

# 朴素贝叶斯与其他模型的联结

随机森林 + 伯努利模型

小胖



# 目录

ONE 朴素贝叶斯模型回顾

伯努利模型

TWO 随机森林模型回顾

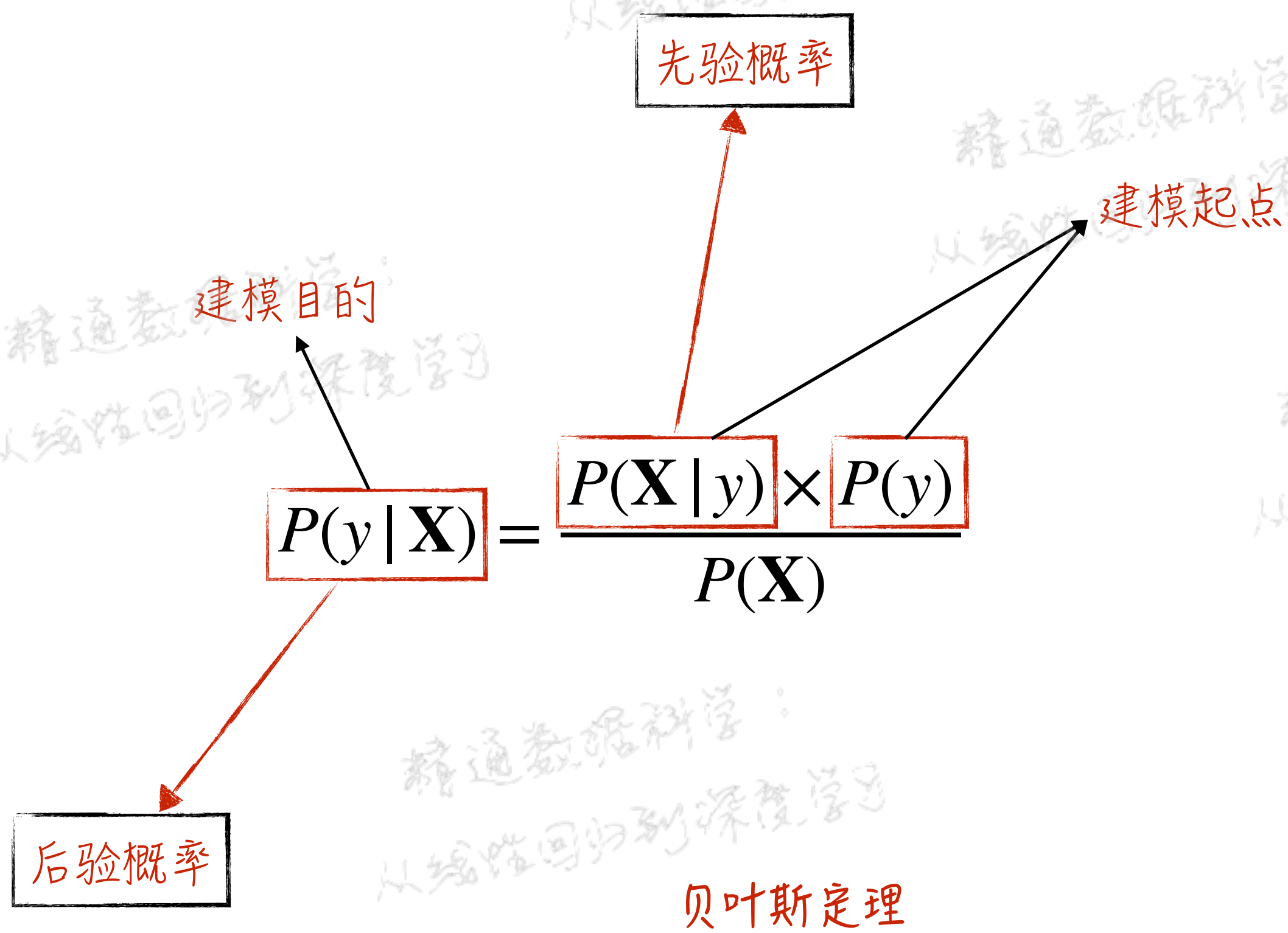
Random Forest Embedding

THREE 代码实现

scikit-learn

# 朴素贝叶斯模型回顾

## 伯努利模型



朴素贝叶斯模型的假设：

- 在给定类别下，各特征相互独立

$$P(\mathbf{X} | y) = P(x_1, x_2, \dots, x_n | y) = \prod_{i=1}^n P(x_i | y)$$

