ 不一样，无法提出为 $W^T\varphi(x_i)\sum_{i=1}^{N}\alpha_iy_i$，只能 $W^T\sum_{i=1}^{N}\alpha_iy_i\varphi(x_i)$

由①：$\frac{1}{2}\|w\|^2 = \frac{1}{2}w^Tw = \frac{1}{2}\left(\sum_{i=1}^{N}\alpha_iy_i\varphi(x_i)\right)^T\left(\sum_{j=1}^{N}\alpha_jy_j\varphi(x_j)\right)$

$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jy_iy_j\underbrace{\varphi^T(x_i)\varphi(x_j)}_{=K(x_i,x_j)，是常数}$

对于 $-\sum_{i=1}^{N}\alpha_iy_iw^T\varphi(x_i)$

$= -\sum_{i=1}^{N}\alpha_iy_i\left(\sum_{j=1}^{N}\alpha_jy_j\varphi(x_j)\right)^T\varphi(x_i)$

$= -\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jy_iy_j\varphi(x_j)^T\varphi(x_i)$ → 由于结果是常数 ∴ $\varphi(x_j)^T\varphi(x_i)$ 交换前后顺序不影响.

对照

最终：$L(w,\alpha,\beta) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jy_iy_j\varphi^T(x_i)\varphi(x_j)$

即：$\Theta(\alpha,\beta) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jy_iy_jK(x_i,x_j)$

限制条件：① $0 \leq \alpha_i \leq C$ ⇐ ∵ 上文得到，$\alpha_i + \beta_i = C$，$\alpha_i \geq 0$，$\beta_i \geq 0$.

② $\sum_{i=1}^{N}\alpha_iy_i = 0$ ⇐ 也是上文的。

已知：标签 y_i, y_j，以及核函数表达式 K. 未知：α.

对于 $W^T\varphi(X) = \sum_{i=1}^{N} [\alpha_i y_i \varphi(x_i)]^T \varphi(x)$

$$= \sum_{i=1}^{N} \alpha_i y_i \varphi(x_i)^T \varphi(x) = \sum_{i=1}^{N} \alpha_i y_i K(X_i, X)$$ ，发现w不用求.

应用 $\alpha_i^* = 0$ 或 $g_i^*(w^*) = 0$ (KKT条件)，有：

① 要么 $\beta_i = 0$，要么 $\xi_i = 0$

② 要么 $\alpha_i = 0$，要么 $1 + \xi_i - y_i W^T\varphi(x_i) - y_i b = 0$

取 $0 < \alpha_i < C \Rightarrow \beta_i = C - \alpha_i > 0$

此时，$\beta_i \neq 0 \Rightarrow \xi_i = 0$.

$$\alpha_i \neq 0 \Rightarrow 1 + \underbrace{\xi_i}_{0} - y_i W^T\varphi(x_i) - y_i b = 1 - y_i W^T \varphi(x_i) - y_i b = 0$$

$$\Rightarrow b = \frac{1 - y_i W^T \varphi(x_i)}{y_i} = \frac{1 - y_i \sum_{i=1}^{N} \alpha_i y_i K(x_i, x_j)}{y_i}$$

测试：替换w即可.

$$\Rightarrow \quad \hat{y}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b\right) \geq 0$$