人工智能与自动化学院

模 式 识 别

第八讲 对偶支撑向量机与核支撑向量机(*Dual SVM & Kernel SVM*)

8.1 对偶支撑向量机动机(*Motivation of Dual SVM*)

8.2 对偶支撑向量机的拉格朗日分析(*Lagrange Dual SVM*)

8.3 求解对偶支撑向量机最佳值(*Solving Dual SVM*)

8.4 对偶支撑向量机讨论(*Messages behind Dual SVM*)

8.5 核函数支撑向量机(*Kernel SVM*)

8.1 对偶支撑向量机动机

线性支撑向量机模型

最佳的$(\boldsymbol{w}, b)$ =?

$$\min_{\boldsymbol{w}} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$$

$$Subject\ to \quad y_n(\boldsymbol{w}^T\mathbf{x}_n + b) \geq 1$$

$$for\ n = 1,2, \ldots N$$

① $\mathbf{Q} = \begin{bmatrix} 0 & \mathbf{0}_d^T \\ \mathbf{0}_d & \boldsymbol{I}_d \end{bmatrix}$, $\mathbf{p} = \mathbf{0}_{d+1}$,

$\mathbf{a}_n^T = y_n[1 \quad \mathbf{x}_n^T]$, $c_n = 1$,

② $\begin{bmatrix} b \\ \boldsymbol{w} \end{bmatrix} \leftarrow \mathrm{QP}(\mathbf{Q}, \mathbf{p}, \mathbf{A}, \mathbf{c})$

③ 返回最终的b和$\boldsymbol{w}$作为学到的g_{SVM}

8.1 对偶支撑向量机动机

非线性支撑向量机模型

最佳的$(\boldsymbol{w}, b)$ =?

$$\min_{\boldsymbol{w}} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$$

$$Subject\ to \quad y_n(\boldsymbol{w}^T\underbrace{\mathbf{z}_n}_{\boldsymbol{\phi}(\mathbf{x}_n)} + b) \geq 1$$

$$for\ n = 1,2, \ldots N$$

① $\mathbf{Q} = \begin{bmatrix} 0 & \mathbf{0}_{\tilde{d}}^T \\ \mathbf{0}_{\tilde{d}} & \boldsymbol{I}_{\tilde{d}} \end{bmatrix}$, $\mathbf{p} = \mathbf{0}_{\tilde{d}+1}$,

$\mathbf{a}_n^T = y_n[1 \quad \mathbf{x}_n^T]$, $c_n = 1$,

② $\begin{bmatrix} b \\ \boldsymbol{w} \end{bmatrix} \leftarrow \mathrm{QP}(\mathbf{Q}, \mathbf{p}, \mathbf{A}, \mathbf{c})$

③ 返回最终的b和$\boldsymbol{w}$作为学到的g_{SVM}

如果 $\tilde{d}$ 很大，甚至无穷大，挑战巨大

QP针对$(\tilde{d} + 1)$ 个变量和N 个约束条件求解

目的：SVM算法求解可以不依赖于$\tilde{d}$ 吗?

8.1 对偶支撑向量机动机

将有约束条件下的寻优转变为无约束条件下的寻优问题

$$\min_{\mathbf{w}} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$\textit{Subject to} \quad y_n(\mathbf{w}^T\mathbf{z}_n + b) \geq 1$$

$$for\ n = 1,2, \ldots N$$

用Lagrange乘子 α_n 构造Lagrange函数

$$\mathcal{L}(b,\mathbf{w},\boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n(1 - y_n(\mathbf{w}^T\mathbf{z}_n + b))$$

约束项隐含在*max*中

$$\text{SVM} \equiv \min_{b,\mathbf{w}}\Big(\max_{all\ \alpha_n \geq 0} \mathcal{L}(b,\mathbf{w},\boldsymbol{\alpha})\Big) = \min_{b,\mathbf{w}}\Big(\infty\ if\ violating\ ,\ \frac{1}{2}\mathbf{w}^T\mathbf{w}\ if\ feasible\ \Big) = \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

- 任何 “*violating*” $(b,\mathbf{w})$: $\max_{all\ \alpha_n \geq 0}\left(\blacksquare + \sum_n \alpha_n(\text{一些正数})\right) \rightarrow \infty$
- 任何 “*feasible*” $(b,\mathbf{w})$: $\max_{all\ \alpha_n \geq 0}\left(\blacksquare + \sum_n \alpha_n(\text{所有非正数})\right) = \blacksquare$

第八讲 对偶支撑向量机与核支撑向量机(*Dual SVM & Kernel SVM*)

8.2 对偶支撑向量机的拉格朗日分析

在所有的$\alpha_n \geq 0$的中挑选任意一个α' （$\because$ $max \geq any$）

$$\min_{b,\mathbf{w}}\left(\max_{all\ \alpha_n\geq 0} \mathcal{L}(b,\mathbf{w},\boldsymbol{\alpha})\right) \geq \min_{b,\mathbf{w}}(\mathcal{L}(b,\mathbf{w},\alpha'))$$

如果$\alpha' \geq 0$是上式右边$\max\limits_{all\ \alpha_n\geq 0} \mathcal{L}(b,\mathbf{w},\boldsymbol{\alpha})$中的最佳值 （$\because$ $best\ is\ one\ of\ any$）

$$\min_{b,\mathbf{w}}\left(\max_{all\ \alpha_n\geq 0} \mathcal{L}(b,\mathbf{w},\boldsymbol{\alpha})\right) \geq \underbrace{\max_{all\ \alpha_n'\geq 0}\left(\min_{b,\mathbf{w}}(\mathcal{L}(b,\mathbf{w},\alpha'))\right)}_{\text{Lagrange Dual Problem}}$$

