## 第三讲（预处理）

1、数据预处理的三个流程按顺序为：（）

A 数据清理、数据归约、数据集成

B 数据集成、数据清理、数据归约

C 数据清理、数据集成、数据归约

D 数据归约、数据清理、数据集成

2、假设12个销售价格记录组已经排序如下：5, 10, 11, 13, 15,35, 50, 55, 72, 92, 204, 215 使用如下每种方法将它们划分成四个箱。等频（等深）划分时，15在第几个箱子内？ ()

A.第一个

B.第二个

C.第三个

D.第四个

3、上题中，等宽划分时（宽度为50），15又在哪个箱子里？ ()

A 第一个

B 第二个

C 第三个

D 第四个

4、只有非零值才重要的二元属性被称作：（）

A 计数属性

B 离散属性

C非对称的二元属性

D 对称属性

5、以下哪种方法不属于特征选择的标准方法?()

A嵌入

B 过滤

C 包装

D 抽样

6、下面不属于创建新属性的相关方法的是?()

A特征提取

B特征修改

C映射数据到新的空间

D特征构造

7、下面哪个属于映射数据到新的空间的方法?()

A 傅立叶变换

B特征加权

C 渐进抽样

D维归约

8、假设属性income的最大最小值分别是12000元和98000元。利用最大最小规范化的方法将属性的值映射到0至1的范围内。对属性income的73600元将被转化为?（）

A 0.821

B 1.224

C 1.458

D 0.716

9、(多选)在现实世界的数据中，元组在某些属性上有缺失值是常见的。描述处理该问题的各种方法有哪些？()

A忽略元组

B使用属性的平均值填充空缺值

C使用一个全局常量填充空缺值

D使用与给定元组属同一类的所有样本的平均值

E使用最可能的值填充空缺值

10、（多选）下面属于维归约常用的线性代数技术的有： ()

A 主成分分析

B 特征提取

C 奇异值分解

D 特征加权

E 离散化

11、如果两个变量的变化趋势相反，那么两个变量间的协方差就是负值（）

A 对

B 错