## 第四讲2（分类）

1. 构造决策树采用的基本方法是（），即它会在每个节点选择当前状态下分类效果最好的属性进行分类

A 贪心

B 动态规划

C 回溯法

D 分支界限

1. 以下关于ID3算法局限性或者特性的说法，错误的选项是（）

A没有考虑连续特征

B 采用信息增益大的特征优先建立决策树的节点

C 没有考虑缺失值的情况

D考虑了过拟合的问题

1. CART分类条件是按照（）最小的原则

A 信息增益

B 增益率

C 基尼系数

1. 一般情况，CART需要计算χ2值来判断分类条件和类别的相关程度，决定是否停止分裂 （）

A 对

B 错

1. 朴素贝叶斯分类基于一种假设：给定样本属性值有条件地相互独立 （）

A 对

B 错

1. Bagging算法和Boosting算法是常见的组合方法 （）

A 正确

B 错误

1. Adaboost算法对于分类错误的元组，增加其权重。（）

A 正确

B 错误

1. 在分类问题中，ROC曲线离对角线越近，模型的准确率越低 （）

A 对

B 错

1. 以下哪一项关于决策树的说法是错误的 ()

A. 冗余属性不会对决策树的准确率造成不利的影响

B. 子树可能在决策树中重复多次

C. 决策树算法对于噪声的干扰非常敏感

D. 寻找最佳决策树是NP完全问题

1. 下列错误的是 （ ）

A 随机森林运用了基于bagging的算法

B 随机森林在构建决策树的过程中不需要剪枝

C 随机森林算法对每次拆分时选择的属性敏感

11.术语“假正（false positive）”的含义是（）

A 分类器正确标记的正元组

B 分类器正确标记的负元组

C 错误标记的负元组

D 错误标记的正元组