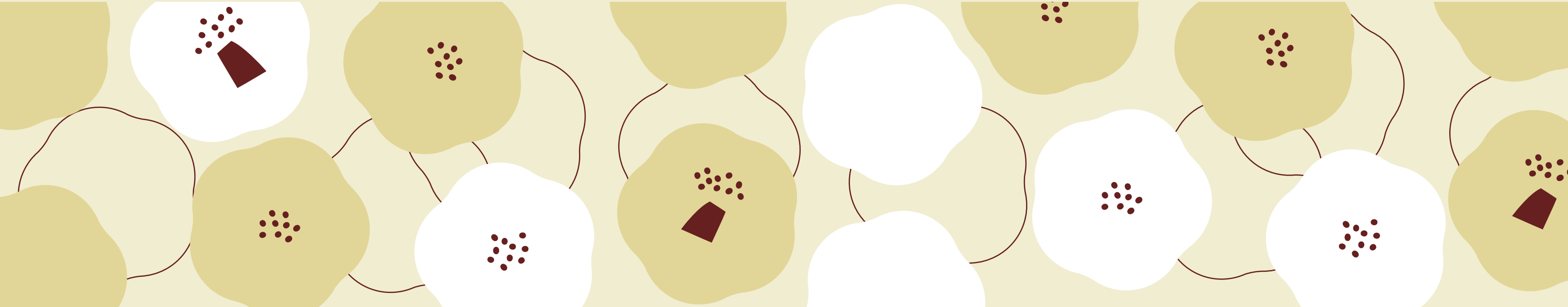

约束优化问题

$$\begin{cases} \min\limits_{x\in\mathbb{R}} f(x) \quad \text{原问题} \\ \\ s.t. \quad m_i(x)\le 0 \quad i=1,2,\cdots,M \\ \qquad\ \ n_j(x)\ge 0 \quad j=1,2,\cdots,N \end{cases} \qquad ①$$

将原问题(带约束)通过引入拉格朗日乘子转为无约束的拉格朗日函数

$$L(x,\lambda,\eta)=f(x)+\sum_{i=1}^{M}\lambda_i m_i(x)+\sum_{j=1}^{N}\eta_j n_j(x) \qquad ②$$

无约束优化问题(与原问题等价)

$$\begin{cases} \min\limits_{x} \max\limits_{\lambda,\eta} L(x,\lambda,\eta) \\ \\ s.t. \quad \lambda_i \ge 0 \end{cases} \qquad ③$$

→ 只包含满足约束的x集合

下面证明，求解③与①等价

证：假设有x_i违反了问题①中的不等式约束

即 $m_i(x_i)>0$

则有 $\max\limits_{\lambda} \lambda_i m_i(x_i) \to +\infty$

若x_i满足不等式约束，即 $m_i(x_i)\le 0$

则有 $\max\limits_{\lambda} \lambda_i m_i(x_i) \ne +\infty$

则问题③等价于：

$$\min_{x}\max_{\lambda} L = \min_{x}\{\underbrace{\max L}_{\text{满足约束的}x}, \underbrace{+\infty}_{\text{不满足约束的}x}\} = \min_{x\text{满足不等式约束}} \max L$$

即使用拉格朗日乘子法相当于对x的取值进行过滤，只取满足约束的x集合

对偶性

写出问题③的对偶形式

$$\begin{cases}\min\limits_{x}\max\limits_{\lambda,\eta} L(x,\lambda,\eta) \\ s.t.\ \lambda_i \geq 0\end{cases} \xleftrightarrow{\text{对偶问题}} \begin{cases}\max\limits_{\lambda,\eta}\min\limits_{x} L(x,\lambda,\eta) \\ s.t.\ \lambda_i \geq 0\end{cases}$$

原问题 p 　　　　　　对偶问题 d

弱对偶性：　$d \leq p$，即

要证：$\max\limits_{\lambda,\eta}\min\limits_{x} L \leq \min\limits_{x}\max\limits_{\lambda,\eta} L$

证：$\underbrace{\min\limits_{x} L(x,\lambda,\eta)}_{\substack{x\text{方向的极小值} \\ \text{记 } A(\lambda,\eta)}} \leq L(x,\lambda,\eta) \leq \underbrace{\max\limits_{\lambda,\eta} L(x,\lambda,\eta)}_{\substack{\lambda,\eta\text{上的极大值} \\ \text{记 } B(x)}}$

即　$A(\lambda,\eta) \leq B(x)$　函数A恒小于等于B

$\Rightarrow \max\limits_{\lambda,\eta} A \leq \min\limits_{x} B$

故　$\max\limits_{\lambda,\eta}\min\limits_{x} L(x,\lambda,\eta) \leq \min\limits_{x}\max\limits_{\lambda,\eta} L(x,\lambda,\eta)$

证毕.

强对偶关系

对偶问题的解 $d^* =$ 原问题的解 p^*

即　$\max\limits_{\lambda,\eta}\min\limits_{x} L(x,\lambda,\eta) = \min\limits_{x}\max\limits_{\lambda,\eta} L(x,\lambda,\eta)$

凸优化 + Slater条件 $\Rightarrow$ 满足强对偶关系

记结论：SVM的优化问题满足强对偶关系.