### 伯努利模型假设

× 只有两种取值  $x_i = 1 \quad or \quad x_i = 0$

× 取值的概率  
随类别的不同而不同

$$P(x_i = 1 | y) = p_{i,y}$$
$$P(x_i = 0 | y) = 1 - p_{i,y}$$

类别的分布  $P(y = l) = \theta_l$

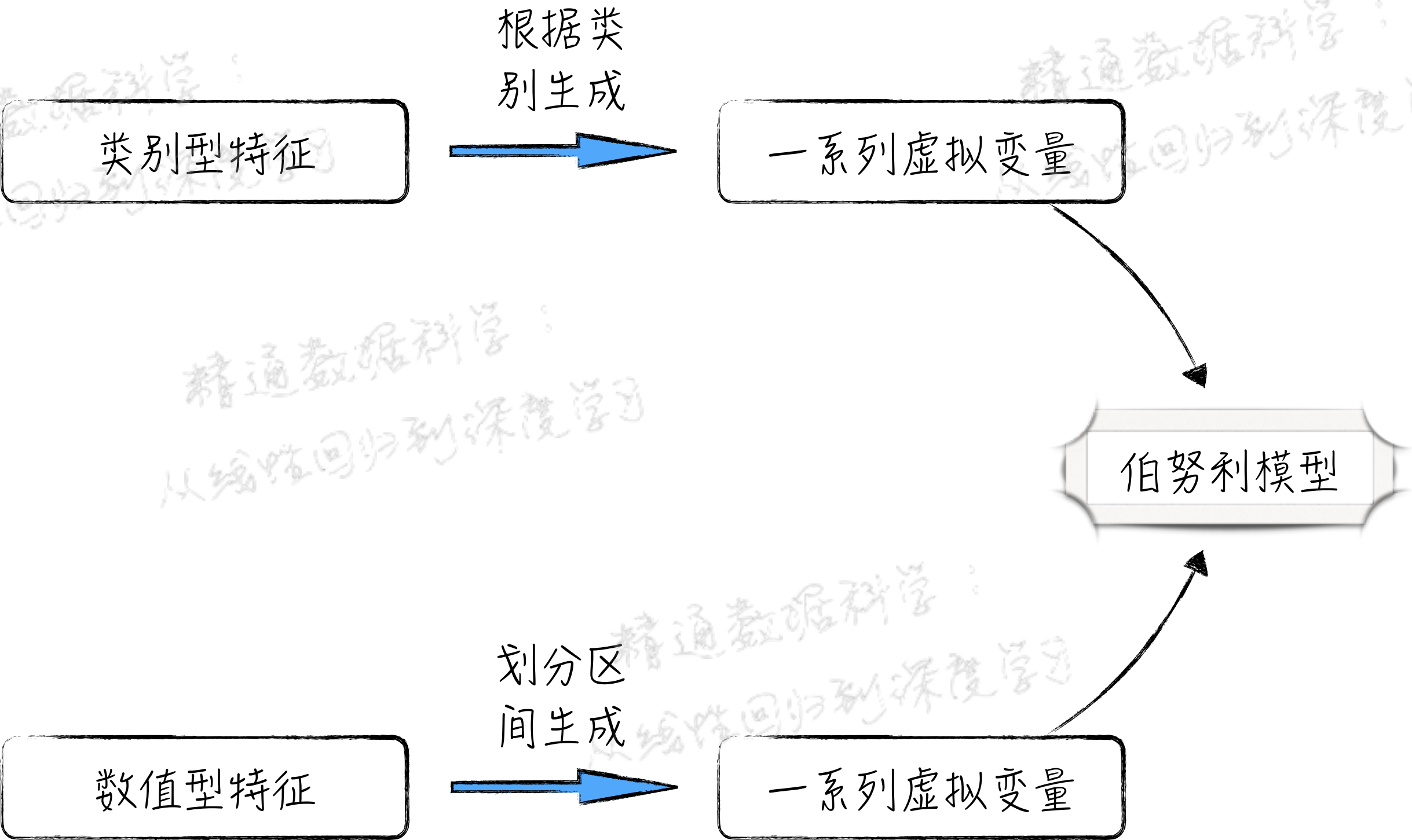
# 朴素贝叶斯模型回顾

伯努利模型

伯努利模型假设

x只有两种取值

$$x_i = 1 \quad or \quad x_i = 0$$



# 目录

## ONE 朴素贝叶斯模型回顾

伯努利模型

