支持向量机 最优间隔分类器

原始优化问题 和 对偶优化问题 (KKT)条件

$$h_{w,b}(x) = g(w^T x + b)$$

$$g(z) = \begin{cases} 1 & z \ge 0 \\ -1 & \text{otherwise} \end{cases}$$

Function margin: $\hat{\gamma}^{(i)} = y^{(i)}(w^T x^{(i)} + b)$

Geometric margin: $\gamma^{(i)} = y^{(i)}\left(\frac{w^T}{\|w\|} x^{(i)} + \frac{b}{\|w\|}\right)$

$\gamma = \min \gamma^{(i)}$

#1

$$\max_{\gamma, w, b} \gamma$$

s.t. $y^{(i)}(w^T x^{(i)} + b) \ge \gamma$

$\|w\| = 1$

难优

非凸优化问题

#2

$$\max_{\hat{\gamma}, w, b} \frac{\hat{\gamma}}{\|w\|}$$

s.t. $y^{(i)}(w^T x^{(i)} + b) \ge \hat{\gamma}$

用现优化软件优化

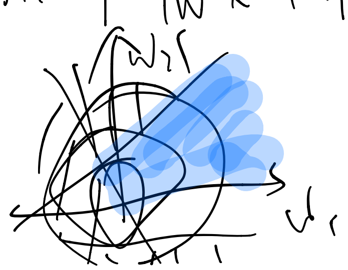
实际就是 找直线线把点分割的情况下 到点最差的距离最大

加约束 $\hat{\gamma} = 1$ $\quad \min_i y^{(i)}(w^T x^{(i)} + b) = 1$

#3

$$\min \|w\|^2$$

s.t. $y^{(i)}(w^T x^{(i)} + b) \ge 1$



去掉多余平衡后，凸优化

↑ 最优间隔分类器

拉格朗日最值，，，高数知识了

$$L(w,\beta)=f(w)+\sum_i \beta_i h_i(w)$$

$$\frac{\partial L}{\partial w}\overset{set}{=}0 \qquad \frac{\partial L}{\partial \beta}=0$$

w^* 存在解，则 $\exists \beta^*$ s.t. $\frac{\partial L(w^*,\beta^*)}{\partial w^*}=0=\frac{\partial L(w^*,\beta^*)}{\partial \beta^*}$ 必要

拉格朗日一般形式

$$\min f(w)$$

s.t. $g_i(w)\le 0$, $h_i(w)=0$ 约束

$$L(w,\alpha,\beta)=f(w)+\sum_{i=1}^{k}\alpha_i g_i(w)+\sum_{i=1}\beta_i h_i(w)$$

定义 $\theta_p(w)=\max_{\alpha,\beta,\ \alpha_i\ge 0} L(w,\alpha,\beta)$

$$p^*=\min_w \max_{\alpha,\beta,\ \alpha_i\ge 0} L(w,\alpha,\beta)=\min \theta_p(w)$$