原问题(求解$b,\mathbf{w}$)与拉格朗日对偶问题(求解α)的关系

8.2 对偶支撑向量机的拉格朗日分析

二次规划(QP)满足强对偶特性

$$\underbrace{\min_{b,\mathbf{w}}\left(\max_{all\ \alpha_n\geq 0} \mathcal{L}(b,\mathbf{w},\boldsymbol{\alpha})\right)}_{\text{equiv.to original SVM Primal Problem}} \geq \underbrace{\max_{all\ \alpha_n\geq 0}\left(\min_{b,\mathbf{w}}(\mathcal{L}(b,\mathbf{w},\boldsymbol{\alpha}))\right)}_{\text{Lagrange Dual Problem}}$$

- “≥”是一种弱对偶关系(weak duality)
- “=”是一种强对偶关系(strong duality)，如果满足：

二次规划(QP)问题

- 原问题是凸函数
- 原问题存在可行解
- 约束条件为线性的

非线性变换$\boldsymbol{\phi}(\mathbf{x}_n)$

8.2 对偶支撑向量机的拉格朗日分析

二次规划(QP)满足强对偶特性

$$\underbrace{\min_{b,\boldsymbol{w}}\left(\max_{all\ \alpha_n\geq 0}\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\alpha})\right)}_{\text{equiv.to original SVM Primal Problem}} = \underbrace{\max_{all\ \alpha_n\geq 0}\left(\min_{b,\boldsymbol{w}}(\mathcal{L}(b,\boldsymbol{w},\alpha))\right)}_{\text{Lagrange Dual Problem}}$$

等式两边对原问题求解和对对偶问题求解都能得到最优$(b,\boldsymbol{w},\alpha)$

- "≥"是一种弱对偶关系(weak duality)
- "="是一种强对偶关系(strong duality)，如果满足：

二次规划(QP)问题 ⇐
- 原问题是凸函数
- 原问题存在可行解
- 约束条件为线性的

非线性变换$\boldsymbol{\phi}(\mathbf{x}_n)$

8.2 对偶支撑向量机的拉格朗日分析

对偶问题求解

$$\max_{all\ \alpha_n\geq 0}\left(\underbrace{\min_{b,\boldsymbol{w}}(\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{n=1}^{N}\alpha_n(1-y_n(\boldsymbol{w}^T\boldsymbol{z}_n+b)))}_{\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\alpha})}\right)$$

- "括号"内的问题(*inner problem*)是无约束条件的优化问题

$$\frac{\partial\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\alpha})}{\partial b}=-\sum_{n=1}^{N}\alpha_n y_n=0$$

$$\max_{all\ \alpha_n\geq 0,\sum_{n=1}^{N}\alpha_n y_n=0}\left(\min_{b,\boldsymbol{w}}(\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{n=1}^{N}\alpha_n(1-y_n(\boldsymbol{w}^T\boldsymbol{z}_n))-\sum_{n=1}^{N}\alpha_n y_n b)\right)$$

8. 2 对偶支撑向量机的拉格朗日分析

对偶问题求解

$$\max_{all\ \alpha_n \geq 0}\left(\underbrace{\min_{b,\boldsymbol{w}}(\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \sum_{n=1}^{N}\alpha_n(1 - y_n(\boldsymbol{w}^T\boldsymbol{z}_n + b)))}_{\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\alpha})}\right)$$

- "括号"内的问题(*inner problem*)是无约束条件的优化问题

$$\frac{\partial\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\alpha})}{\partial b} = -\sum_{n=1}^{N}\alpha_n y_n = 0$$

"括号"内的问题b取最佳解时：

$$\max_{all\ \alpha_n \geq 0, \sum_{n=1}^{N}\alpha_n y_n = 0}\left(\min_{b,\boldsymbol{w}}(\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \sum_{n=1}^{N}\alpha_n(1 - y_n(\boldsymbol{w}^T\boldsymbol{z}_n))\right)$$

8. 2 对偶支撑向量机的拉格朗日分析

对偶问题求解

$$\max_{all\ \alpha_n \geq 0, \sum_{n=1}^{N}\alpha_n y_n = 0}\left(\min_{b,\boldsymbol{w}}(\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \sum_{n=1}^{N}\alpha_n(1 - y_n(\boldsymbol{w}^T\boldsymbol{z}_n))\right)$$

- "括号"内的问题(*inner problem*)是无约束条件的优化问题

$$\frac{\partial\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\alpha})}{\partial w_i} = w_i - \sum_{n=1}^{N}\alpha_n y_n z_{n,i} = 0 \Rightarrow \boldsymbol{w} = \sum_{n=1}^{N}\alpha_n y_n \boldsymbol{z}_n$$

"括号"内的问题$\boldsymbol{w}$取最佳解时：

$$\max_{all\ \alpha_n \geq 0,\ \sum_{n=1}^{N}\alpha_n y_n = 0,\ \boldsymbol{w} = \sum_{n=1}^{N}\alpha_n y_n \boldsymbol{z}_n}\left(\min_{b,\boldsymbol{w}}(\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \sum_{n=1}^{N}\alpha_n - \boldsymbol{w}^T\boldsymbol{w})\right)$$

8. 2 对偶支撑向量机的拉格朗日分析

对偶问题求解

$$\max_{all\ \alpha_n\geq 0, \sum_{n=1}^{N}\alpha_n y_n=0}\left(\min_{b,w}(\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{n=1}^{N}\alpha_n(1-y_n(\boldsymbol{w}^T\boldsymbol{z}_n))\right)$$

- “括号”内的问题(*inner problem*)是无约束条件的优化问题

$$\frac{\partial\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\alpha})}{\partial b w_i}=w_i-\sum_{n=1}^{N}\alpha_n y_n z_{n,i}=0 \Rightarrow \boldsymbol{w}=\sum_{n=1}^{N}\alpha_n y_n \boldsymbol{z}_n$$

“括号”内的问题$\boldsymbol{w}$取最佳解时：

$$\max_{all\ \alpha_n\geq 0,\ \sum_{n=1}^{N}\alpha_n y_n=0,\ \boldsymbol{w}=\sum_{n=1}^{N}\alpha_n y_n \boldsymbol{z}_n}\left(\min_{b,w}(-\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{n=1}^{N}\alpha_n)\right)$$