原问题：

$$\begin{cases} \min\limits_{x \in R} \ f(x) \\ s.t \quad m_i(x) \geq 0 \\ \qquad n_j(x) \leq 0 \end{cases}$$

拉格朗日对偶问题

$$g(\lambda, \eta) = \min_x L(x, \lambda, \eta)$$

$$\begin{cases} \max\limits_{\lambda, \eta} \ g(\lambda, \eta) \\ s.t. \quad \lambda_i \geq 0 \end{cases}$$

充要条件

已知问题③满足强对偶关系 $\Longleftrightarrow$ 满足KKT条件

KKT条件

①可行域条件

$$m_i(x^*) \leq 0$$
$$n_j(x^*) = 0$$
$$\lambda^* \geq 0$$

②互补松弛条件

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

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

$$\leq f(x^*) = p^*$$

由于原问题满足强对偶关系，有 $d^* = p^*$

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

$\Rightarrow \lambda_i m_i(x^*) = 0, \forall i = 1, 2, \cdots, M$　（$\lambda_i \geq 0$, $m_i(x^*) \leq 0$, 所以每项都必须为0）

③梯度为0

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

$$\Rightarrow \left. \frac{\partial L(x, \lambda^*, \eta^*)}{\partial x} \right|_{x = x^*} = 0$$

SVM：最大硬间隔分类器

函数间隔 $\hat{\gamma}_i = y_i(wx_i+b)$　　$x_i \in R$，$y_i \in \{-1,+1\}$

$\hat{\gamma} = \min \hat{\gamma}_i$

几何间隔 $\gamma_i = \dfrac{y_i(wx_i+b)}{\|w\|}$

$\gamma = \min \gamma_i$

SVM优化问题：最大化训练集的点到超平面的几何间隔

$$\begin{cases} \max\limits_{w,b} \gamma \\ s.t. \ \dfrac{y_i(wx_i+b)}{\|w\|} \ge \gamma \end{cases} \quad ①$$

几何间隔与函数间隔关系

$$\Rightarrow \begin{cases} \max\limits_{w,b} \dfrac{\hat{\gamma}}{\|w\|} \\ s.t. \ y_i(wx_i+b) \ge \hat{\gamma} \end{cases} \quad ②$$

$\hat{\gamma}$的取值不影响优化问题，可以令 $\hat{\gamma}=1$

问题②转化为

$$\begin{cases} \min\limits_{w} \frac{1}{2}\|w\|^2 \\ s.t. \ 1-y_i(wx_i+b) \le 0, \quad i=1,2,\cdots,N \end{cases} \quad ③$$

问题③是凸二次规化问题，利用拉格朗日和强对偶 KKT 进行求解

支持向量：$y_i(wx_i+b)=1$

只有 support vector 决定分类超平面

③的拉格朗日函数

$$L(w,b,\alpha) = \frac{1}{2}w^Tw + \sum_{i=1}^{N}\alpha_i(1-y_i(wx_i+b))$$

③的无约束问题

$$\begin{cases} \min\limits_{w,b} \max\limits_{\alpha} L(w,b,\alpha) \\ s.t. \ \alpha_i \ge 0 \end{cases} \quad ④$$

本质上丢弃了 $1-y_i(wx_i+b)>0$ 的情况

证明可见上两页的内容

利用强对偶可知，问题④等价于其对偶问题

$$\begin{cases} \min\limits_{w,b} \max\limits_{\alpha} L(w,b,\alpha) & ④ \\ s.t. \quad \alpha_i \geq 0 \end{cases}$$

对偶 $\Leftrightarrow$

$$\begin{cases} \max\limits_{\alpha} \min\limits_{w,b} L(w,b,\alpha) & ⑤ \\ s.t. \quad \alpha_i \geq 0 \end{cases}$$

对⑤进行求偏导求解有：

$$\begin{cases} \min\limits_{\alpha} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i & ⑥ \\ s.t. \quad \sum_{i=1}^{N}\alpha_i y_i = 0 \\ \alpha_i \geq 0 \end{cases}$$

对⑥使用KKT条件，得． α^*是⑥的解，b^*是使用某个支持向量求解得到

$$\begin{cases} w^* = \sum_{i=1}^{N}\alpha_i^* y_i x_i & ⑦ \\ b^* = y_j - \sum_{i=1}^{N}\alpha_i^* y_i (x_i \cdot x_j) \end{cases}$$

$\alpha_i^* > 0$ 的是支持向量

对 w^*，b^* 有影响

软间隔 SVM

$$\begin{cases} \min\limits_{w,b,\xi} \frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i & ⑧ \\ s.t. \quad y_i(wx_i+b) \geq 1-\xi_i \\ \xi_i \geq 0 \end{cases}$$

对每个松弛变量支付代价

函数间隔放宽松

问题⑧的拉格朗日函数：

$$L(w,b,\xi,\alpha,\mu) = \frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\alpha_i\left[1-\xi_i - y_i(wx_i+b)\right] - \sum_{i=1}^{N}\mu_i\xi_i$$

