如何处理非线性的分类问题

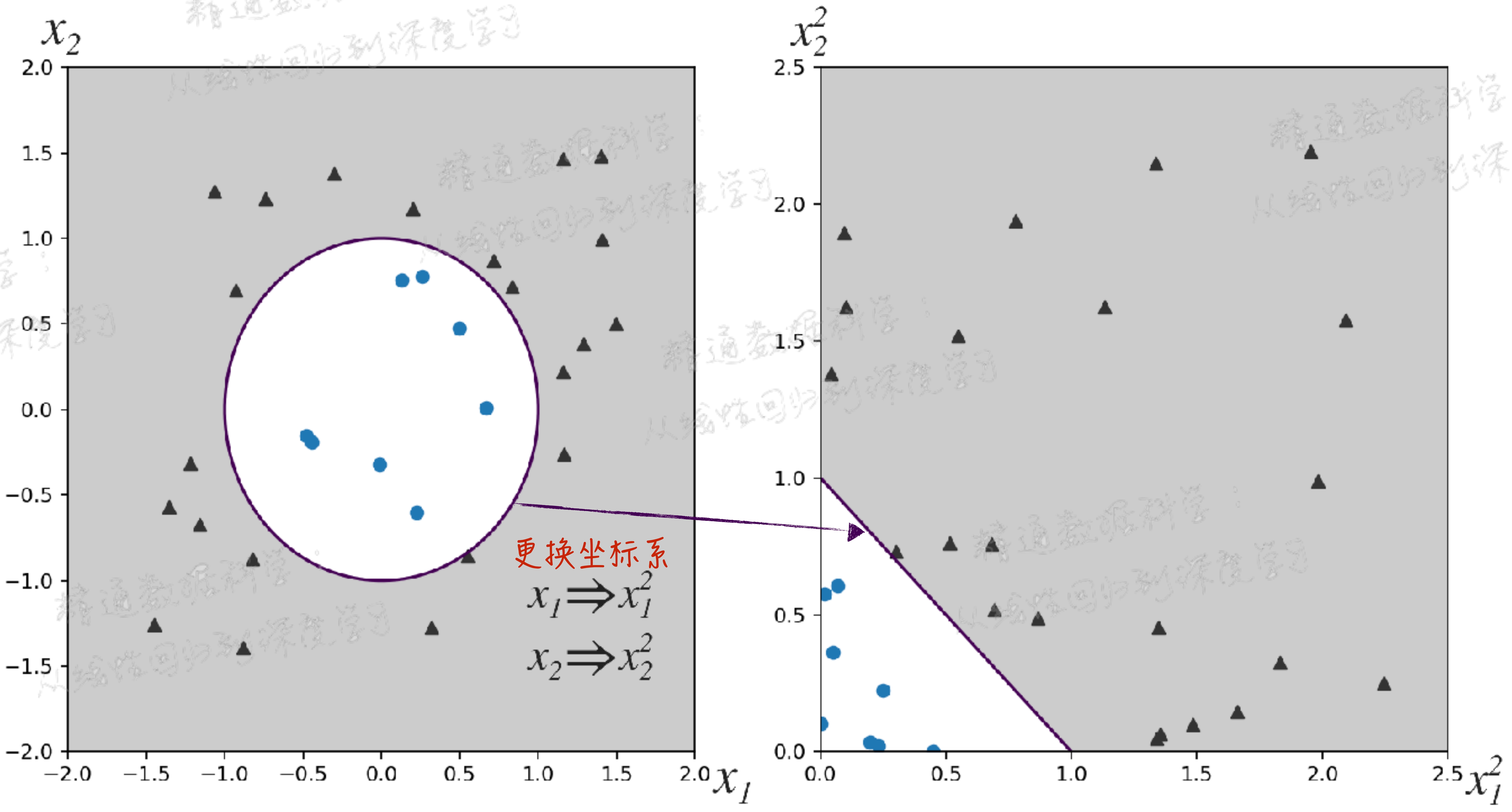
拉格朗日对偶

小胖

目录

从非线性到线性

空间变换



从非线性到线性

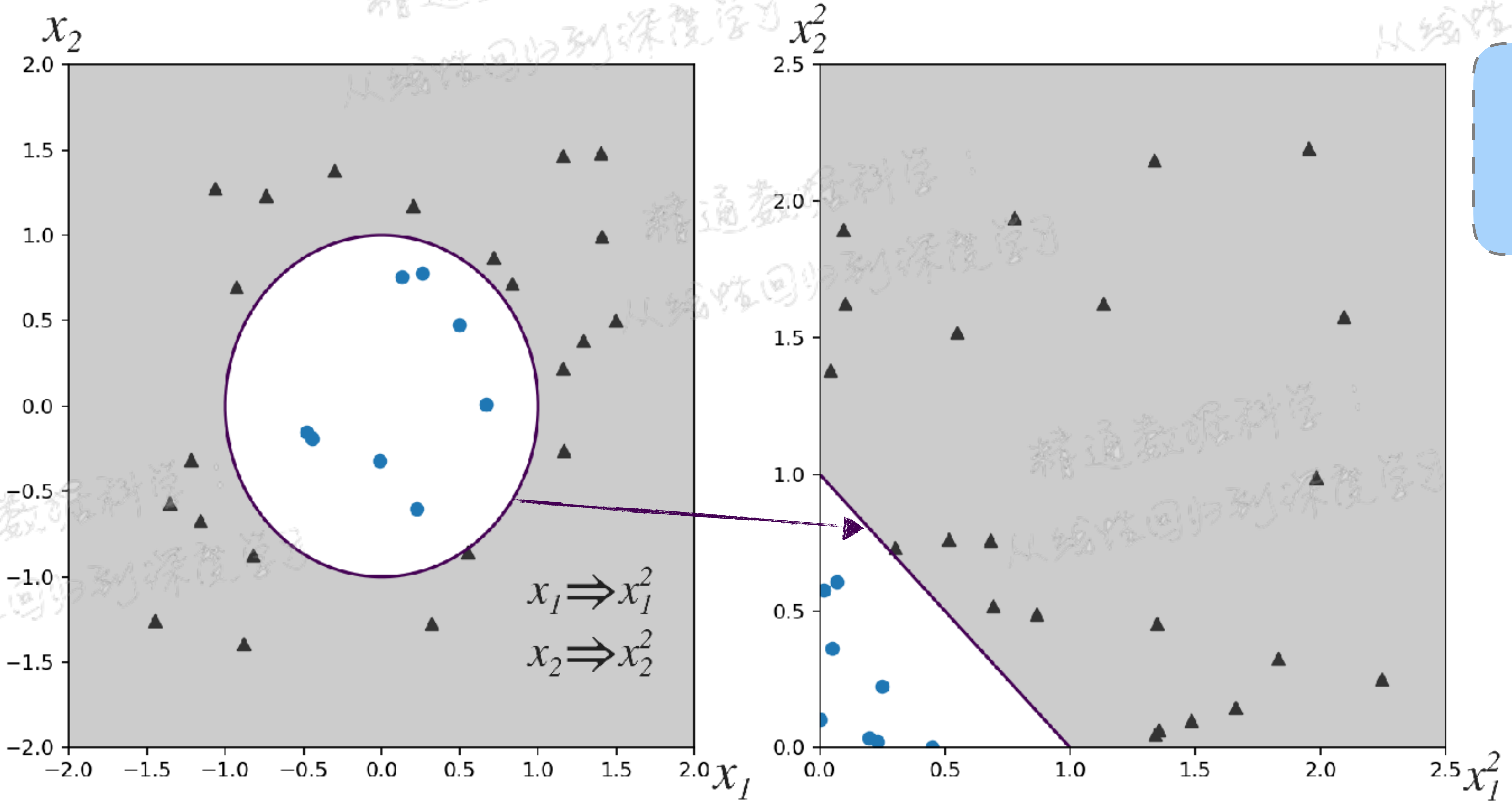
空间变换

非线性变换：φ

$$\varphi(x_1) = x_1^2, \quad \varphi(x_2) = x_2^2$$

原始数据 $\mathbf{X}_i, y_i$ 线性不可分

变换后的数据 $\varphi(\mathbf{X}_i), y_i$ 线性可分



- 变换函数很难定义
- 运算复杂程度高

核函数解决的问题

目录

拉格朗日对偶

什么是拉格朗日对偶

伟大的法国数学家

Joseph Louis de Lagrange

与乾隆皇帝同时代

拉格朗日对偶问题：

- 解决约束条件下的最优化问题

拉格朗日对偶

定义原始问题

SVM模型对应的最优化问题

最优化问题 $$\min \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_i \xi_i$$

限制条件 $$y_i(\mathbf{w} \cdot \mathbf{X}_i + c) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

$$\Rightarrow$$

$$\min \underbrace{\boxed{\frac{1}{2} \| \mathbf{w} \|^2}}_{\text{惩罚项}} + \underbrace{\boxed{C \sum_i \max(0,1 - y_i(\mathbf{w} \cdot \mathbf{X}_i + c))}}_{\text{模型的预测损失(hinge loss)}}$$

无限制条件的最优化问题

$$\min f(\theta)$$

$$g_i(\theta) \leq 0, \quad i = 1,\ldots,k$$

