

## 第四次作业

---

林之阳 3118311071

### 1. 选择题

1. D
2. A
3. A B C
4. B
5. B C D

### 2. 简答题

#### 1. 参数方法

1. 通过首先给出目标函数，待定参数，可以极大的简化学习过程

2. 优点：

- 简单：这些算法很容易理解和解释结果
- 快速：参数模型可以很快从数据中学习
- 训练集可以不多：它们不需要太多的训练数据，甚至可以很好地拟合有缺陷的数据

3. 缺点

- 约束：算法选择一种函数形式只有参数可变，这限制模型本身
- 复杂度低：算法难以解决关系复杂的问题

#### 2. 决策树

1. 通过将信息增益由大到小的属性分支依次分布，完成对位置数据的分类判断

2. 优点：

- 决策树易于理解和实现. 人们在通过解释后都有能力去理解决策树所表达的意义
- 每一次预测的最大计算次数不超过决策树的深度

3. 缺点：

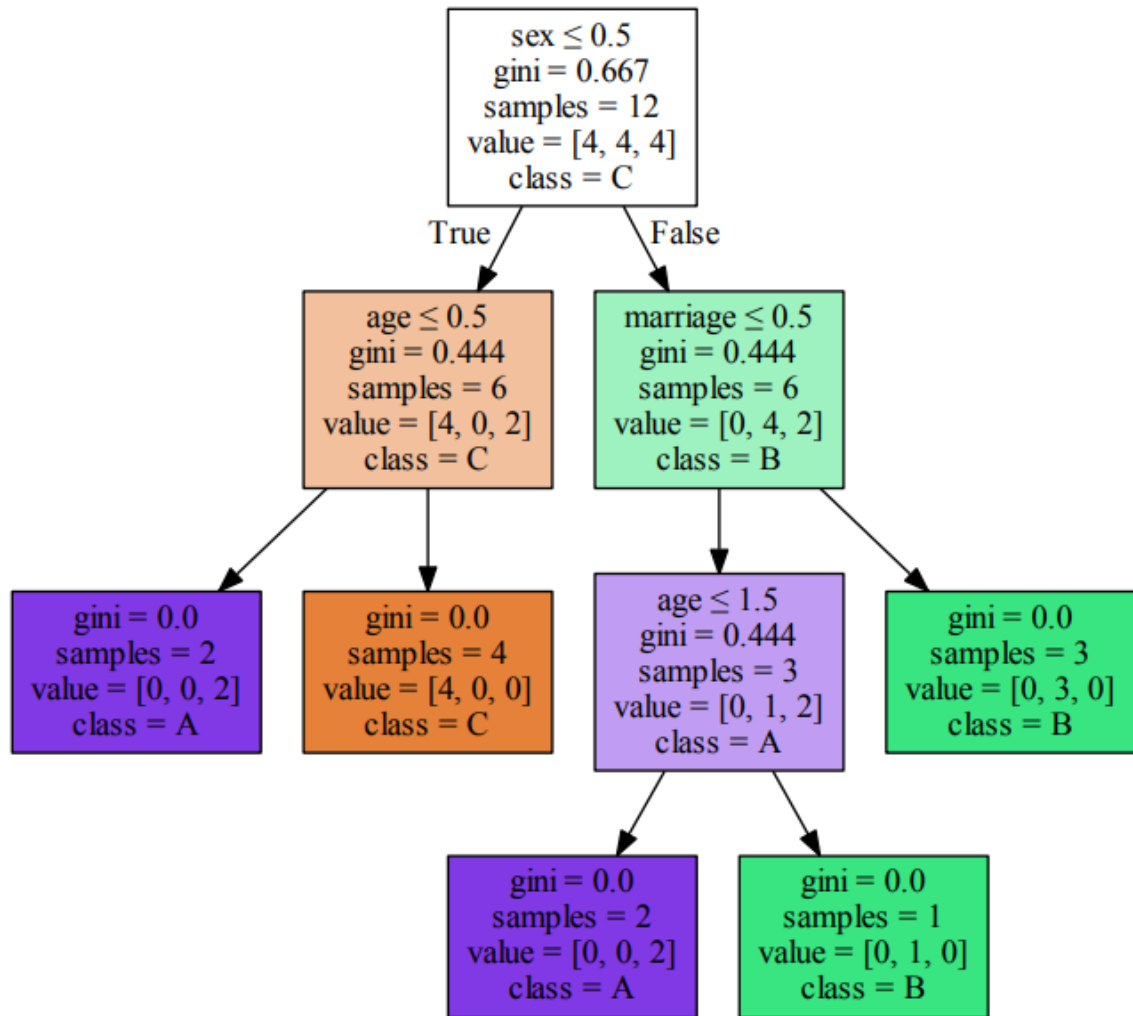
- 对特征关联性强的数据不敏感
- 可能造成过拟合

#### 3. 计算题

1. 性别信息增益：

$$\begin{aligned} Gain(D, sex) &= Info(D) - \left( \frac{D^{boy}}{D} \times Info(sex_{boy}) + \frac{D^{girl}}{D} \times Info(sex_{girl}) \right) \\ &= -(P_A \log_2 P_A + P_B \log_2 P_B + P_C \log_2 P_C) - (P_{boy}(P(B|boy) \log_2 P(B|boy) + \dots \\ &\quad P(C|boy) \log_2 P(C|boy)) + P_{girl}(P(A|girl) \log_2 P(A|girl) + P(C|girl) \log_2 P(C|girl))) \\ &= 5.67 \end{aligned}$$

.....依次求取最大增益为判别



#### 4. 编程题

对米兰大教堂不同时刻图片进行k-means聚类分析：

- 聚类个数k值决定最终通过图片RGB信息，聚类为几类，聚类数目多，则特征更多，但信息更杂乱
- 对于建筑与背景色差较大的图片，可以较好地通过聚类提取建筑物特征，反之则不易，比如下图米兰大教堂黄昏时刻建筑物与背景，RGB信息接近，故提取到建筑特征模糊