人工智能与自动化学院 13

8. 2 对偶支撑向量机的拉格朗日分析

对偶问题求解

$$\max_{all\ \alpha_n\geq 0, \sum_{n=1}^{N}\alpha_n y_n=0}\left(\min_{b,w}(\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{n=1}^{N}\alpha_n(1-y_n(\boldsymbol{w}^T\boldsymbol{z}_n))\right)$$

- “括号”内的问题(*inner problem*)是无约束条件的优化问题

$$\frac{\partial\mathcal{L}(b,\boldsymbol{w},\boldsymbol{\alpha})}{\partial b w_i}=w_i-\sum_{n=1}^{N}\alpha_n y_n z_{n,i}=0 \Rightarrow \boldsymbol{w}=\sum_{n=1}^{N}\alpha_n y_n \boldsymbol{z}_n$$

“括号”内的问题$\boldsymbol{w}$取最佳解时：

$$\max_{all\ \alpha_n\geq 0,\ \sum_{n=1}^{N}\alpha_n y_n=0,\ \boldsymbol{w}=\sum_{n=1}^{N}\alpha_n y_n \boldsymbol{z}_n}\left(-\frac{1}{2}\left\|\sum_{n=1}^{N}\alpha_n y_n \boldsymbol{z}_n\right\|^2+\sum_{n=1}^{N}\alpha_n\right)$$

人工智能与自动化学院 14

8.2 对偶支撑向量机的拉格朗日分析

KKT条件

$$\max_{\text{all } \alpha_n \ge 0,\ \sum_{n=1}^{N} \alpha_n y_n = 0,\ \mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{z}_n} \left(-\frac{1}{2}\left\| \sum_{n=1}^{N} \alpha_n y_n \mathbf{z}_n \right\|^2 + \sum_{n=1}^{N} \alpha_n \right)$$

如果$(b, \mathbf{w}, \boldsymbol{\alpha})$是原问题-对偶问题的最佳解(*primal-dual* optimal)：

- 原问题可行解(*primal feasible*)： $y_n(\mathbf{w}^T\mathbf{z}_n + b) \ge 1$
- 对偶问题可行解(*dual feasible*)： $\alpha_n \ge 0$
- 对偶“括号”内优化解(*dual-inner* optimal)：$\sum_{n=1}^{N} \alpha_n y_n = 0, \mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{z}_n$
- 原问题“括号”内优化解(*primal-inner* optimal)： $\alpha_n(1 - y_n(\mathbf{w}^T\mathbf{z}_n + b)) = 0$

----称为KKT条件，为优化的充要条件

利用对偶问题求解最佳$\boldsymbol{\alpha}$后，
再用KKT条件得到$(b, \mathbf{w})$

第八讲 对偶支撑向量机与核支撑向量机(*Dual SVM & Kernel SVM*)

8.3 求解对偶支撑向量机最佳值

支撑向量机的对偶问题求解

$$\max_{\text{all } \alpha_n \ge 0,\ \sum_{n=1}^{N}\alpha_n y_n = 0,\ \mathbf{w}=\sum_{n=1}^{N}\alpha_n y_n \mathbf{z}_n} \left(-\frac{1}{2}\left\|\sum_{n=1}^{N}\alpha_n y_n \mathbf{z}_n\right\|^2 + \sum_{n=1}^{N}\alpha_n\right)$$

标准的硬间隔SVM的对偶问题----求解最佳$\boldsymbol{\alpha}$：

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \mathbf{z}_n^T\mathbf{z}_m - \sum_{n=1}^{N}\alpha_n$$

$$\textit{Subject to} \quad \sum_{n=1}^{N}\alpha_n y_n = 0$$

$$\alpha_n \ge 0, \textit{for } n = 1,2,\ldots,N$$

- $\boldsymbol{\alpha}$的目标函数是二次函数、凸函数！N个变量
- $\boldsymbol{\alpha}$的约束条件是线性函数！N+1个约束条件

----二次规划(QP)问题！

QP有成熟方便的办法求优化解！

8.3 求解对偶支撑向量机最佳值

SVM对偶问题的一般求解：

最佳的$\boldsymbol{\alpha}$ =?

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \mathbf{z}_n^T\mathbf{z}_m - \sum_{n=1}^{N}\alpha_n$$

$$\textit{Subject to} \quad \sum_{n=1}^{N}\alpha_n y_n = 0$$

$$\alpha_n \ge 0, \textit{for } n = 1,2,\ldots,N$$

$$\mathbf{u} = [\alpha_1 \ \cdots \ \alpha_n \ \cdots \ \alpha_N]^T \qquad \mathbf{p} = -\mathbf{1}_N$$

$$\mathbf{Q} = \begin{bmatrix} y_1 y_1 \mathbf{z}_1^T\mathbf{z}_1 & \cdots & y_1 y_N \mathbf{z}_1^T\mathbf{z}_N \\ \vdots & y_n y_m \mathbf{z}_n^T\mathbf{z}_m & \vdots \\ y_N y_1 \mathbf{z}_N^T\mathbf{z}_1 & \cdots & y_N y_N \mathbf{z}_N^T\mathbf{z}_N \end{bmatrix}$$