$$h_i(\theta) = 0, \quad i = 1,\ldots,l$$

原始问题
primal optimization problem

$$\Updownarrow$$

$$\min_\theta P(\theta) = \min_\theta \max_{\alpha \geq 0, \beta} L(\theta, \alpha, \beta)$$

定义拉格朗日函数

$$L(\theta, \alpha, \beta) = f(\theta) + \sum_i \alpha_i g_i(\theta) + \sum_i \beta_i h_i(\theta)$$

$$P(\theta) = \max_{\alpha \geq 0, \beta} L(\theta, \alpha, \beta)$$

$$\xrightarrow{\text{满足限制条件}} P(\theta) = f(\theta)$$

$$\xrightarrow[\text{不满足限制条件}]{} P(\theta) = +\infty$$

拉格朗日对偶

定义对偶问题

两个问题在一定条件下等价

两个问题的最优解一定存在，且满足KKT条件

$$\min f(\theta)$$

$$g_i(\theta) \le 0, \quad i = 1,\ldots,k$$

$$h_i(\theta) = 0, \quad i = 1,\ldots,l$$

$$L(\theta,\alpha,\beta) = f(\theta) + \sum_i \alpha_i g_i(\theta) + \sum_i \beta_i h_i(\theta)$$

原始问题

primal optimization problem

$$\min_{\theta} P(\theta) = \min_{\theta} \max_{\alpha\ge 0,\beta} L(\theta,\alpha,\beta)$$