对偶问题形式

$$\begin{cases} \min\limits_{\alpha} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i & ⑨ \\ s.t. \quad \sum_{i=1}^{N}\alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C, \quad i=1,2,\cdots,N \end{cases}$$

对⑨使用KKT求解得到与⑦同样形式的解

出了实例 x_i 到间隔边界的距离 $\frac{\xi_i}{\|w\|}$。

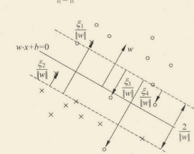

图 7.5 软间隔的支持向量

软间隔的支持向量 x_i 或者在间隔边界上，或者在间隔边界与分离超平面之间，或者在分离超平面误分一侧。若 $\alpha_i^* < C$，则 $\xi_i = 0$，支持向量 x_i 恰好落在间隔边界上；若 $\alpha_i^* = C$，$0 < \xi_i < 1$，则分类正确，x_i 在间隔边界与分离超平面之间；若 $\alpha_i^* = C$，$\xi_i = 1$，则 x_i 在分离超平面上；若 $\alpha_i^* = C$，$\xi_i > 1$，则 x_i 位于分离超平面误分一侧。

153 / 488

从合页损失理解软间隔SVM

问题⑨等价于优化:

$$\min_{w,b} \sum_{i=1}^{N} [1-y_i(wx_i+b)]_+ + \lambda \|w\|_2^2 \quad ⑩$$

$$[z]_+ = \begin{cases} z, & z>0 \\ 0, & z\le 0 \end{cases}$$

SMO算法

求解如下凸二次规划的对偶问题:

$$\begin{cases} \min_{\alpha} \dfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j K(x_i,x_j) - \sum_{i=1}^{N}\alpha_i \\ \text{s.t.} \quad \sum_{i=1}^{N}\alpha_i y_i = 0 \\ \quad 0 \le \alpha_i \le C,\ i=1,2,\cdots,N \end{cases} \quad ⑪$$

SMO算法的思想是每次只优化两个变量α_1, α_2,其余当做常量

则⑪变成

$$\begin{cases} \min_{\alpha_1,\alpha_2} \dfrac{1}{2}K_{11}\alpha_1^2 + \dfrac{1}{2}K_{22}\alpha_2^2 + y_1y_2K_{12}\alpha_1\alpha_2 - (\alpha_1+\alpha_2) \\ \qquad + y_1\alpha_1\sum_{i=3}^{m} y_i\alpha_i K_{i1} + y_2\alpha_2\sum_{i=3}^{N} y_i\alpha_i K_{i2} \\ \text{s.t.} \quad \alpha_1y_1+\alpha_2y_2 = -\sum_{i=3}^{N}\alpha_i y_i = k \\ \quad 0 \le \alpha_i \le C,\ i=1,2,\cdots,N \end{cases} \quad ⑫$$

由⑫中的约束 $\begin{cases} \alpha_1y_1+\alpha_2y_2 = k, \text{ 且 } y_1, y_2 \in \{-1,+1\} \\ 0 \le \alpha_i \le C \end{cases}$

$$\alpha_1^{new}y_1 + \alpha_2^{new}y_2 = \alpha_1^{old}y_1 + \alpha_2^{old}y_2$$

当$y_1=y_2$时,有 $\alpha_1+\alpha_2=k$ (k指代常数,不区分符号)

有: $\alpha_1^{old}+\alpha_2^{old} = \alpha_1^{new}+\alpha_2^{new}$

$\Rightarrow \alpha_2^{new} = \alpha_1^{old}+\alpha_2^{old}-\alpha_1^{new}$

又∵ $0 \le \alpha_1^{new} \le C$

∴ $\alpha_1^{old}+\alpha_2^{old}-C \le \alpha_2^{new} \le \alpha_1^{old}+\alpha_2^{old}$

∵ $0 \le \alpha_2^{new} \le C$

$\therefore \max(0, \alpha_1^{old}+\alpha_2^{old}-C) \le \alpha_2^{new} \le \min(C, \alpha_1^{old}+\alpha_2^{old})$

同理，当 $y_1 \neq y_2$ 时，$\alpha_1 - \alpha_2 = k$

有 $\alpha_1^{old} - \alpha_2^{old} = \alpha_1^{new} - \alpha_2^{new}$

$\Rightarrow \alpha_2^{new} = \alpha_1^{new} - (\alpha_1^{old} - \alpha_2^{old})$

$\because \quad 0 \le \alpha_1^{new} \le C$

$\therefore \quad -(\alpha_1^{old} - \alpha_2^{old}) \le \alpha_2^{new} \le C - (\alpha_1^{old} - \alpha_2^{old})$

又$\because \quad 0 \le \alpha_2^{new} \le C$

$\therefore \max(0, \alpha_2^{old} - \alpha_1^{old}) \le \alpha_2^{new} \le \min(C, C+\alpha_2^{old} - \alpha_1^{old})$