二次规划(QP)的求解：

最佳的$\mathbf{u} \leftarrow \mathrm{QP}(\mathbf{Q}, \mathbf{p}, \mathbf{A}, \mathbf{c})$

$$\min_{\mathbf{u}} \quad \frac{1}{2}\mathbf{u}^T\mathbf{Q}\mathbf{u} + \mathbf{p}^T\mathbf{u}$$

$$\textit{Subject to} \quad \mathbf{a}_m^T\mathbf{u} \ge c_m$$

$$\textit{for } m = 1,2,\ldots M$$

$$\mathbf{a}_{\ge}^T = [y_1 \ \cdots \ y_n \ \cdots \ y_N], \qquad c_{\ge} = 0$$

$$\mathbf{a}_{\le}^T = [-y_1 \ \cdots \ -y_n \ \cdots \ -y_N], \qquad c_{\le} = 0$$

$$\mathbf{a}_1^T = [1 \ \cdots \ 0 \ \cdots \ 0],$$

$$\mathbf{a}_n^T = [0 \ \cdots \ 1 \ \cdots \ 0], \quad c_n = 0, \quad M = N$$

$$\mathbf{a}_N^T = [0 \ \cdots \ 0 \ \cdots \ 1],$$

8. 3 求解对偶支撑向量机最佳值

利用二次规划(QP)实现支撑向量机对偶问题求解

① $\mathbf{Q}=\begin{bmatrix} y_1y_1\mathbf{z}_1^T\mathbf{z}_1 & \cdots & y_1y_N\mathbf{z}_1^T\mathbf{z}_N \\ \vdots & y_ny_m\mathbf{z}_n^T\mathbf{z}_m & \vdots \\ y_Ny_1\mathbf{z}_N^T\mathbf{z}_1 & \cdots & y_Ny_N\mathbf{z}_N^T\mathbf{z}_N \end{bmatrix}$, $\mathbf{p}=-\mathbf{1}_N$,

$\mathbf{a}_n^T=[0 \quad \cdots \quad 1 \quad \cdots \quad 0]$, $c_n=0$,

$\mathbf{a}_{\geq}^T=[y_1 \quad \cdots \quad y_n \quad \cdots \quad y_N]$, $c_{\geq}=0$,

$\mathbf{a}_{\leq}^T=[-y_1 \quad \cdots \quad -y_n \quad \cdots \quad -y_N]$, $c_{\leq}=0$,

② $[\alpha_1 \quad \cdots \quad \alpha_n \quad \cdots \quad \alpha_N]^T \leftarrow \mathrm{QP}(\mathbf{Q},\mathbf{p},\mathbf{A},\mathbf{c})$

③ 返回最终的$\boldsymbol{\alpha}$

8. 3 求解对偶支撑向量机最佳值

求解最佳的$(b,\mathbf{w})$

如果$(b,\mathbf{w},\boldsymbol{\alpha})$是原问题-对偶问题的最佳解(*primal-dual optimal*)：

- 原问题可行解(*primal feasible*)： $y_n(\mathbf{w}^T\mathbf{z}_n+b)\geq 1$
- 对偶问题可行解(*dual feasible*)： $\alpha_n\geq 0$
- 对偶“括号”内优化解(*dual-inner optimal*)：$\sum_{n=1}^{N}\alpha_n y_n=0$, $\mathbf{w}=\sum_{n=1}^{N}\alpha_n y_n \mathbf{z}_n$
- 原问题“括号”内优化解(*primal-inner optimal*)： $\alpha_n(1-y_n(\mathbf{w}^T\mathbf{z}_n+b))=0$

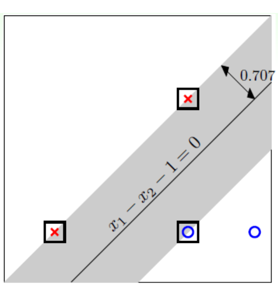
----称为KKT条件，为优化的充要条件

挑选任意一个 $\alpha_n>0$ 的样本， $b=y_n-\mathbf{w}^T\mathbf{z}_n$ 　 $\alpha_n>0$ 的样本是支撑向量

第八讲 对偶支撑向量机与核支撑向量机(*Dual SVM & Kernel SVM*)

7.3 支撑向量机

为什么叫支撑向量机?



$$\mathbf{X} = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \\ 3 & 0 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} -1 \\ -1 \\ +1 \\ +1 \end{bmatrix}$$

$$\min_{w} \quad \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w}$$

Subject to $\quad y_n(\boldsymbol{w}^T \mathbf{x}_n + b) \geq 1, for\ all\ n$

优化得到的解为:

$w_1 = 1$, $w_2 = -1$, $b = -1$

$$g(\mathbf{x}) = \text{sign}(x_1 - x_2 - 1)$$

$$\text{margin}(\boldsymbol{w}) = \frac{1}{\|\boldsymbol{w}\|} = \frac{1}{\sqrt{2}}$$

- 分类面由边界上的样本确定，其他样本不起作用
- 边界上的样本被称为支撑向量(候选)

支撑向量机(SVM)—Support Vector Machine
----借助支撑向量学到间隔最大分类面

8. 4 对偶支撑向量机讨论

SVM原问题

- 分类面由边界上的样本确定，其他样本不起作用
- 边界上的样本被称为支撑向量(候选)

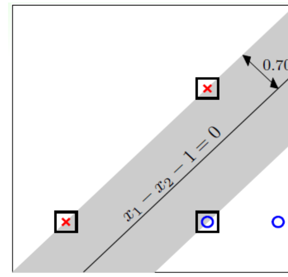
SVM对偶问题

- $\alpha_n > 0$ 的样本落在分类面的边界上
- $\alpha_n > 0$ 的样本$(\mathbf{z}_n, y_n)$被称为支撑向量(~~候选~~)

$SV(\alpha_n > 0$的样本$) \subseteq SV($边界上的样本$)$

求解$\boldsymbol{w}$时，只需要支撑向量: $\boldsymbol{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{z}_n = \sum_{\mathrm{SV}} \alpha_n y_n \mathbf{z}_n$

求解b时，只需要任一个支撑向量$(\mathbf{z}_n, y_n)$： $b = y_n - \boldsymbol{w}^T \mathbf{z}_n$



SVM
----利用对偶问题的最佳解确定支撑向量，从而找到间隔最大的分类面

8. 4 对偶支撑向量机讨论

SVM：

$$\boldsymbol{w}_{SVM} = \sum_{n=1}^{N} \alpha_n (y_n \mathbf{z}_n)$$

α_n由对偶问题的解确定

PLA：

$$\boldsymbol{w}_{PLA} = \sum_{n=1}^{N} \beta_n (y_n \mathbf{z}_n)$$

β_n由分错的样本确定

$\boldsymbol{w}$ 是 $y_n \mathbf{z}_n$ 的线性组合

$\boldsymbol{w}$ 被数据 $y_n \mathbf{z}_n$ 所表达

$\boldsymbol{w}$ 体现出“模式”