对偶问题

dual optimization problem

$$D(\alpha,\beta) = \min_{\theta} L(\theta,\alpha,\beta)$$

$$\max_{\alpha\ge 0,\beta} D(\alpha,\beta) = \max_{\alpha\ge 0,\beta} \min_{\theta} L(\theta,\alpha,\beta)$$

$$L(\hat\theta,\hat\alpha,\hat\beta) = \max_{\alpha\ge 0,\beta} \min_{\theta} L(\theta,\alpha,\beta) = \min_{\theta} \max_{\alpha\ge 0,\beta} L(\theta,\alpha,\beta)$$

KKT条件

$$\frac{\partial L}{\partial \theta}(\hat\theta,\hat{\boldsymbol\alpha},\hat{\boldsymbol\beta}) = 0$$

$$\frac{\partial L}{\partial \beta}(\hat\theta,\hat{\boldsymbol\alpha},\hat{\boldsymbol\beta}) = 0$$

$$\hat\alpha_i g_i(\hat\theta) = 0$$

$$g_i(\hat\theta) \leqslant 0$$

$$\hat\alpha_i \geqslant 0$$

拉格朗日对偶

SVM的对偶问题

$$L(\hat{\theta}, \hat{\alpha}, \hat{\beta}) = \max_{\alpha \geq 0, \beta} \min_{\theta} L(\theta, \alpha, \beta) = \min_{\theta} \max_{\alpha \geq 0, \beta} L(\theta, \alpha, \beta)$$