## TWO 随机森林模型回顾

Random Forest Embedding

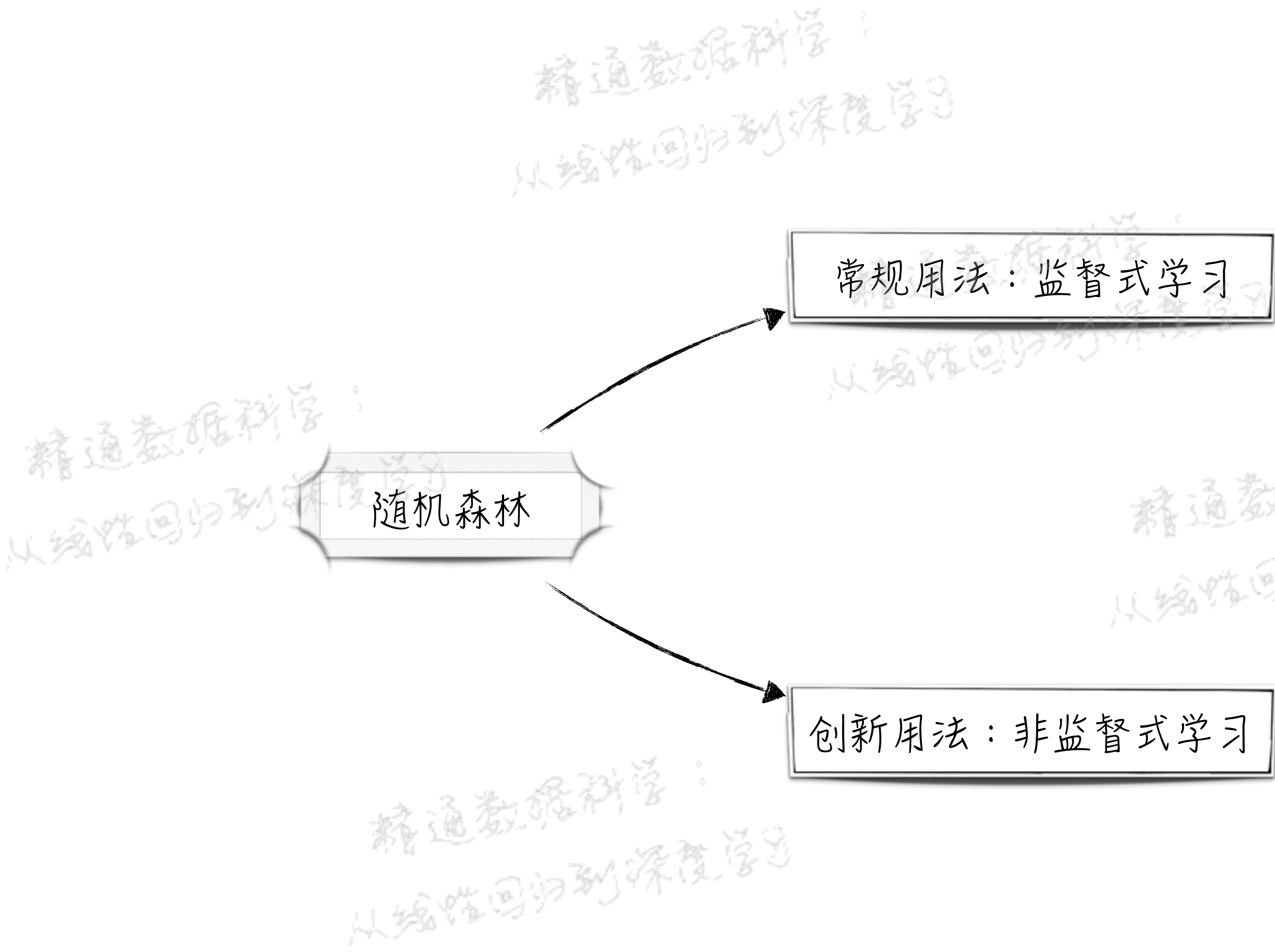
## THREE 代码实现

scikit-learn



# 随机森林模型回顾

随机森林的非监督学习



精通数据科学：  
从线性回归到深度学习

$(x_{11}, x_{12}, x_{13})$	$y_1$
$(x_{21}, x_{22}, x_{23})$	$y_2$
$(x_{31}, x_{32}, x_{33})$	$y_3$
$(x_{41}, x_{42}, x_{43})$	$y_4$

