

机器学习实验报告

决策树

朱天泽

(日期: 2022 年 3 月 28 日)

摘要 在《机器学习》第四章中,我学习了决策树模型。在此次实验中,我尝试用 DFS 和 BFS 实现了多个用于分类任务的决策树模型,包括针对离散值的单变量决策树、基于对数几率回归的多变量决策树。在训练好决策树模型的基础上,我使用了 4 个 UCI 数据集,比较了未剪枝、预剪枝、后剪枝三种类型决策树的准确率。此外,我尝试使用了机器学习库 scikit-learn,同自己实现的算法进行了比较。

关键词 决策树; 剪枝; 分类; 搜索; scikit-learn

1 习题 1

1.1 编程题目理解

题目要求:任意选择 4 个 UCI 数据集,对基于信息增益划分选择 (ID3)、基于基尼指数划分选择 (CART),基于对率回归划分选择的决策树算法 (包括未剪枝、预剪枝、后剪枝三种) 进行实验比较。

ID3、CART 两种划分选择,可用于实现单变量的决策树;对率回归是一个线性分类模型,基于对率回归,可以实现多变量决策树。

1.2 决策树模型原理阐述

1.3 决策树模型设计思路

1.4 核心代码详解

1.5 实验结果分析

1.6 学习收获

1.7 参考资料

•

2 习题 2

- 2.1 编程题目理解
- 2.2 交叉验证法阐述
- 2.3 核心代码分析
- 2.4 实验结果分析
- 2.5 学习收获
- 2.6 参考资料

•

3 习题 3

- 3.1 编程题目理解
 - 3.1.1 原理阐述
- 3.2 设计思路
- 3.3 核心代码详解
 - 3.3.1 数据集格式
- 3.4 实验结果分析
- 3.5 学习收获
- 3.6 参考资料

•