$$L(\theta, \alpha, \beta) = f(\theta) + \sum_i \alpha_i g_i(\theta) + \sum_i \beta_i h_i(\theta)$$

$$\xrightarrow{\text{KKT条件}}$$

$$\frac{\partial L}{\partial \theta}(\hat{\theta}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = 0$$

$$\frac{\partial L}{\partial \beta}(\hat{\theta}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = 0$$

$$\hat{\alpha}_i g_i(\hat{\theta}) = 0$$

$$g_i(\hat{\theta}) \leqslant 0$$

$$\hat{\alpha}_i \geqslant 0$$

SVM的拉格朗日函数

$$L(w, c, \xi, \alpha, \gamma) = \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_i \xi_i - \sum_i \alpha_i [y_i (w \cdot X_i + c) - 1 + \xi_i] - \sum_i \gamma_i \xi_i$$

$$\frac{\partial L}{\partial w} = 0 \Rightarrow \hat{\boldsymbol{w}} = \sum_i \hat{\alpha}_i y_i \boldsymbol{X}_i$$

$$\frac{\partial L}{\partial \xi} = 0 \Rightarrow C - \hat{\alpha}_i - \hat{\gamma}_i = 0$$

$$\frac{\partial L}{\partial c} = 0 \Rightarrow \sum_i y_i \hat{\alpha}_i = 0$$

$$D(\alpha, \gamma) = \min_{\theta} L(\theta, \alpha, \beta)$$

$$D(\alpha, \gamma) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j (X_i \cdot X_j)$$

$$0 \leq \alpha_i \leq C; \quad \sum_i \alpha_i y_i = 0$$

拉格朗日对偶

SVM的对偶问题

Original SVM

参数估计公式

$$\min_{w,c} \frac{1}{2}\|w\|^2 + C\sum_i \xi_i$$
$$s.t.\ \ y_i(w \cdot X_i + c) \geq 1 - \xi_i$$
$$\xi_i \geq 0, \forall i$$

预测公式

$$\hat{y}_j = \text{sign}(\hat{w} \cdot X_j + \hat{c})$$

Dual problem

参数估计公式

$$\max_{\alpha} \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} y_i y_j \alpha_i \alpha_j (X_i \cdot X_j)$$
$$s.t.\ \ 0 \leq \alpha_i \leq C$$
$$\sum_i \alpha_i y_i = 0, \forall i$$

预测公式

$$\hat{y}_j = \text{sign}(\sum_i \hat{\alpha}_i y_i (X_i \cdot X_j) + \hat{c})$$

$$D(\alpha, \gamma) = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} y_i y_j \alpha_i \alpha_j (X_i \cdot X_j)$$

