如何处理非线性的分类问题

核函数

小胖

目录

核函数的定义

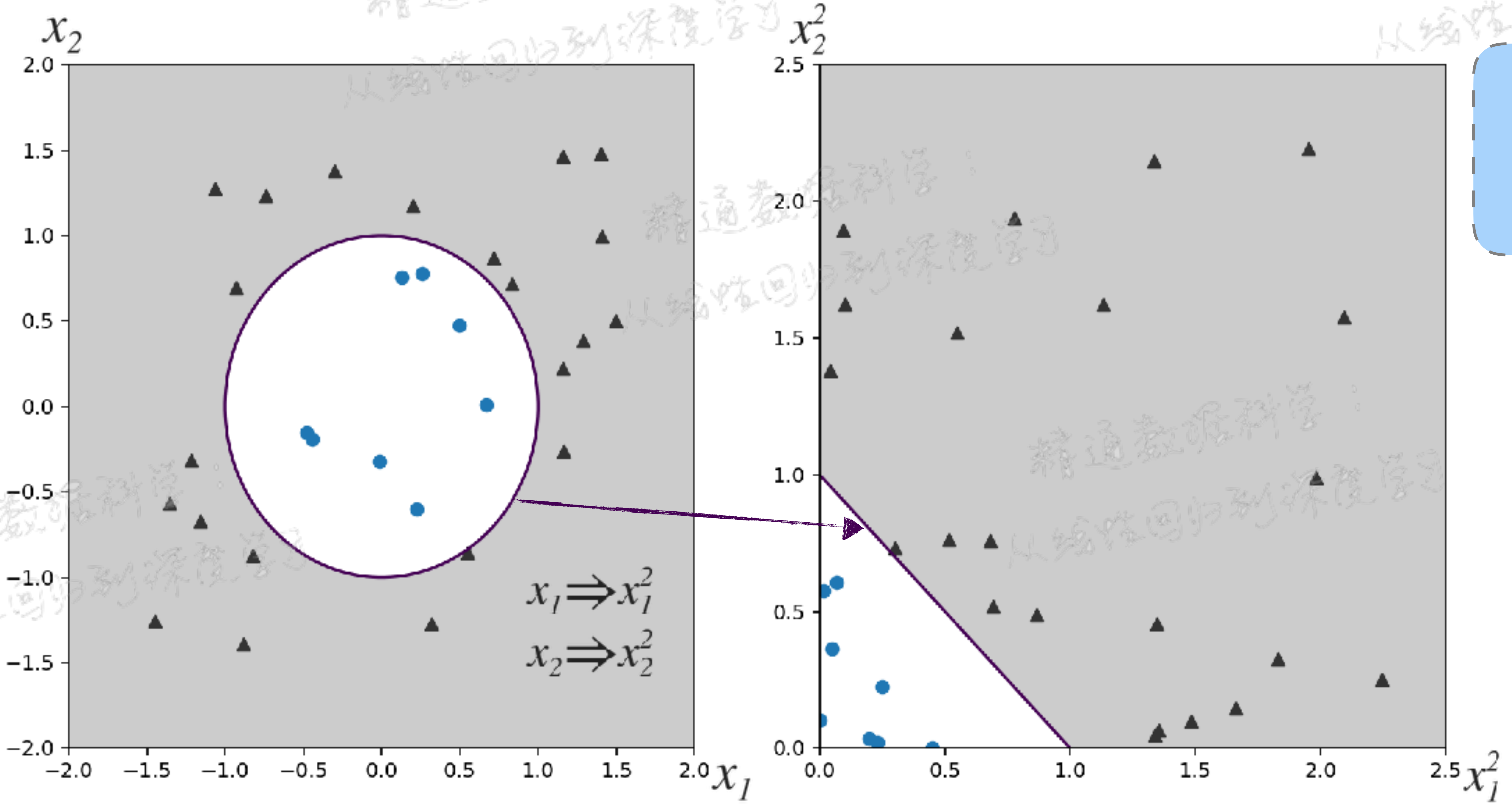
空间变换

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
- 运算复杂度高

核函数解决的问题

核函数的定义

SVM的对偶问题

Dual problem

$$\max_{\boldsymbol{\alpha}} \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} y_i y_j \alpha_i \alpha_j (\boldsymbol{X}_i \cdot \boldsymbol{X}_j)$$

$$s.t.\ \ 0 \le \alpha_i \le C$$

$$\sum_i \alpha_i y_i = 0, \forall i$$

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

- SVM模型在训练和预测时，只会用到内积运算
- 我们只关心非线性变换后，“新数据”的内积

$$\{\mathbf{X}_i, y_i\} \xrightarrow{\text{非线性变换：}\varphi} \{\varphi(\mathbf{X}_i), y_i\}$$

