# 概述

决策树呈树形结构，是一种基本的回归和分类方法。决策树模型的优点在于可读性强、分类速度快。

训练数据集用于训练决策树模型，然后将已经训练好的模型运用在预测数据集上，给出预测结果。

# 算法

主要包括两类算法：

（1）Quinlan系列：ID3算法、C4.5算法、C5.0算法

（2）CART决策树

上述一系列算法的步骤总体可以概括为建树与剪树：

在建树步骤中，首先选择最有解释力度的变量，接着对于每个变量选择最优的分割点进行建树。在剪树方法中，分为前剪枝和后剪枝。前剪枝用于控制树的生成规模，后剪枝用于删除没有意义的分组。

早期决策树算法前剪枝和后剪枝都会使用，而Python的scikit-learn包中统一使用交叉验证的算法对所有分类模型进行超参数的筛选，因此不需要再使用后剪枝算法。

## ID3生成算法

（1）信息增益

（2）建树原理

（3）分类规则

## C4.5生成算法

# 剪枝算法

# CART树

# 连续值、缺失值处理

# 多变量决策树

# 特点

# 应用