SVM： 仅通过SV表达$\boldsymbol{w}$

8.4 对偶支撑向量机讨论

SVM的原问题求解：

$$\min_{\mathbf{w}} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$Subject\ to \quad y_n(\mathbf{w}^T\mathbf{z}_n + b) \geq 1$$

$$for\ n = 1,2,\ldots N$$

- $(\tilde{d}+1)$ 个变量和N 个约束条件
- 求解最佳$(b, \mathbf{w})$

SVM的对偶问题求解：

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n\alpha_m y_n y_m \mathbf{z}_n^T\mathbf{z}_m - \sum_{n=1}^{N}\alpha_n$$

$$Subject\ to \quad \sum_{n=1}^{N}\alpha_n y_n = 0$$

$$\alpha_n \geq 0, for\ n = 1,2,\ldots,N$$

- N 个变量和 $N+1$ 个约束条件
- 求解最佳$\boldsymbol{\alpha}$，确定支撑向量

两种方法都能得到最佳解$(b, \mathbf{w})$获得最大间隔分类面$g_{SVM} = \mathrm{sign}(\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_n) + b)$

第八讲 对偶支撑向量机与核支撑向量机(*Dual SVM & Kernel SVM*)

8.5 核函数支撑向量机

研究对偶SVM的动机是不想依赖$\tilde{d}$

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \mathbf{z}_n^T\mathbf{z}_m - \sum_{n=1}^{N}\alpha_n$$

$$Subject\ to \quad \sum_{n=1}^{N}\alpha_n y_n = 0$$

$$\alpha_n \geq 0, for\ n = 1,2,\ldots,N$$

8.5 核函数支撑向量机

研究对偶SVM的动机是不想依赖$\tilde{d}$

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha} - \mathbf{1}^T\boldsymbol{\alpha}$$

$$Subject\ to \quad \mathbf{y}^T\boldsymbol{\alpha} = 0$$

$$\alpha_n \geq 0, for\ n = 1,2,\ldots,N$$

$$\mathbf{Q} = \begin{bmatrix} y_1y_1\mathbf{z}_1^T\mathbf{z}_1 & \cdots & y_1y_N\mathbf{z}_1^T\mathbf{z}_N \\ \vdots & y_ny_m\mathbf{z}_n^T\mathbf{z}_m & \vdots \\ y_Ny_1\mathbf{z}_N^T\mathbf{z}_1 & \cdots & y_Ny_N\mathbf{z}_N^T\mathbf{z}_N \end{bmatrix}$$

对偶SVM真的不依赖于$\tilde{d}$?

- N 个变量和 $N+1$ 个约束条件
- 对偶问题的$\mathbf{Q}$ 是个稠密矩阵
- 每个元素$q_{n,m} = y_ny_m\mathbf{z}_n^T\mathbf{z}_m$都要做内积运算，计算代价$O(\tilde{d})$

能提高 $\mathbf{z}_n^T\mathbf{z}_m = \Phi(\mathbf{x}_n)^T\Phi(\mathbf{x}_m)$ 计算效率吗?

8.5 核函数支撑向量机

二次多项式$\boldsymbol{\Phi}_2(\mathbf{x})$的快速内积计算

$$\boldsymbol{\Phi}_2(\mathbf{x}) = (1, x_1, x_2, \ldots, x_d, x_1^2, x_1x_2, \ldots, x_1x_d, x_2x_1, x_2^2, \ldots, x_2x_d, x_dx_1, x_dx_2, \ldots, x_d^2)$$

$$\begin{aligned}\boldsymbol{\Phi}_2^T(\mathbf{x})\boldsymbol{\Phi}_2(\mathbf{x}') &= 1 + \sum_{i=1}^{d} x_i x_i' + \sum_{i=1}^{d}\sum_{j=1}^{d} x_i x_j x_i' x_j' \\ &= 1 + \sum_{i=1}^{d} x_i x_i' + \sum_{i=1}^{d} x_i x_i' \sum_{j=1}^{d} x_j x_j' \\ &= 1 + \mathbf{x}^T\mathbf{x}' + (\mathbf{x}^T\mathbf{x}')(\mathbf{x}^T\mathbf{x}')\end{aligned}$$

计算复杂度从$O(\tilde{d})$下降到$O(d)$

8.5 核函数支撑向量机

核函数： Φ变换+内积计算 ➡ $K_{\boldsymbol{\Phi}}(\mathbf{x}, \mathbf{x}') = \boldsymbol{\Phi}^{\boldsymbol{T}}(\mathbf{x})\boldsymbol{\Phi}(\mathbf{x}')$

$$K_{\boldsymbol{\Phi}_2}(\mathbf{x}, \mathbf{x}') = \boldsymbol{\Phi}_2^T(\mathbf{x})\boldsymbol{\Phi}_2(\mathbf{x}') = 1 + \mathbf{x}^T\mathbf{x}' + (\mathbf{x}^T\mathbf{x}')(\mathbf{x}^T\mathbf{x}')$$

$$q_{n,m} = y_n y_m \mathbf{z}_n^T \mathbf{z}_m \quad = y_n y_m \boldsymbol{\Phi}^{\boldsymbol{T}}(\mathbf{x}_n)\boldsymbol{\Phi}(\mathbf{x}_m) = y_n y_m K_{\boldsymbol{\Phi}}(\mathbf{x}_n, \mathbf{x}_m)$$

$$\boldsymbol{w} = \sum_{n=1}^{N} \alpha_n y_n \boldsymbol{z}_n = \sum_{\text{SV}} \alpha_n y_n \boldsymbol{z}_n$$

支撑SV$(\boldsymbol{z}_m, y_m)$： $b = y_m - \boldsymbol{w}^T\boldsymbol{z}_m = y_m - (\sum_{\text{SV}} \alpha_n y_n \boldsymbol{z}_n)^T \boldsymbol{z}_m$

