**第6章 决策树**

练一练

**一、判断题**

1、 事件发生的可能性越大，其所包含的信息量也越大。 （ ）

2、C5.0决策树算法剪枝方法中，节点错误率是使用该节点错误率区间估计的下限来代替点估计错误率。 （ ）

3、 使用Bagging集成方法时，可以使用并行化的方法提高运算效率。 （ ）

**二、单选题**

1、熵表示为消除不确定性所需的信息量，投掷一枚硬币的熵是（ ）比特。

A. 1 B. 2.6 C. 3.2 D. 1.8

2、决策树中不包含（ ）。

A. 根节点 B. 内部节点 C. 外部节点 D. 叶子节点

3、下列属于C5.0算法特点的是（ ）。

(1)集成了MDLP方法对连续属性离散化 (2)使用悲观剪枝法以避免过拟合

(3)能处理带有缺失值的数据集 (4)偏向于选择取值数目多的属性

A. (1) (2) (3) B. (1) (2) (4) C. (2) (3) (4) D. (1) (2) (3) (4)

4、以下对于ID3算法描述，不正确的是（ ）

A. ID3算法使用信息增益作为最佳分裂属性的选择标准

B. ID3算法不能处理连续型属性

C. ID3算法不能处理带有缺失值的数据集

D. ID3算法偏向于选择取值数目少的属性

练一练答案

**一、判断题**

1、**×**

**2、×**

**3、√**

**二、单选题**

1、A

2、C

3、A

4、D