两个不满足前提 有问题？

so $p^*=f(w)$

对偶问题

$$\theta_D(\alpha,\beta)=\min L(w,\alpha,\beta)$$

$$d^*=\max \nearrow = \max_{\alpha\ge 0,\ \beta}\theta_D(\alpha,\beta)$$

$d^*\le p^*$ —— 有时对偶问题更简单

f 是凸函数

h_i 是仿射函数 $h_i(w)=a_i^T w+b_i$

g_i 严格可执行的 $[\exists w$ 使 $\forall i\ g_i(w)<0]$

在这种情况下 $\exists w^*, \alpha^*, \beta^*$ 使 w^* 是原问题的解

α^*, β^* 是对偶问题解

$g_i(w^*)\le 0$ $\alpha_i^* \ge 0$

且 $p^* = d^* = \mathcal{L}(w^*, \alpha^*, \beta^*)$

原问题解 对偶解

且 $\frac{\partial \mathcal{L}(w^*, \alpha^*, \beta^*)}{\partial w}=0$,

$\frac{\partial \mathcal{L}}{\partial \beta}=0$, $\alpha^* g_i(w)=0$

至少其中一个 $=0$

KKT 互补条件

KKT 蕴含着 if $\alpha_i \neq 0 \Rightarrow g_i(w^*)=0$

$\alpha_i > 0$

g_i 是一个活动约束

返回去 最优间隔分类器的原始问题和对偶问题 拉格朗日乘数

$\alpha_i, \beta_i \rightarrow \alpha_i$

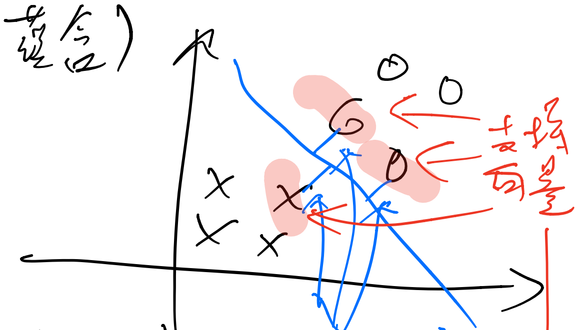
只需一组

$w \rightarrow w, b$

$\min \frac{1}{2}\|w\|^2$ $\quad y^{(i)}(w^T x^{(i)}+b) \geq 1$ $\quad g_i(w,b)=1-y^{(i)}(w^T x^{(i)}+b) \le 0$

$\alpha_i \rightarrow g_i(w,b)=0$ （KKT 条件）

$\Longleftrightarrow$ 函数间隔 $=1$



$$\mathcal{L}(w,b,\alpha)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{m}\alpha_i\left(y^{(i)}\left(w^Tx^{(i)}+b\right)-1\right)$$

对偶问题

$$\theta_D(\alpha)=\min_{w,b}\mathcal{L}(w,b,\alpha)$$

$\downarrow$ 导 $=0$ $\quad \nabla_w \mathcal{L}=w-\sum_{i=1}^{m}\alpha_i y^{(i)}x^{(i)} \stackrel{set}{=} 0$

约束 1: $\quad \Rightarrow w=\sum_{i=1}^{m}\alpha_i y^{(i)}x^{(i)}$

约束 2 $\quad \frac{\partial \mathcal{L}}{\partial b}=-\sum_{i=1}^{m}y^{(i)}\alpha^{(i)} \stackrel{set}{=} 0$

$$\mathcal{L}=\frac{1}{2}w^Tw-\sum_{i=1}^{m}\alpha_i\left(y^{(i)}\left(w^Tx^{(i)}+b\right)-1\right)$$

$$=-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}y^{(i)}y^{(j)}\alpha_i\alpha_j\underbrace{\langle x^{(i)},x^{(j)}\rangle}_{(x^{(i)})^T x^{(j)}}+\sum_{i=1}^{n}\alpha_i$$

$=W(\alpha)$

对偶问题

$\max W(\alpha) \quad \alpha_i \geq 0 \quad \sum_i y_i\alpha_i=0 \quad \leftarrow \frac{\partial \mathcal{L}}{\partial b}=0$

if $\left(\sum_i y_i\alpha_i \neq 0\right)$

$\theta_D(\alpha)=-\infty$

$\max \uparrow$

如果求出 α^* 则求出 w 再 → b 目标 max

$$w = \sum_{i}^{m} \alpha_i y^{(i)} x^{(i)}$$

$$b = \frac{\max_{i:y^{(i)}=-1} w^T x^{(i)} + \min_{i:y^{(i)}=1} w^T x^{(i)}}{2}$$

$$h_{w,b}(x) = g(w^T x + b)$$

$$w^T x + b = \sum_{i=1}^{m} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle + b$$

只用算非0

kernels 高维 $x^{(i)} \in \mathbb{R}^\infty$ → 高效计算的结构的
无限维的向量
（下一讲）

这一章：无约束极小 → 导 $=0$ $\nabla = 0$

↓

$=0$ 约束极值 → 拉格朗日 $\mathcal{L}(\ ,\alpha)$

↓

≤ 0
$=0$ 极值 → KKT

↓

+ 约 ? → 支持向量机

+ 对偶

（目前都假设样本线性可分割）

对偶问题的定义