输入测试样本 $\mathbf{x}$： $g_{SVM} = \text{sign}(\boldsymbol{w}^T\boldsymbol{\Phi}(\mathbf{x}) + b)$

8.5 核函数支撑向量机

核函数：Φ变换+内积计算 ➡ $K_{\Phi}(\mathbf{x},\mathbf{x}') = \boldsymbol{\Phi}^{T}(\mathbf{x})\boldsymbol{\Phi}(\mathbf{x}')$

$$K_{\Phi_2}(\mathbf{x},\mathbf{x}') = \boldsymbol{\Phi}_2^{T}(\mathbf{x})\boldsymbol{\Phi}_2(\mathbf{x}') = 1 + \mathbf{x}^T\mathbf{x}' + (\mathbf{x}^T\mathbf{x}')(\mathbf{x}^T\mathbf{x}')$$

$$q_{n,m} = y_n y_m \mathbf{z}_n^T \mathbf{z}_m = y_n y_m \boldsymbol{\Phi}^{T}(\mathbf{x}_n)\boldsymbol{\Phi}(\mathbf{x}_m) = y_n y_m K_{\Phi}(\mathbf{x}_n,\mathbf{x}_m)$$

$$\boldsymbol{w} = \sum_{n=1}^{N} \alpha_n y_n \boldsymbol{z}_n = \sum_{\mathrm{SV}} \alpha_n y_n \boldsymbol{z}_n$$

支撑SV$(\boldsymbol{z}_m, y_m)$：$b = y_m - \boldsymbol{w}^T\boldsymbol{z}_m = y_m - \sum_{\mathrm{SV}} \alpha_n y_n \boldsymbol{\Phi}^{T}(\mathbf{x}_n)\,\boldsymbol{\Phi}(\mathbf{x}_m)$

输入测试样本 $\mathbf{x}$： $g_{SVM} = \mathrm{sign}(\boldsymbol{w}^T\boldsymbol{\Phi}(\mathbf{x}) + b)$

8.5 核函数支撑向量机

核函数：Φ变换+内积计算 ➡ $K_{\Phi}(\mathbf{x},\mathbf{x}') = \boldsymbol{\Phi}^{T}(\mathbf{x})\boldsymbol{\Phi}(\mathbf{x}')$

$$K_{\Phi_2}(\mathbf{x},\mathbf{x}') = \boldsymbol{\Phi}_2^{T}(\mathbf{x})\boldsymbol{\Phi}_2(\mathbf{x}') = 1 + \mathbf{x}^T\mathbf{x}' + (\mathbf{x}^T\mathbf{x}')(\mathbf{x}^T\mathbf{x}')$$

$$q_{n,m} = y_n y_m \mathbf{z}_n^T \mathbf{z}_m = y_n y_m \boldsymbol{\Phi}^{T}(\mathbf{x}_n)\boldsymbol{\Phi}(\mathbf{x}_m) = y_n y_m K_{\Phi}(\mathbf{x}_n,\mathbf{x}_m)$$

$$\boldsymbol{w} = \sum_{n=1}^{N} \alpha_n y_n \boldsymbol{z}_n = \sum_{\mathrm{SV}} \alpha_n y_n \boldsymbol{z}_n$$

支撑SV$(\boldsymbol{z}_m, y_m)$：$b = y_m - \boldsymbol{w}^T\boldsymbol{z}_m = y_m - \sum_{\mathrm{SV}} \alpha_n y_n K_{\Phi}(\mathbf{x}_n,\mathbf{x}_m)$

输入测试样本 $\mathbf{x}$： $g_{SVM} = \mathrm{sign}(\boldsymbol{w}^T\boldsymbol{\Phi}(\mathbf{x}) + b) = sign((\sum_{\mathrm{SV}} \alpha_n y_n \boldsymbol{z}_n)^T\boldsymbol{\Phi}(\mathbf{x}) + b)$

8.5 核函数支撑向量机

核函数：Φ变换+内积计算 ⟶ $K_{\Phi}(\mathbf{x},\mathbf{x}') = \boldsymbol{\Phi}^{T}(\mathbf{x})\boldsymbol{\Phi}(\mathbf{x}')$

$$K_{\Phi_2}(\mathbf{x},\mathbf{x}') = \boldsymbol{\Phi}_2^{T}(\mathbf{x})\boldsymbol{\Phi}_2(\mathbf{x}') = 1 + \mathbf{x}^T\mathbf{x}' + (\mathbf{x}^T\mathbf{x}')(\mathbf{x}^T\mathbf{x}')$$

$$q_{n,m} = y_n y_m \mathbf{z}_n^T \mathbf{z}_m = y_n y_m \boldsymbol{\Phi}^{T}(\mathbf{x}_n)\boldsymbol{\Phi}(\mathbf{x}_m) = y_n y_m K_{\Phi}(\mathbf{x}_n,\mathbf{x}_m)$$

$$\boldsymbol{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{z}_n = \sum_{\mathrm{SV}} \alpha_n y_n \mathbf{z}_n$$

支撑SV$(\mathbf{z}_m, y_m)$：$b = y_m - \boldsymbol{w}^T\mathbf{z}_m = y_m - \sum_{\mathrm{SV}} \alpha_n y_n K_{\Phi}(\mathbf{x}_n,\mathbf{x}_m)$

输入测试样本 $\mathbf{x}$：$g_{SVM} = \mathrm{sign}(\boldsymbol{w}^T\boldsymbol{\Phi}(\mathbf{x}) + b) = sign(\sum_{\mathrm{SV}} \alpha_n y_n \boldsymbol{\Phi}^{T}(\mathbf{x}_n)\boldsymbol{\Phi}(\mathbf{x}) + b)$