$(x_{11}, x_{12}, x_{13})$	$y_1$
$(x_{21}, x_{22}, x_{23})$	$y_2$
$(x_{31}, x_{32}, x_{33})$	$y_3$
$(x_{41}, x_{42}, x_{43})$	$y_4$

# 随机森林模型回顾

随机森林的非监督学习

生成数据

原始数据

$$\begin{aligned} X_1 &= (a_1, b_1) \\ X_2 &= (a_2, b_2) \\ X_3 &= (a_3, b_3) \end{aligned}$$

生成的y

类别0

合成数据

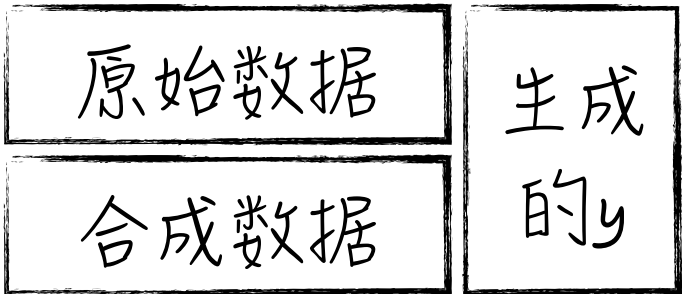
$$\begin{aligned} X_4 &= (a_1, b_3) \\ X_5 &= (a_3, b_2) \\ X_6 &= (a_2, b_1) \end{aligned}$$

