

模型的联结

实际中，决策树的常用方式

小胖

目录

ONE 决策树的预测算法

决策树的优缺点

TWO 模型的联结

一个风控领域常用的方式

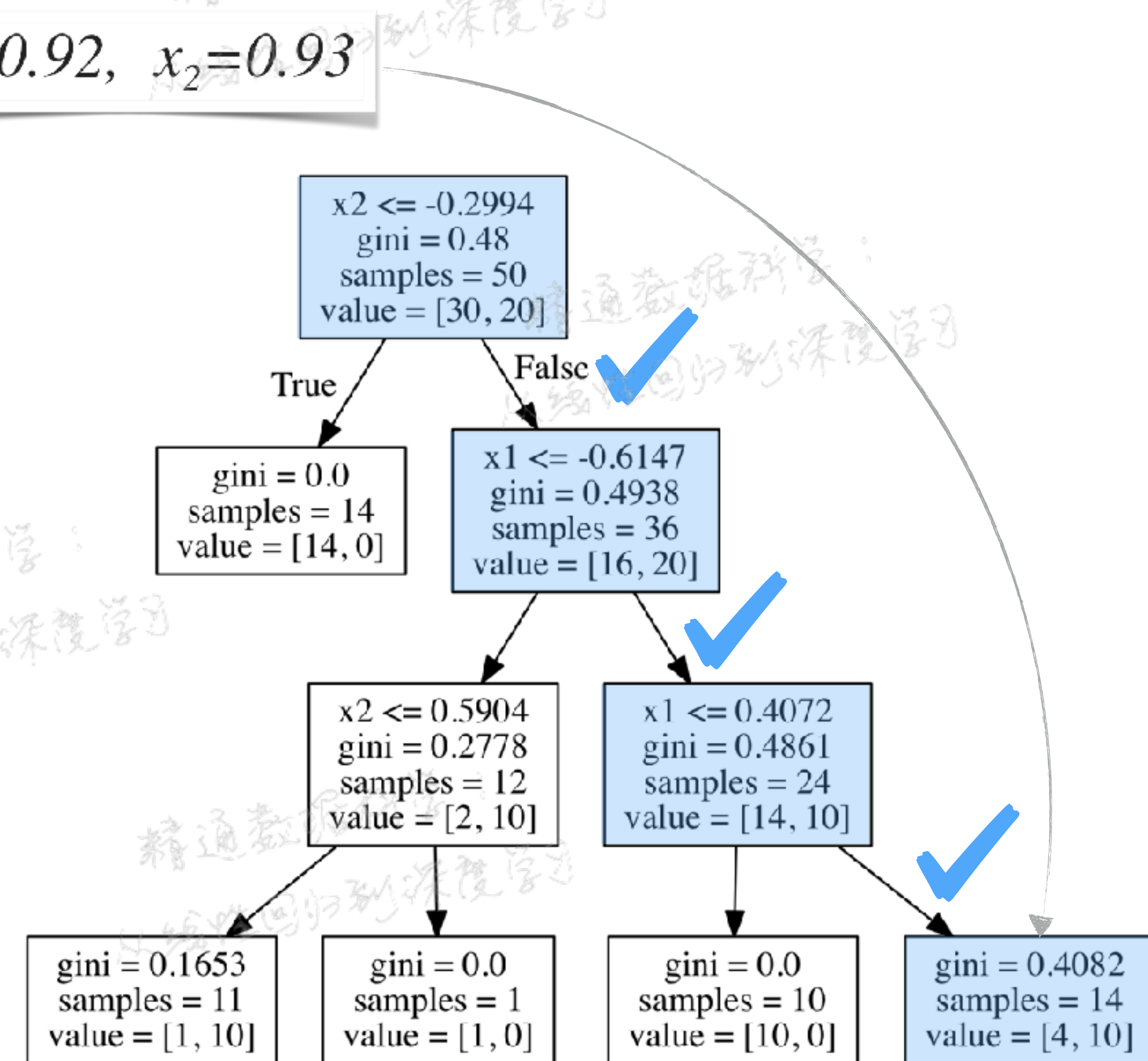
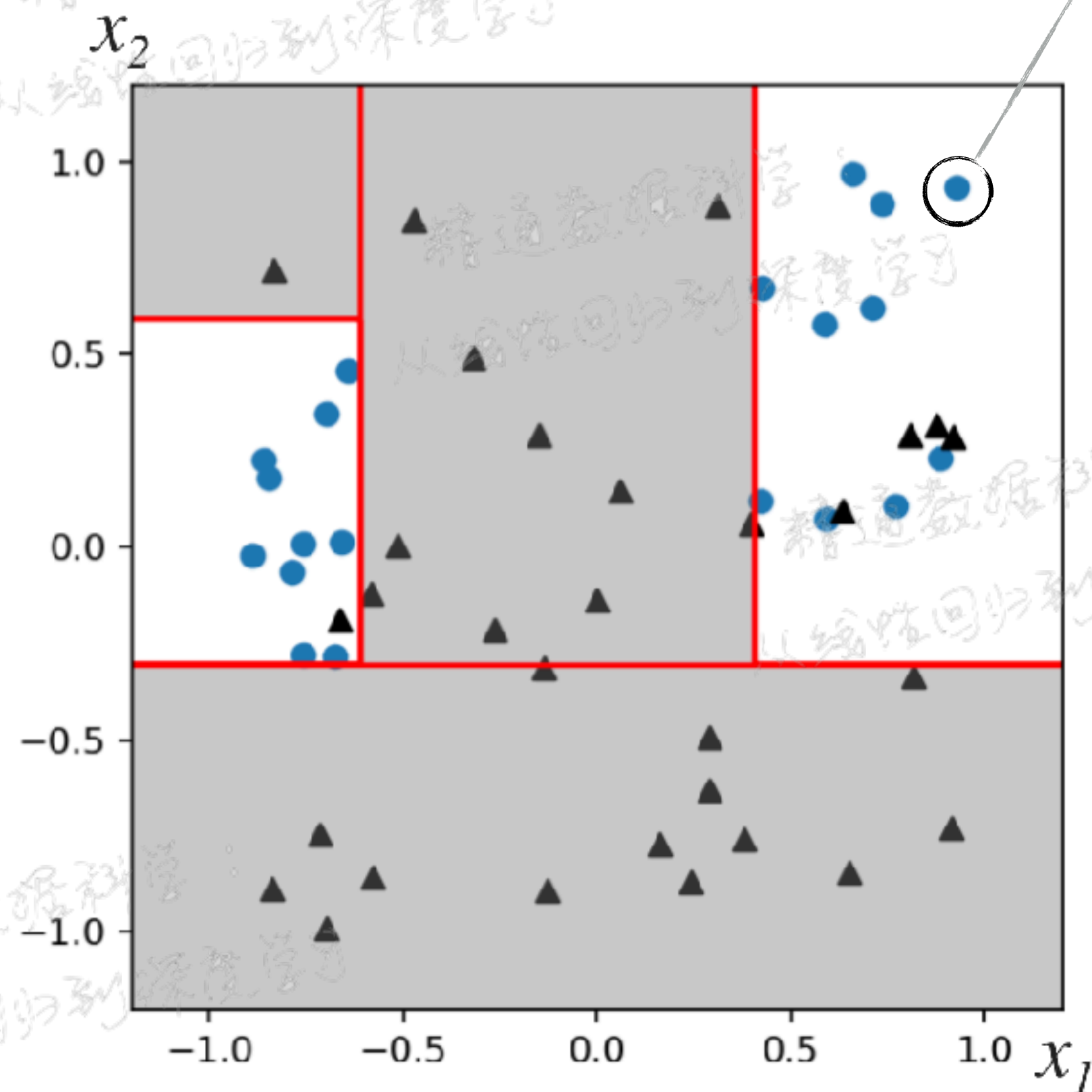
THREE 代码实现