$$0 \leq \alpha_i \leq C; \quad \sum_i \alpha_i y_i = 0$$

目录

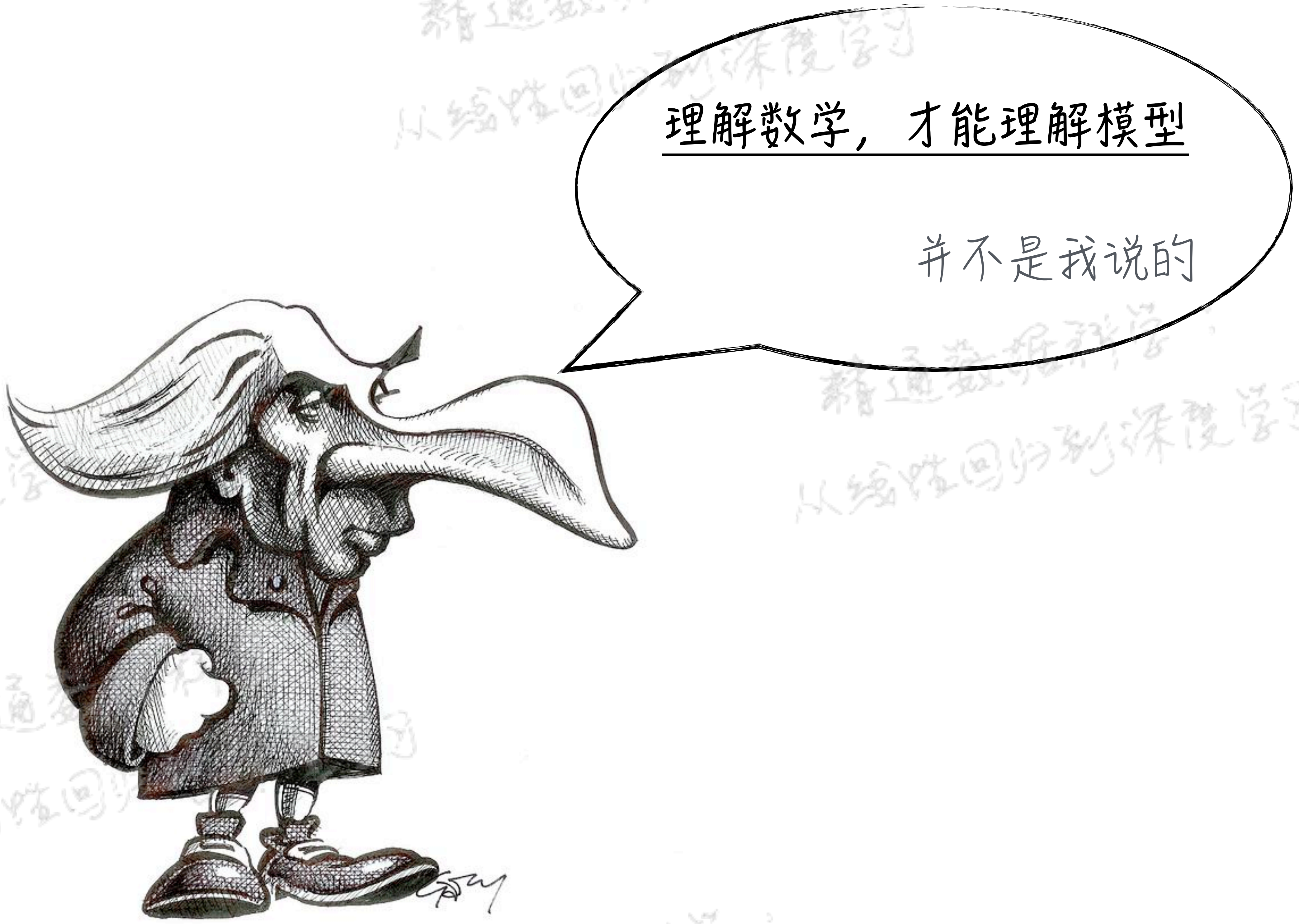
支持向量

SVM名字来源

Dual problem

$$\max_{\alpha} \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} y_i y_j \alpha_i \alpha_j (\boldsymbol{X_i} \cdot \boldsymbol{X_j})$$
$$s.t.\ \ 0 \le \alpha_i \le C$$
$$\sum_i \alpha_i y_i = 0, \forall i$$

$$\hat{y}_j = \mathrm{sign}(\sum_i \hat{\alpha}_i y_i (\boldsymbol{X_i} \cdot \boldsymbol{X_j}) + \hat{c})$$



- 用内积表示“点”与“点”之间的相似度
- 用内积作为权重去平均被预测量y，得到预测结果

支持向量

SVM名字来源

$$\min_{w,c} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_i \xi_i$$

$$s.t.\ \ y_i(\boldsymbol{w}\cdot \boldsymbol{X}_i + c) \geq 1 - \xi_i$$

$$\xi_i \geq 0, \forall i$$

$$\frac{\partial L}{\partial \theta}(\hat{\theta}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = 0$$

$$\frac{\partial L}{\partial \beta}(\hat{\theta}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = 0$$