类别1

生成的y

随机抽取

训练模型



随机森林

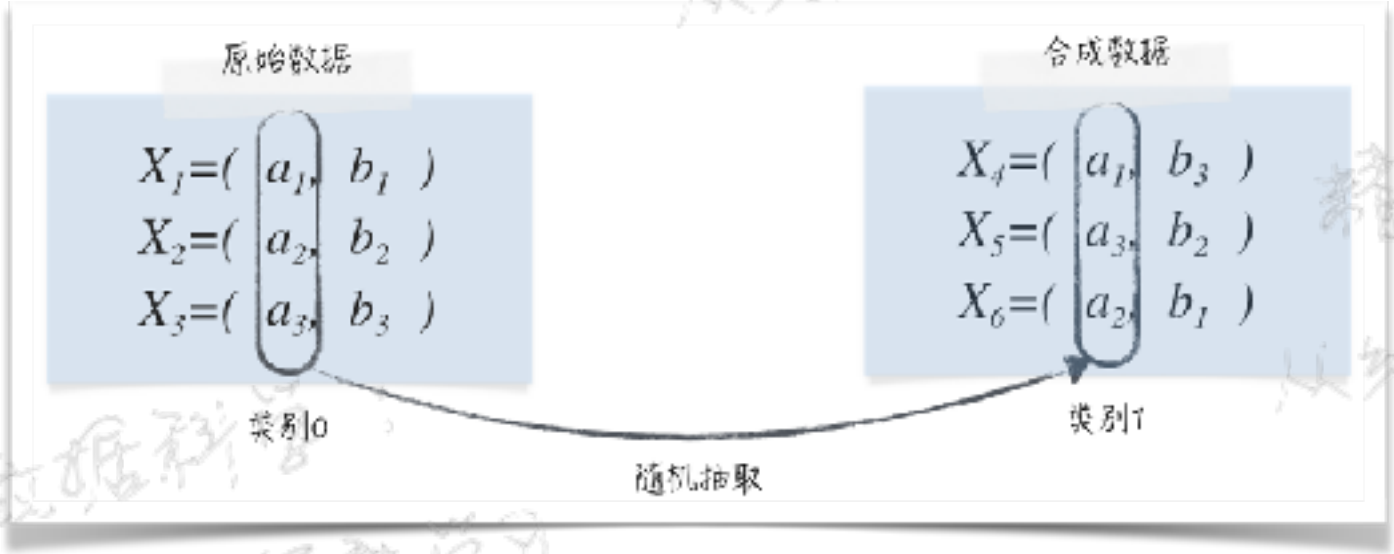
目的

- 分析数据变量之间的相关关系
- 将低维数据映射到高维空间（Random Forest Embedding），方便与其他模型联结

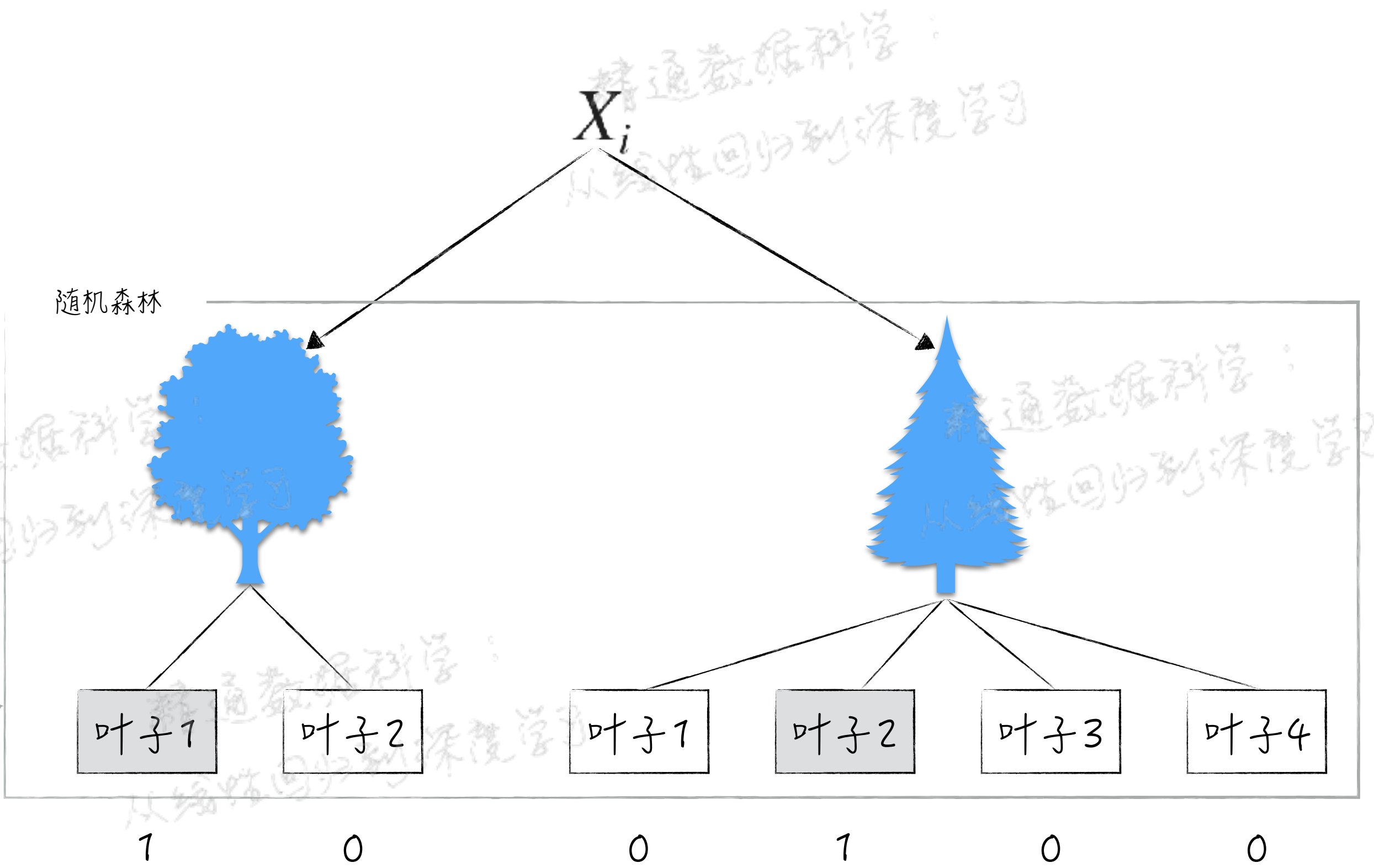