Python

决策树的预测算法

决策树的预测算法

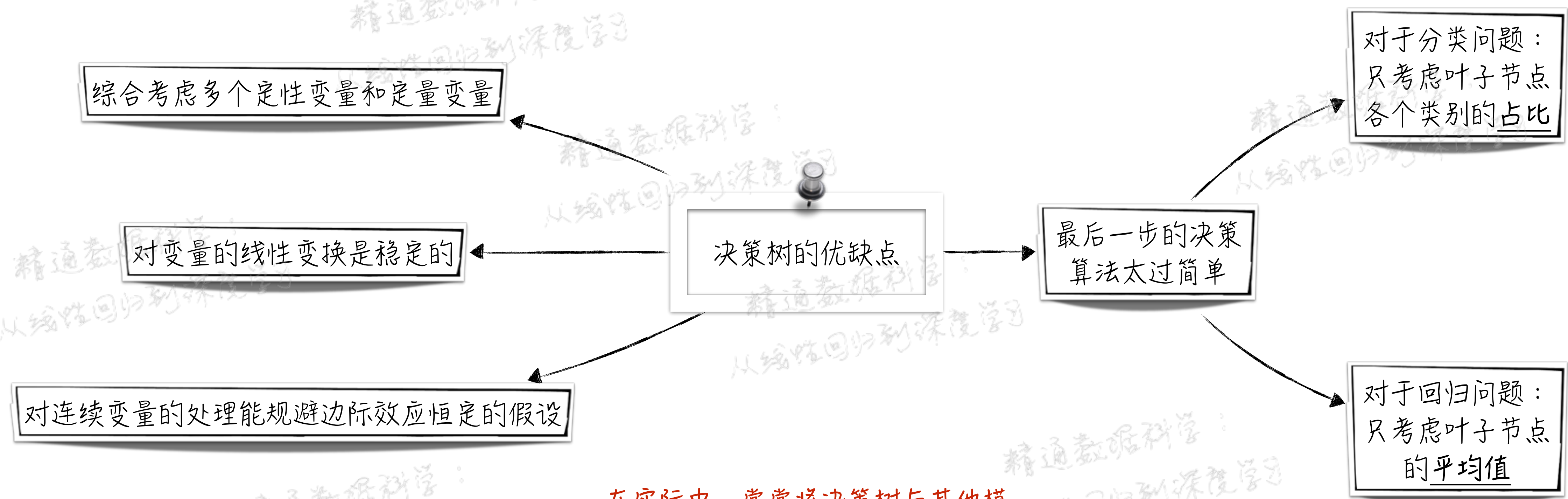
- 根据决策规则，找到决策路径
- 根据叶子节点类别占比，决定预测结果



$$P(y=1) = \frac{10}{14}, \quad P(y=0) = \frac{4}{14}$$

决策树的预测算法

决策树的优缺点



