

# 模式识别导论第六次作业——简述题

杨登天-202028015926089-微电子研究所

1. ID3、C4.5 和 CART 是三种典型的决策树分类算法，请分别说明这三种算法具体使用的决策特征选择方法（注：CART 分别说明分类和回归两种情况）。

**解答：**

ID3 算法：利用信息增益最大的准则进行特征选择。

C4.5 算法：结合信息增益，利用增益率最大的准则进行特征选择。

CART 分类树：利用基尼指数最小的准则进行特征选择。

CART 回归树：利用最小平方误差的准则进行特征选择。

**解答完毕**

2. 简述构造决策树的一般技术路线。

**解答：**

步骤一、构建根结点，将所有训练数据都放在根结点。

步骤二、选择一个“最优”特征，按照它的离散取值，将上一节点中数据分割成不同子集。

步骤三、对于分割的子集，如果能够基本正确分类（子集中大部分数据都属于同一类），那么构建叶结点。

步骤四、对于不能构建叶结点的子集，继续 2-3 步，直到所有子集都能构建叶结点。

**解答完毕**

3. 简述决策树的特点。

**解答：**

特点 1、分而治之的基本策略

特点 2、以树形结构，根据事先定好的“判定问题”，进行（分类）决策

特点 3、叶结点对应决策结果

特点 4、内部结点对应属性测试（回答“判定问题”）

**解答完毕**

4. 简述决策树剪枝的目的，以及两种常用的判断决策树是否需要剪枝的准则。

**解答：**

问题一、简述决策树剪枝的目的

基于无剪枝决策树容易过拟合的问题——学习过程过多的考虑如何提高分类正确率，忽略了推广性，构建出过于复杂的决策树；对训练数据很准确；对未知测试数据的分类性能很差。

因此需要通过剪枝来降低决策树复杂度，在模型复杂度和对训练数据的拟合之间进行平衡。

问题二、两种常用的判断决策树是否需要剪枝的准则

准则一、根据决策树泛化性能是否有提升来确定是否剪枝；

准则二、根据整体代价是否会减少来确定是否剪枝。

**解答完毕**

5. 决策树的剪枝策略一般有哪些？各有什么优缺点？

**解答：** 一般有预剪枝和后剪枝两种。

预剪枝：在决策树生成的过程进行剪枝，只有泛化性能提升的结点划分才被采纳。其优点是计算速度快，其缺点是有欠拟合的风险。

后剪枝：先生成决策树，再自底向上对所有结点进行检验，判断剪枝前后是否会带来泛化性能的提升，以此决定是否剪枝。其优点是欠拟合风险低，泛化能力强，其缺点是仍存在欠拟合的风险，计算量大。

**解答完毕**