8.5 核函数支撑向量机

核函数：Φ变换+内积计算 ⟶ $K_{\Phi}(\mathbf{x},\mathbf{x}') = \boldsymbol{\Phi}^{T}(\mathbf{x})\boldsymbol{\Phi}(\mathbf{x}')$

$$K_{\Phi_2}(\mathbf{x},\mathbf{x}') = \boldsymbol{\Phi}_2^{T}(\mathbf{x})\boldsymbol{\Phi}_2(\mathbf{x}') = 1 + \mathbf{x}^T\mathbf{x}' + (\mathbf{x}^T\mathbf{x}')(\mathbf{x}^T\mathbf{x}')$$

$$q_{n,m} = y_n y_m \mathbf{z}_n^T \mathbf{z}_m = y_n y_m \boldsymbol{\Phi}^{T}(\mathbf{x}_n)\boldsymbol{\Phi}(\mathbf{x}_m) = y_n y_m K_{\Phi}(\mathbf{x}_n,\mathbf{x}_m)$$

$$\boldsymbol{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{z}_n = \sum_{\mathrm{SV}} \alpha_n y_n \mathbf{z}_n$$

利用核函数，将计算复杂度从$O(\tilde{d})$下降到$O(d)$避免依赖$\tilde{d}$

支撑SV$(\mathbf{z}_m, y_m)$：$b = y_m - \boldsymbol{w}^T\mathbf{z}_m = y_m - \sum_{\mathrm{SV}} \alpha_n y_n K_{\Phi}(\mathbf{x}_n,\mathbf{x}_m)$

输入测试样本 $\mathbf{x}$：$g_{SVM} = \mathrm{sign}(\boldsymbol{w}^T\boldsymbol{\Phi}(\mathbf{x}) + b) = sign(\sum_{\mathrm{SV}} \alpha_n y_n K_{\Phi}(\mathbf{x}_n,\mathbf{x}) + b)$

8.5 核函数支撑向量机

利用二次规划(QP)实现核函数支撑向量机求解

① $q_{n,m} = y_n y_m K_{\boldsymbol{\Phi}}(\mathbf{x}_n, \mathbf{x}_m)$， $\mathbf{p} = -\mathbf{1}_N$，由约束条件得到$(\mathbf{A}, \mathbf{c})$

② $\boldsymbol{\alpha} \leftarrow \mathrm{QP}(\mathbf{Q}, \mathbf{p}, \mathbf{A}, \mathbf{c})$

③ 任选一支撑SV$(\mathbf{x}_m, y_m)$： $b \leftarrow \left(y_m - \sum_{\mathrm{SV}} \alpha_n y_n K_{\boldsymbol{\Phi}}(\mathbf{x}_n, \mathbf{x}_m)\right)$

④ 对于测试样本 $\mathbf{x}$： $g_{SVM} = sign\left(\sum_{\mathrm{SV}} \alpha_n y_n K_{\boldsymbol{\Phi}}(\mathbf{x}_n, \mathbf{x}) + b\right)$

- 步骤①的时间复杂度： $O(N^2) \cdot (kernel\ evaluation)$
- 步骤②的开销 ： N个变量，$N+1$个约束
- 步骤③和④的复杂度： $O(\#\mathrm{SV}) \cdot (kernel\ evaluation)$

8.5 核函数支撑向量机

二次多项式核函数的一般表达式

$\boldsymbol{\Phi}_2(\mathbf{x})$	$K_2(\mathbf{x}, \mathbf{x}')$
$(1, x_1, \dots, x_d, x_1^2, x_1x_2, \dots, x_d^2)$	$1 + \mathbf{x}^T\mathbf{x}' + (\mathbf{x}^T\mathbf{x}')^2$
$(1, \sqrt{2}x_1, \dots, \sqrt{2}x_d, x_1^2, x_1x_2, \dots, x_d^2)$	$1 + 2\mathbf{x}^T\mathbf{x}' + (\mathbf{x}^T\mathbf{x}')^2$
$(1, \sqrt{2\gamma}x_1, \dots, \sqrt{2\gamma}x_d, x_1^2, x_1x_2, \dots, x_d^2)$	$1 + 2\gamma\mathbf{x}^T\mathbf{x}' + \gamma^2(\mathbf{x}^T\mathbf{x}')^2$

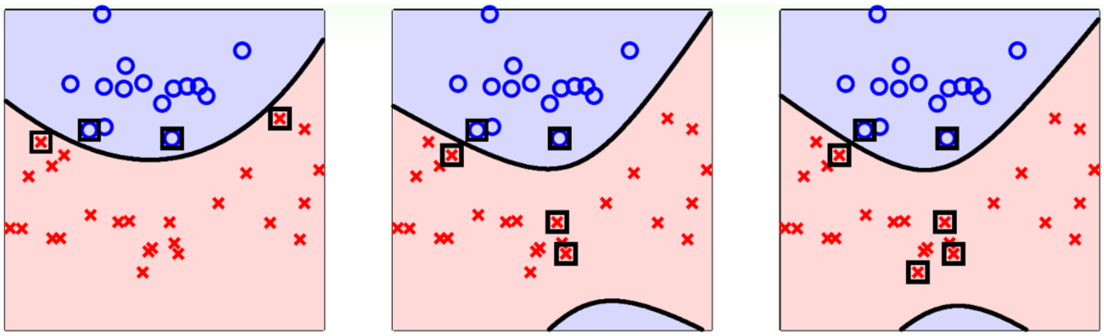
$$K_2(\mathbf{x}, \mathbf{x}') = (1 + \gamma\mathbf{x}^T\mathbf{x}')^2 \qquad \gamma > 0$$

不同的二次多项式变换：

- 升维次数相同
- 内积结果不同 ➡ 分类面不同

8.5 核函数支撑向量机

二次多项式核函数的一般表达式



$(1+0.001\mathbf{x}^T\mathbf{x}')^2$　　$1+\mathbf{x}^T\mathbf{x}'+(\mathbf{x}^T\mathbf{x}')^2$　　$(1+1000\mathbf{x}^T\mathbf{x}')^2$

