## 习题6答案

1. 练习题

1.判断题

聚类的目标是最小化簇间的相似性, 最大化簇内的相似性。（对）

2.判断题

采用一组标注好的数据进行训练，再对所有的未知点做出预测，属于无监督学习（错）。

3.判断题

基于已有的房屋销售信息，来预测房价的具体数值属于回归问题。（对）

4.单选题

按照某种指定的属性特征，划分成两个或多个类别，属于（ C ）问题。

1. 回归 B.关联 C.分类

5.多选题

数据清洗包括（ABCD）。

A.缺失值处理

B.异常值处理

C.去除重复

D.修改格式错误

二、求解

1. 假设在某个局部地区细胞识别中正常（）和异常（）两类的先验概率分别为：

正常状态：=0.9

异常状态：=0.1

现有以待识别的细胞，其观察值为，从类条件概率密度分布曲线上查得

=0.2，=0.4

试对该细胞进行分类。

解：利用贝叶斯公式，分别计算出及的后验概率：

根据贝叶斯决策规则式，有

所以合理的决策是把归类于正常状态。

1. 在上题条件的基础上，利用下表为决策表，按最小风险贝叶斯决策进行分类。

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | 0 | 6 |
| 1 | 0 |

解：已知条件为

=0.9，=0.1

=0.2，=0.4

*，*

*，*

根据上题的计算结果可知后验概率为：

，

因此计算条件风险：

由于，即决策为的条件风险小于决策为的条件风险，因此我们采取决策行动，即判断待识别的细胞为类——异常细胞。

本题的结果与上题相反，这是因为“损失”影响了决策结果。