

常见面试问题

决策树、随机森林、GBDT

小胖

目录

ONE 决策树

TWO 随机森林

THREE GBDT

决策树

面试题目：简述决策树模型的算法步骤，以及决策树有哪些常用的启发函数（划分评判标准）？

精通数据科学：
从线性回归到深度学习

难度：★★★★☆

解答要点：

- 决策树模型进行模型训练和预测的步骤
- 决策树节点不纯度定义：Gini、Entropy、Misclassification
- 决策树如何选定划分特征和划分标准：ID3、C4.5、CART

涉及课程：

- 9_1 什么是决策树：条理清晰地做决定

决策树

面试题目：如何对决策树进行剪枝？

精通数据科学：
从线性回归到深度学习

难度：★★★★☆

解答要点：

- 决策树进行剪枝的原因和理论基础
- 决策树的剪枝方法：前剪枝、后剪枝
- 前剪枝的算法思路以及实现细节、常用的模型超参数
- 后剪枝的算法思路以及步骤

涉及课程：

- 9_1 什么是决策树：条理清晰地做决定

目录

ONE 决策树

TWO 随机森林

THREE GBDT

随机森林

面试题目：集成学习分哪几种？它们有何异同？

精通数据科学：
从线性回归到深度学习

难度：★★★★☆

解答要点：

- 集成学习的基本概念：Boosting、Bagging和基分类器
- Boosting和Bagging两种方法提升模型效果的理论基础、步骤以及需要特别注意的假设
- 在树模型领域，这两种方法的代表模型

涉及课程：

- 9_3 树的集成：随机森林
- 9_4 树的集成：GBDT

随机森林

面试题目：可否将随机森林中的基分类器（决策树模型）替换成为线性分类器？请解释为什么？

精通数据科学：
从线性回归到深度学习

难度：★★★★☆

解答要点：

- 随机森林提升模型效果的关键假设和数学原理
- 决策树模型的稳定性和产生随机性的方法
- 线性分类器的稳定性和产生随机性的方法

涉及课程：

- 5_1 如何解决二元分类问题
- 9_1 什么是决策树：条理清晰地做决定
- 9_3 树的集成：随机森林

精通数据科学：
从线性回归到深度学习

精通数据科学：
从线性回归到深度学习

精通数据科学：
从线性回归到深度学习

精通数据科学：
从线性回归到深度学习

目录

ONE 决策树

TWO 随机森林

THREE GBDT

GBDT

面试题目：GBDT的基本原理是什么？模型的算法步骤是如何的？

精通数据科学：
从线性回归到深度学习

难度：★★★★☆

解答要点：

- 梯度提升的数学原理
- 梯度提升和梯度下降的区别和联系
- GBDT模型的训练步骤

涉及课程：

- 9_4 树的集成：GBDT

GBDT

面试题目：GBDT和XGBoost的联系和区别有哪些？

精通数据科学：
从线性回归到深度学习

难度：★★★★☆

解答要点：

- GBDT的模型步骤和模型的优缺点
- XGBoost的模型步骤
- XGBoost相对于GBDT的改进点

涉及课程：

- 9_4 树的集成：GBDT

THANK YOU

精通数据挖掘科学：
从线性回归到深度学习