- 核函数不同 ➡ 支撑向量(SVs)不同，分类面(g_{SVM})不同
- 核函数变化 ➡ Margin也会变化

8.5 核函数支撑向量机

多项式核函数的一般表达式

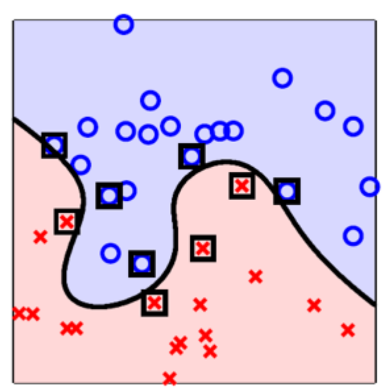
$$K_2(\mathbf{x},\mathbf{x}')=(\zeta+\gamma\mathbf{x}^T\mathbf{x}')^2 \qquad \gamma>0,\zeta\geq 0$$

$$K_3(\mathbf{x},\mathbf{x}')=(\zeta+\gamma\mathbf{x}^T\mathbf{x}')^3 \qquad \gamma>0,\zeta\geq 0$$

$$\vdots$$

$$K_Q(\mathbf{x},\mathbf{x}')=(\zeta+\gamma\mathbf{x}^T\mathbf{x}')^Q \qquad \gamma>0,\zeta\geq 0$$

- Q、γ、ζ 确定了多项式核函数的形式
- 利用核函数可不依赖于$\tilde{d}$ 获得大间隔分类面



*Margin*为1的10次多项式

SVM + *Polynomial Kernel* = *Polynomial SVM*

8.5 核函数支撑向量机

核函数能实现无穷维变换?

当$\boldsymbol{\Phi}(\mathbf{x})$为无穷维变换时，借助核函数$\boldsymbol{K}(\mathbf{x},\mathbf{x}')$技巧仍可进行有效计算

为简单起见，考虑一维变量，即：$\mathbf{x}=(x)$

$$K(x,x') = \exp(-(x-x')^2) = \exp(-(x)^2)\exp(-(x')^2)\exp(2xx')$$

$$= \exp(-(x)^2)\exp(-(x')^2)\left(\sum_{i=0}^{\infty}\frac{(2xx')^i}{i!}\right)$$

$$= \sum_{i=0}^{\infty}\left(\exp(-(x)^2)\exp(-(x')^2)\sqrt{\frac{2^i}{i!}}\sqrt{\frac{2^i}{i!}}(x)^i(x')^i\right)$$

$$\boldsymbol{\Phi}(x) = \exp(-(x)^2)\cdot\left(1, \sqrt{\frac{2^1}{1!}}x, \sqrt{\frac{2^2}{2!}}x^2, \sqrt{\frac{2^3}{3!}}x^3, \dots, \dots\right)$$

8.5 核函数支撑向量机

核函数能实现无穷维变换?

当$\boldsymbol{\Phi}(\mathbf{x})$为无穷维变换时，借助核函数$\boldsymbol{K}(\mathbf{x},\mathbf{x}')$技巧仍可进行有效计算

为简单起见，考虑一维变量，即：$\mathbf{x}=(x)$

$$K(x,x') = \boldsymbol{\Phi}(x)^T\boldsymbol{\Phi}(x') = \exp(-(x-x')^2) = \exp(-(x)^2)\exp(-(x')^2)\exp(2xx')$$

$$= \exp(-(x)^2)\exp(-(x')^2)\left(\sum_{i=0}^{\infty}\frac{(2xx')^i}{i!}\right)$$

高斯核函数$\boldsymbol{K}(\mathbf{x},\mathbf{x}')$的一般式

$$K(\mathbf{x},\mathbf{x}') = \exp(-\gamma\|\mathbf{x}-\mathbf{x}'\|^2)$$

$$\gamma > 0$$

$$= \sum_{i=0}^{\infty}\left(\exp(-(x)^2)\exp(-(x')^2)\sqrt{\frac{2^i}{i!}}\sqrt{\frac{2^i}{i!}}(x)^i(x')^i\right)$$

$$\boldsymbol{\Phi}(x) = \exp(-(x)^2)\cdot\left(1, \sqrt{\frac{2^1}{1!}}x, \sqrt{\frac{2^2}{2!}}x^2, \sqrt{\frac{2^3}{3!}}x^3, \dots, \dots\right)$$

8.5 核函数支撑向量机

高斯核函数SVM

$$K(\mathbf{x},\mathbf{x}') = \exp(-\gamma\|\mathbf{x}-\mathbf{x}'\|^2) \qquad \gamma > 0$$

输入测试样本 $\mathbf{x}$：
$$g_{SVM} = sign(\sum_{SV} \alpha_n y_n K_{\Phi}(\mathbf{x}_n,\mathbf{x}) + b)$$
$$= sign(\sum_{SV} \alpha_n y_n \exp(-\gamma\|\mathbf{x}-\mathbf{x}_n\|^2) + b)$$

- 以所有支撑向量 $\mathbf{x}_n$ 为中心的高斯函数的线性组合
- 也被称之为径向基核函数(Radial Basis Function, RBF)

高斯核函数SVM：通过求取α_n确定所有支撑向量 $\mathbf{x}_n$ ，构造以支撑向量为中心的高斯函数的线性组合，实现在无穷维空间获得最大间隔分类面

第八讲 对偶支撑向量机与核支撑向量机(*Dual SVM & Kernel SVM*)

8.1 对偶支撑向量机动机(*Motivation of Dual SVM*)

希望不依赖于非线性变换后升维的d̃

8.2 对偶支撑向量机的拉格朗日分析(*Lagrange Dual SVM*)

通过KKT 条件 将原问题和对偶问题相关联获得最佳分类面

8.3 求解对偶支撑向量机最佳值(*Solving Dual SVM*)

仍然是二次规划问题，可以方便求解

8.4 对偶支撑向量机讨论(*Messages behind Dual SVM*)

由支撑向量确定最大间隔分类面

8.5 核函数支撑向量机(*Kernel SVM*)

利用核函数避免了对升维后d̃ 的依赖，且高效求解