SVM只需要： $\varphi(\mathbf{X}_i) \cdot \varphi(\mathbf{X}_j)$

$$K(\mathbf{X}_i, \mathbf{X}_j) = \varphi(\mathbf{X}_i) \cdot \varphi(\mathbf{X}_j)$$ 这就是核函数

核函数的定义

一个例子

一个例子

$$K(\mathbf{X}_i, \mathbf{X}_j) = \varphi(\mathbf{X}_i) \cdot \varphi(\mathbf{X}_j)$$

$$\mathbf{X}_i = (x_{1,i}, x_{2,i})$$

$$\varphi(\mathbf{X}_i) = (x_{1,i}^2, x_{2,i}^2, \sqrt{2}x_{1,i}x_{2,i})$$

$$\varphi(\mathbf{X}_i) \cdot \varphi(\mathbf{X}_j) = x_{1,i}^2 x_{1,j}^2 + x_{2,i}^2 x_{2,j}^2 + 2x_{1,i}x_{2,i}x_{1,j}x_{2,j} = K(\mathbf{X}_i, \mathbf{X}_j)$$

$$\varphi(\mathbf{X}_i) \cdot \varphi(\mathbf{X}_j) = (x_{1,i}x_{1,j} + x_{2,i}x_{2,j})^2 = (\mathbf{X}_i \cdot \mathbf{X}_j)^2 = K(\mathbf{X}_i, \mathbf{X}_j)$$

起点

对于已知的空间变换 $\varphi(\mathbf{X}_i)$ → 优化计算过程，减少模型运算量

在模型里使用某个核函数 → 不知不觉间完成了数据的空间转换

- 如何知道一个函数是核函数？
- 如何选择核函数？

目录

ONE 核函数的定义

优化运算

TWO 常用的核函数

Gaussian、Laplacian、etc

THREE Scale variant

线性变换不稳定

常用的核函数

常用核函数

- 核函数的证明通常十分困难，需要用到Mercer定理
- 数学家已经找到了一些对大多数场景都适用的**常用核函数**

常用核函数

Linear kernel：$K(\boldsymbol{X}_i, \boldsymbol{X}_j) = \boldsymbol{X}_i \cdot \boldsymbol{X}_j$

没做任何空间变换，对应着最经典的线性支持向量学习机

Polynomial kernel：$K(\boldsymbol{X}_i, \boldsymbol{X}_j) = (gamma(\boldsymbol{X}_i \cdot \boldsymbol{X}_j) + coef0)^{degree}$

Sigmoid kernel：$K(\boldsymbol{X}_i, \boldsymbol{X}_j) = \tanh\,(gamma(\boldsymbol{X}_i \cdot \boldsymbol{X}_j) + coef0)$

Laplacian kernel：$K(\boldsymbol{X}_i, \boldsymbol{X}_j) = \exp\,(-gamma\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_1)$

RBF kernel：$K(\boldsymbol{X}_i, \boldsymbol{X}_j) = \exp\,(-gamma\|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2)$