# 随机森林模型回顾

Random Forest Embedding

Random forest embedding



训练数据



其他模型：SVM、Logit、naive Bayes、etc



# 目录

ONE 朴素贝叶斯模型回顾

伯努利模型

TWO 随机森林模型回顾

Random Forest Embedding

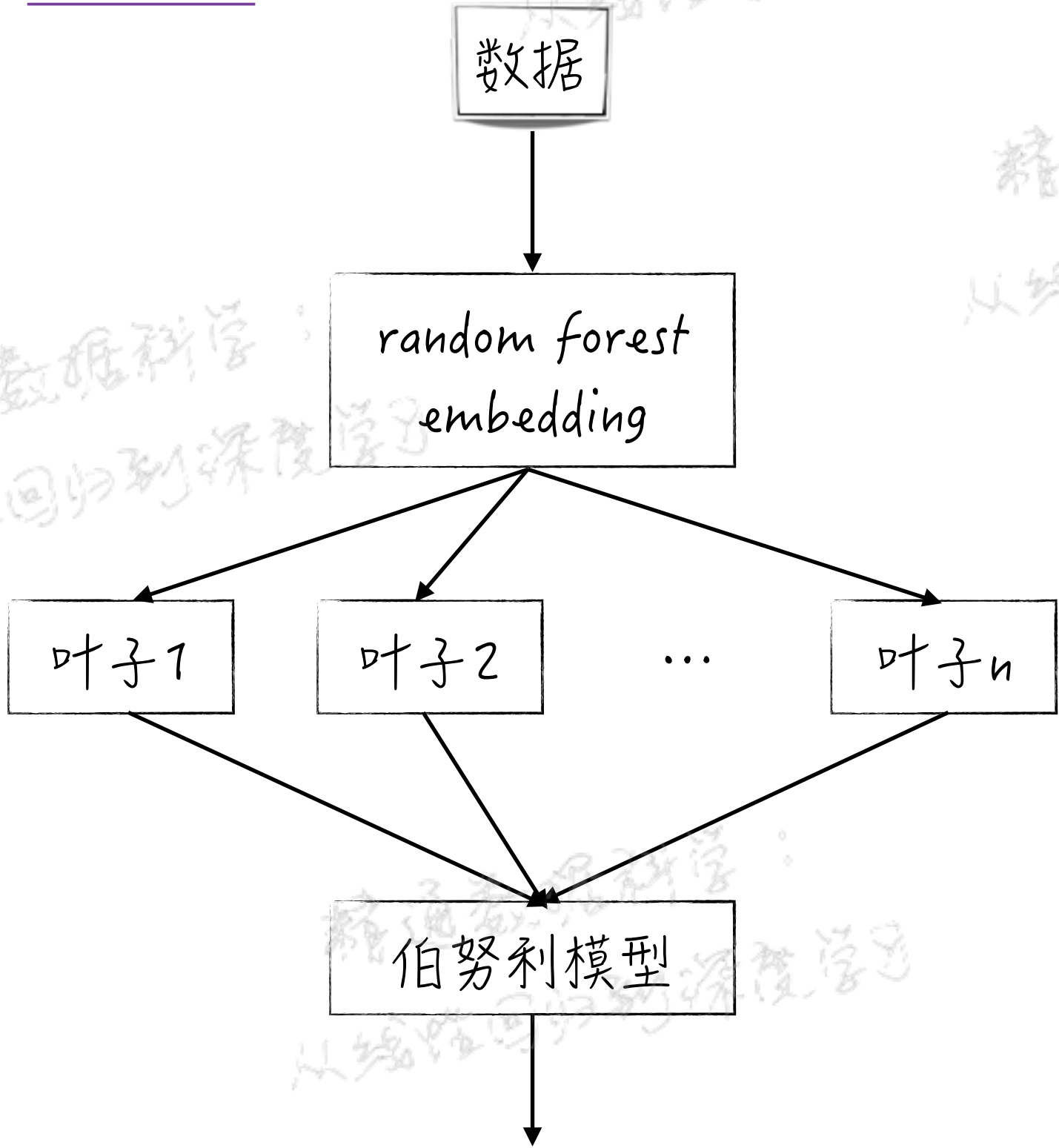
THREE 代码实现

scikit-learn

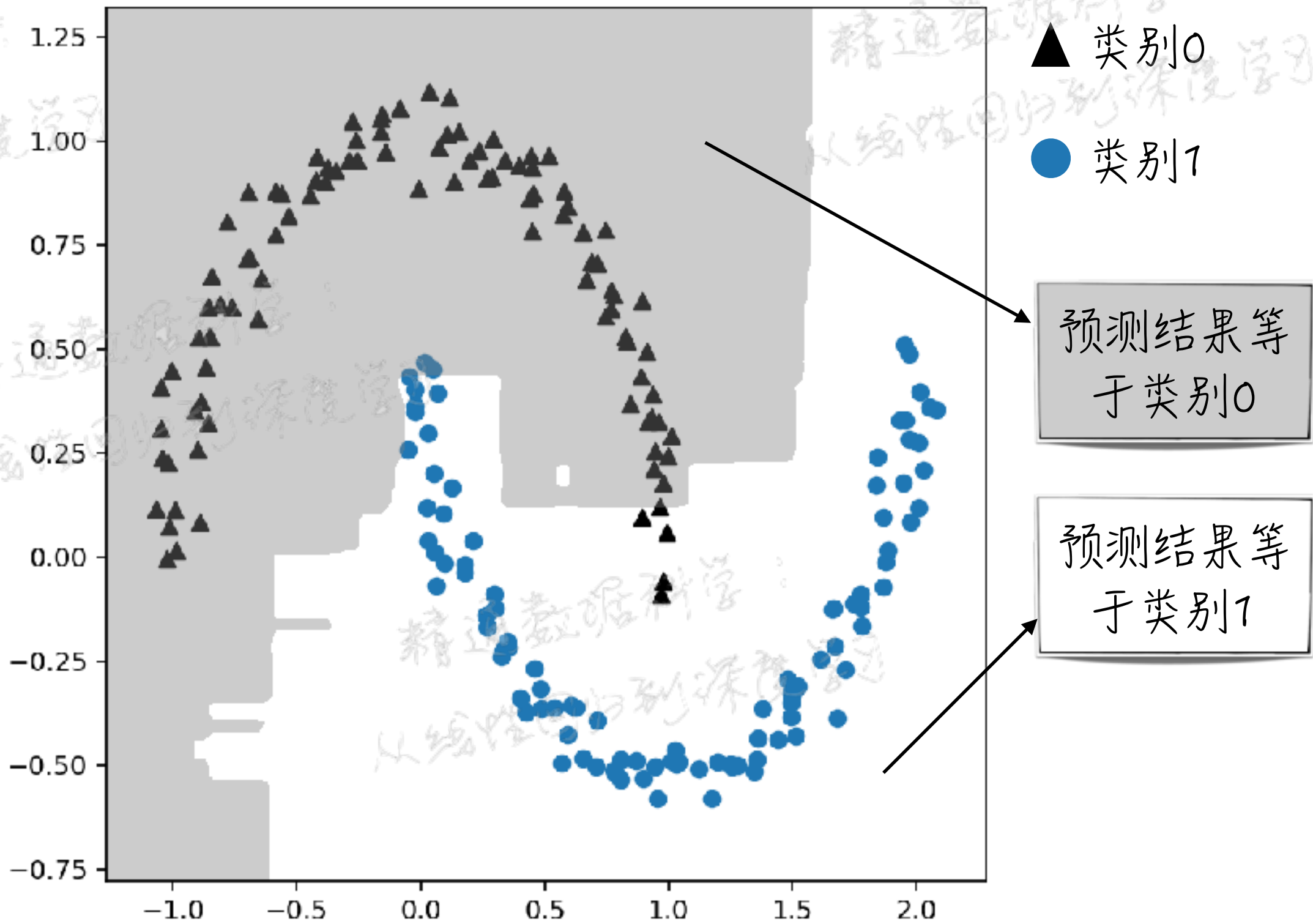
# 代码实现

模型架构与模型结果

模型架构



模型结果





# THANK YOU

精通数据挖掘科学：  
从线性回归到深度学习