$$\hat{\alpha}_i g_i(\hat{\theta}) = 0$$

$$g_i(\hat{\theta}) \leqslant 0$$

$$\hat{\alpha}_i \geqslant 0$$

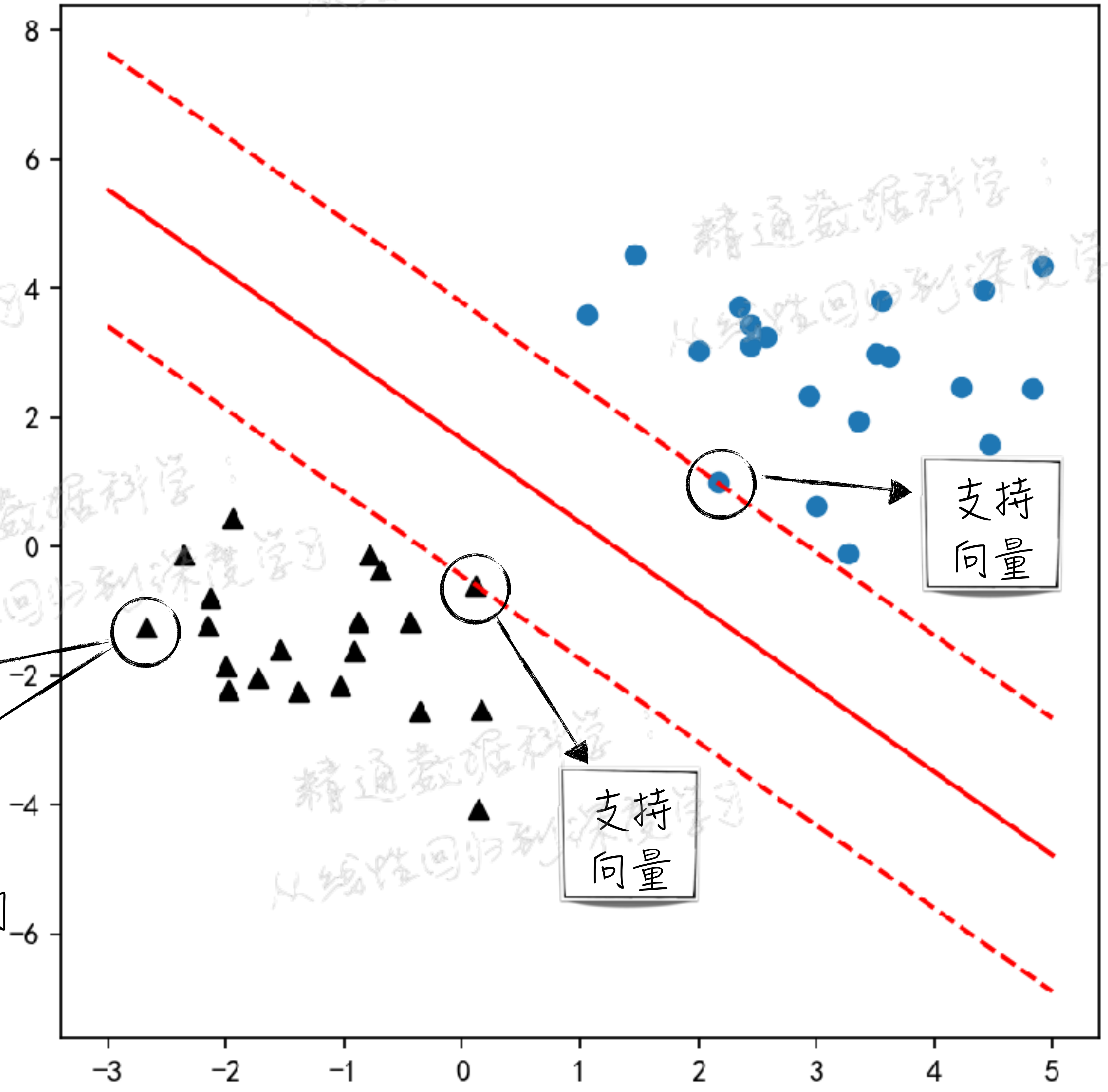
$$g_i = 1 - y_i(w\cdot X_i + c) - \xi_i$$

$$\hat{\alpha}_i g_i = 0$$

$$g_i < 0$$
$$\hat{\alpha}_i = 0$$

对预测结果的
权重等于0

$$\hat{y}_j = \mathrm{sign}(\sum_i \hat{\alpha}_i y_i(\boldsymbol{X}_i \cdot \boldsymbol{X}_j) + \hat{c})$$



THANK YOU