将数据映射到无限维空间

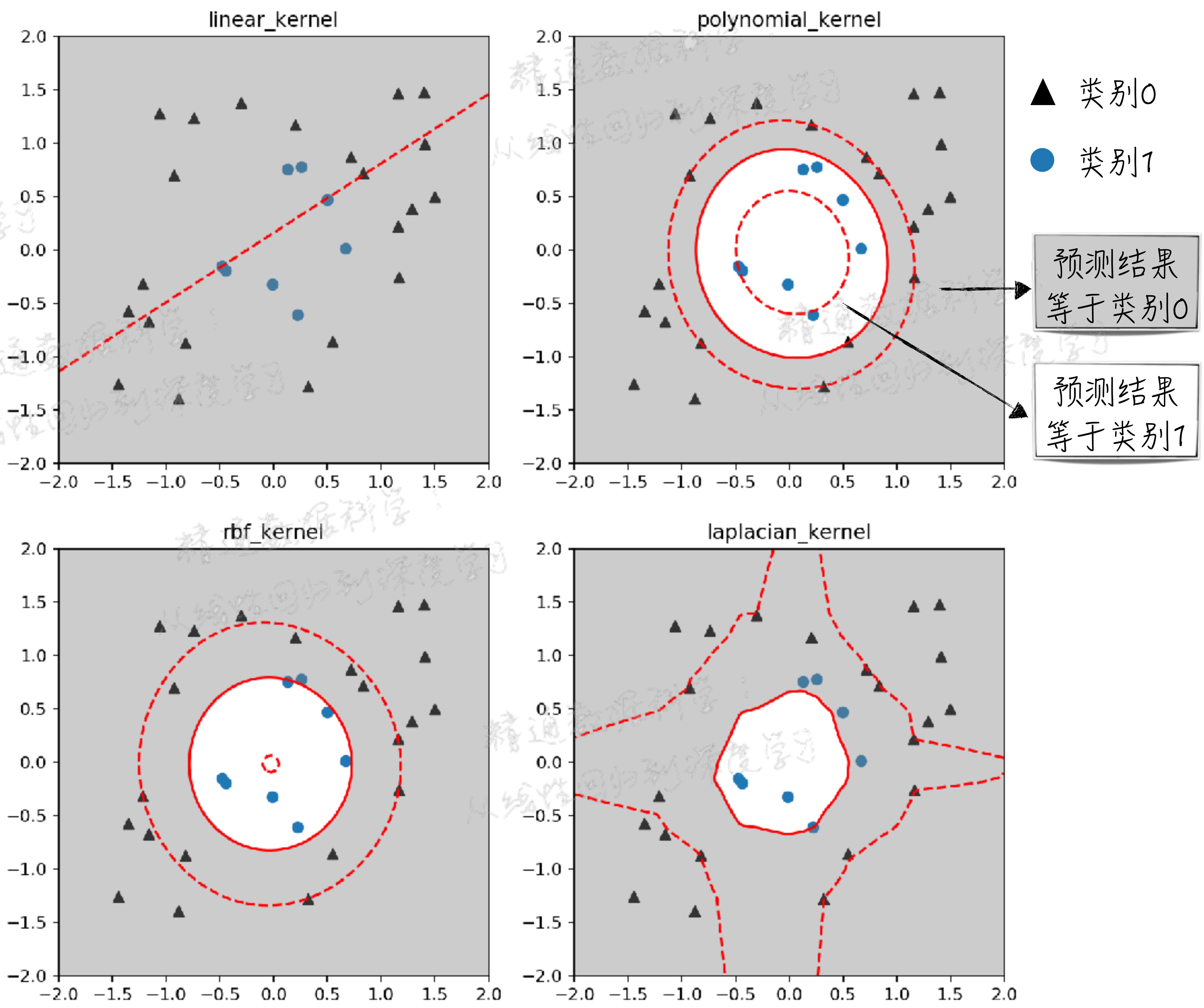
也称为Gaussian kernel，是魔力十足的核函数

常用的核函数

效果示例

对于同一数据使用不同的核函数进行建模：

- 多项式核函数、高斯核函数、拉普拉斯核函数的效果都很好
- 高斯核函数和拉普拉斯核函数将数据映射到无限维空间



目录

Scale variant

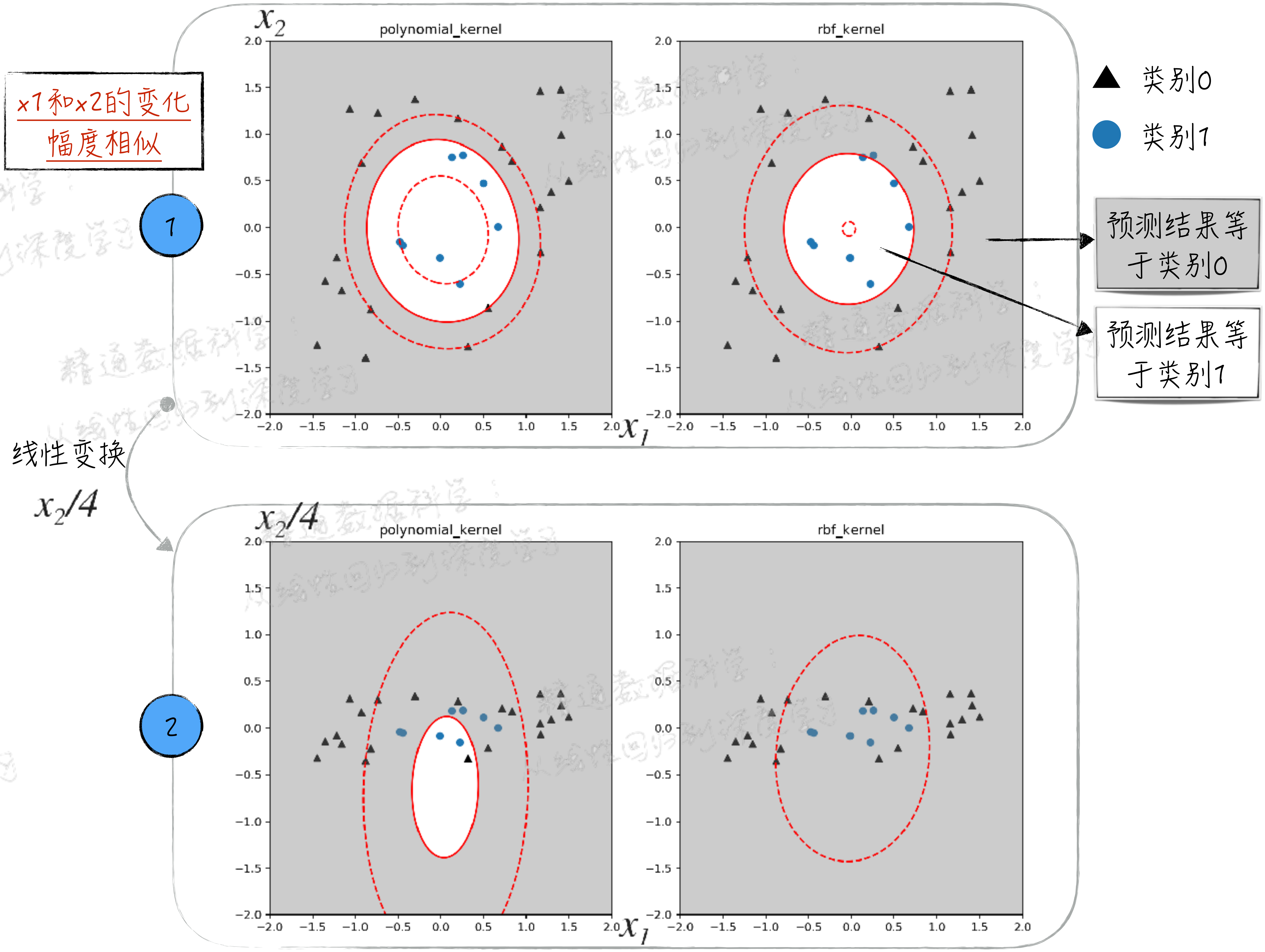
线性变换不稳定

之前讨论的线性回归和逻辑回归对**自变量的线性变换**是稳定的：

- 对自变量做线性变换（平移、伸缩）不影响模型结果；这被称为 Scale invariant

$$xx = ax + b$$

- 但SVM模型不是，它对**自变量的线性变换不稳定**



Scale variant

数学原理

Dual problem

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

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

原始数据：

$$\mathbf{X}_i = (x_{1,i}, x_{2,i}) \qquad \mathbf{X}_i\mathbf{X}_j = x_{1,i}x_{1,j} + x_{2,i}x_{2,j}$$

线性变换后：

变量的权重被扭曲，且模型无力修复

$$\mathbf{X}_i = (x_{1,i}, 10x_{2,i}) \qquad \mathbf{X}_i\mathbf{X}_j = x_{1,i}x_{1,j} + 100x_{2,i}x_{2,j}$$

应对方案：变量归一化（各变量的变化幅度相当）

THANK YOU