在实际中，常常将决策树与其他模型联结起来使用，扬长避短

目录

ONE 决策树的预测算法

决策树的优缺点

TWO 模型的联结

一个风控领域常用的方式

THREE 代码实现

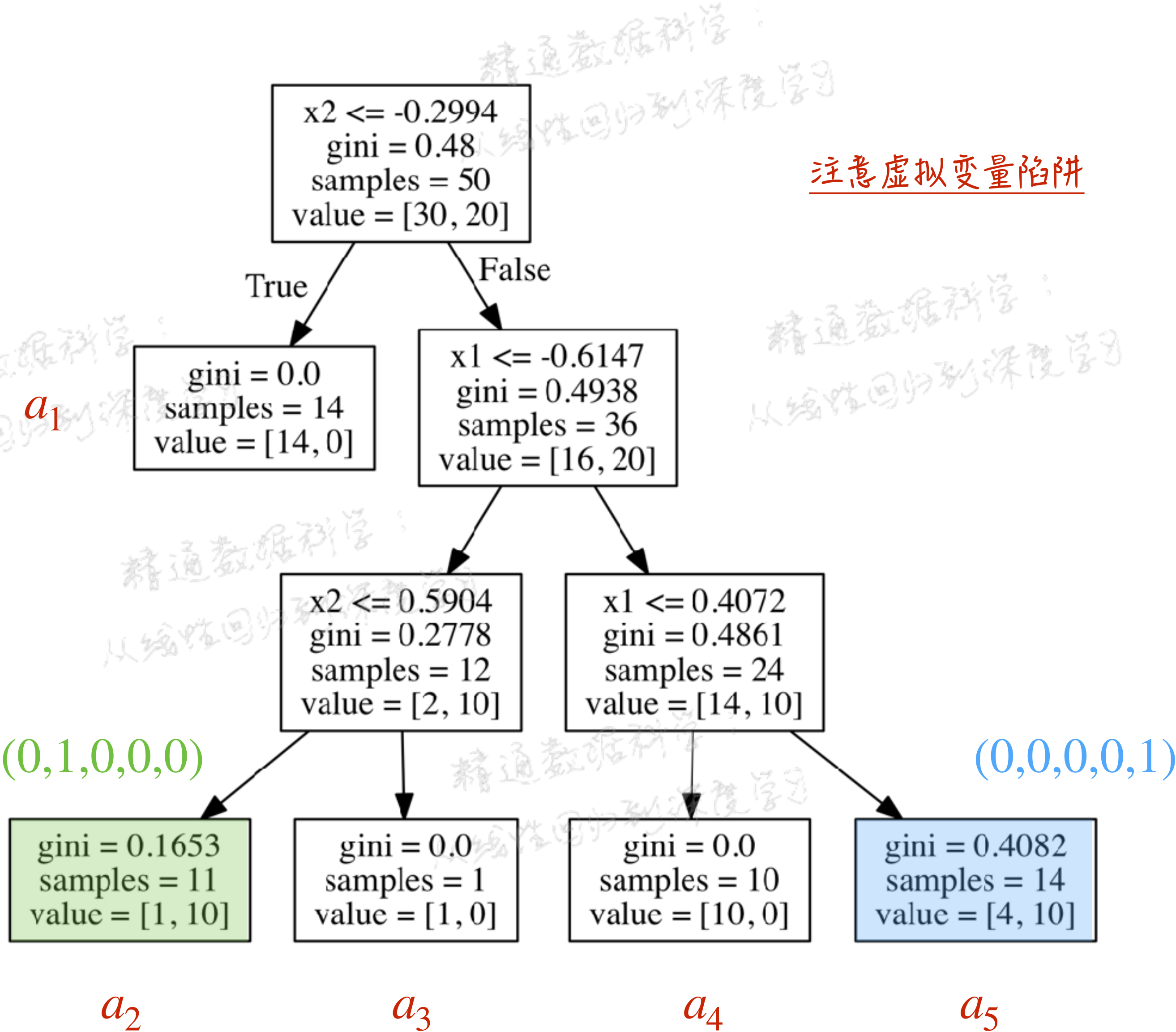
Python

模型的联结

一个风控领域的例子

并不直接使用决策树的预测结果，而是利用数据所在叶子节点带来的信息：

