**第4章 数据预处理**

练一练

**一、判断题**

1、对于不平衡数据，使用欠采样可以增加少数类样本的数量，以达到平衡的要求。 （ ）

2、等深分箱法是指每个箱子具有相同的区间范围。 （ ）

3、在数据预处理中，只要是异常数据，都应该将其去除。 （ ）

4、 聚集主要指对数据进行不同维度的汇总，使用聚集后高粒度的值代替原始值，可实现数值归约的目的。 （ ）

**二、单选题**

1、如果属性销售额的最小、最大值分别为285万元和3255万元，利用最大最小规范化的方法将属性值映射到0~1之间，则属性销售额为1245万元将被转化为（ ）。

A. 0.6768 B. 0.3825 C. 0.3232 D. 0.4192

2、主成分分析法是解决原始数据中存在的哪个问题的常用方法（ ）

A. 数据异常 B. 数据不平衡

C. 数据高维 D. 数据不一致

3、假设浙江工商大学经济统计学专业2019级16名学生的月支出（千元）如下：2.1、2.7、1.9、4.2、3.7、4.3、2.6、3.4、2.9、3.2、2.0、1.8、3.5、3.6、2.8、3.1，若采用等深分箱法分成4箱，则3.5在第几个箱内（ ）。

A. 第一个 B. 第二个 C. 第三个 D. 第四个

练一练答案

**一、判断题**

1、**×**

**2、×**

**3、×**

**4、√**

**二、单选题**

1、C

2、C

3、C