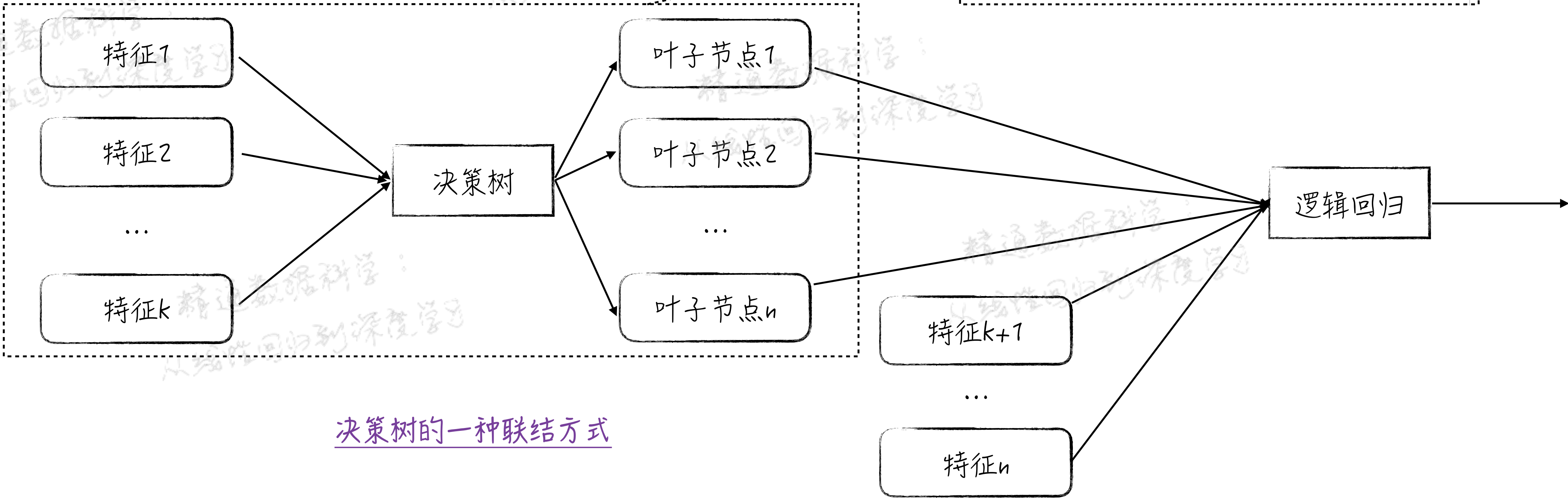
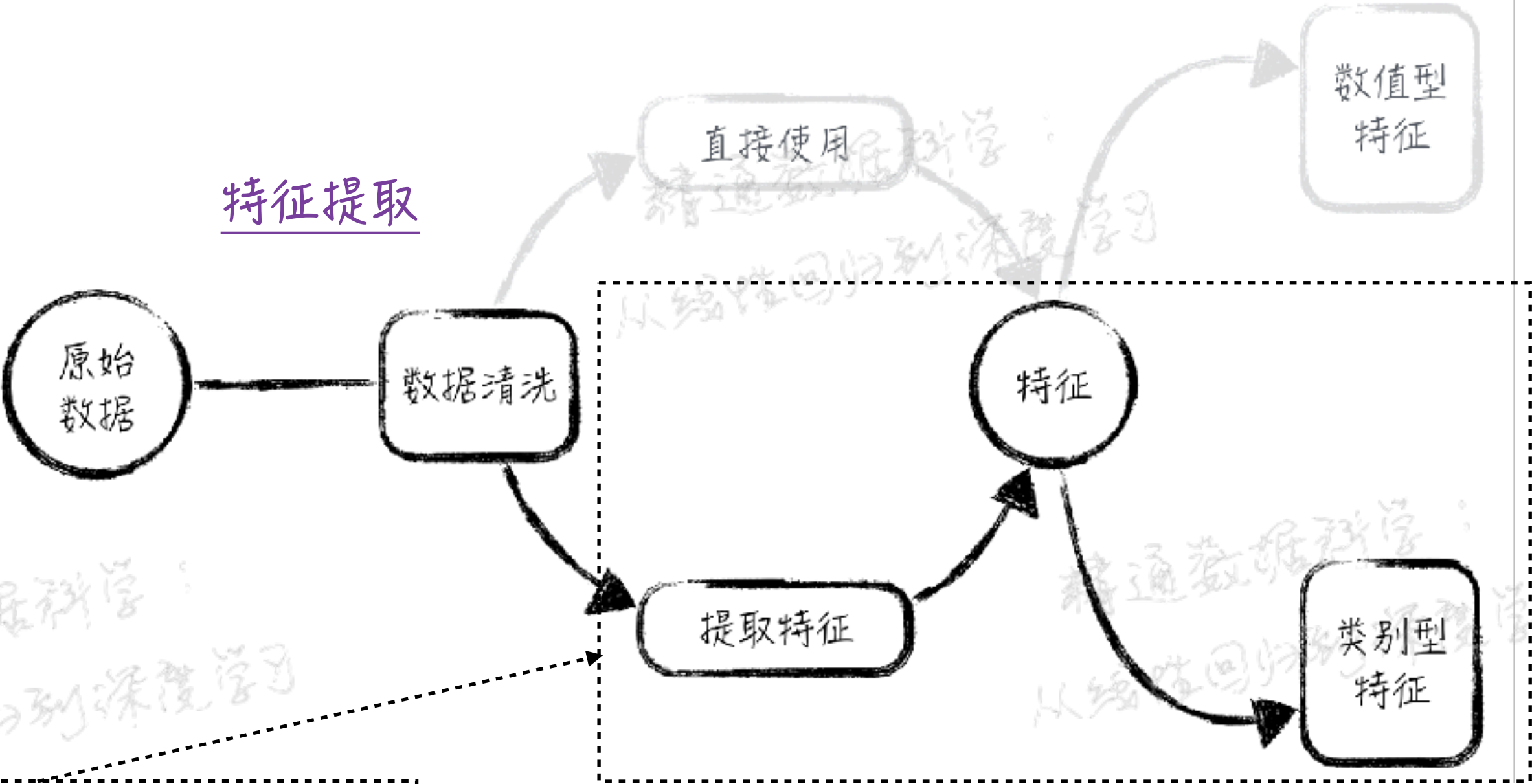
- 使用一系列的虚拟变量（dummy variable）来表示各个叶子节点
- 根据数据所在叶子节点，得到数据的新特征
- 新特征和原有特征可以混合在一起进行建模



模型的联结

一个风控领域的例子

特征提取



决策树的一种联结方式

目录

ONE 决策树的预测算法

决策树的优缺点

TWO 模型的联结

一个风控领域常用的方式

THREE 代码实现

Python

代码实现

DT + Logit

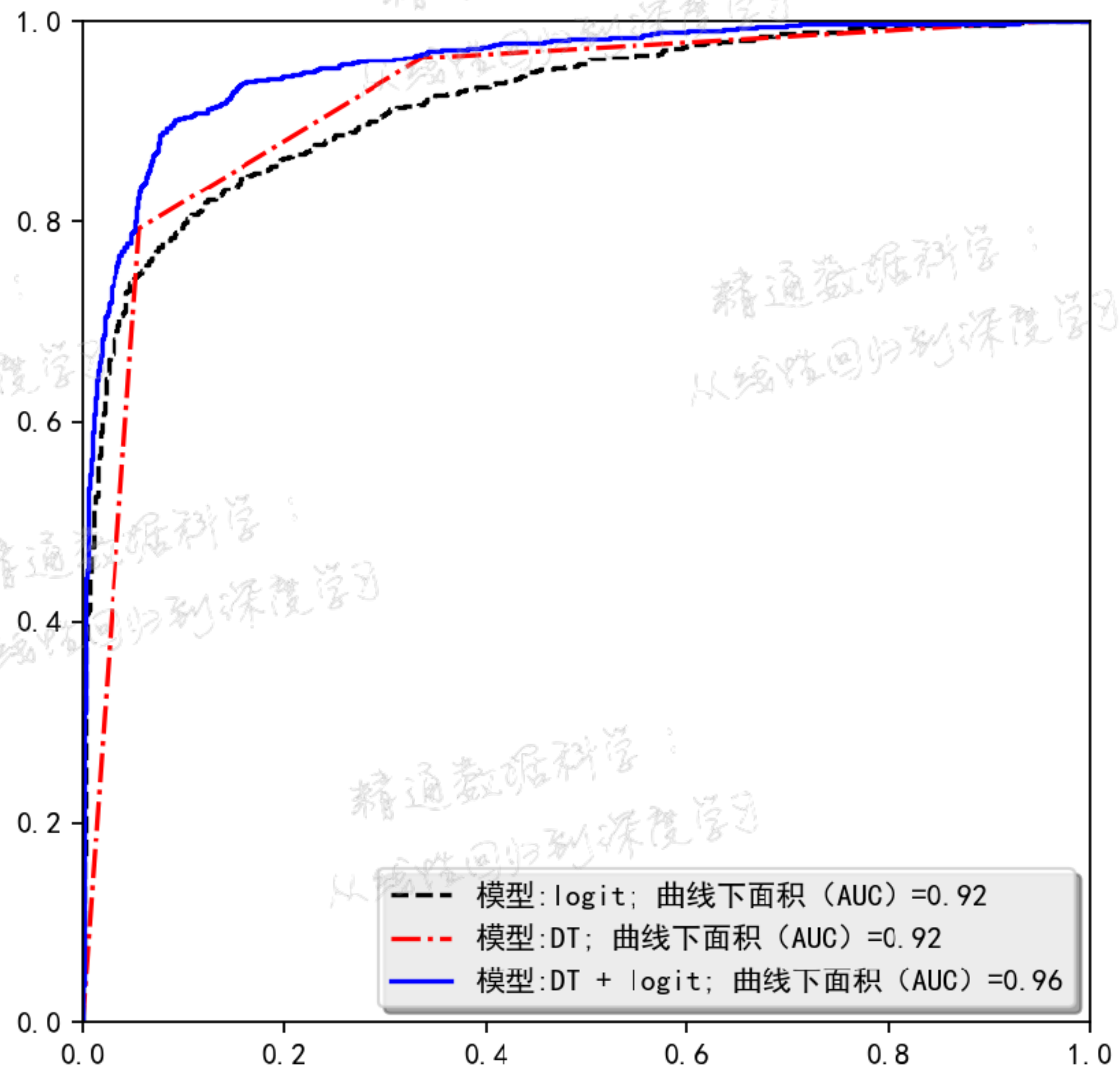
将数据分为训练集和测试集

将训练集进一步划分为决策树训练集
和逻辑回归训练集

使用训练好的决策树对数据进行转换，
得到新特征

使用新特征和旧特征搭建逻辑回归模型

3种模型效果比较



THANK YOU

精通数据挖掘科学：
从线性回归到深度学习