

参考答案

1. 数据误差依据产生原因分类可分为哪几种？

答：数据误差依据产生原因分类一般可分为来源误差、处理误差、语义误差。

2. 多源异构数据误差一般具有哪些特点？

答：多源异构数据误差可能具有以下特点：(1)多样性：由于数据来源多样，可能包括不同的系统、设备、应用程序或组织，因此误差的类型和原因也多种多样。(2)复杂性：多源异构数据在结构、格式、语义上存在差异，这使得在数据整合和处理过程中更容易出现误差。(3)不确定性：由于数据源的多样性和复杂性，以及处理过程中的各种因素(如数据清洗、转换、集成等)，误差的出现具有不确定性。

3. 测量误差根据其成因和特性分为哪些类别，各有什么显著特点？

答：测量误差可分类为系统误差、随机误差和过失误差，每种误差都有其独特的特点。系统误差：是一种经常误差，误差值在整个测量过程中保持一定规律，而且是可以避免的。随机误差：又称偶然误差，误差的绝对值和符号变化无常，但当测量数据较多时，随机误差的数值分布符合一定的统计规律，一般是正态分布，随机误差是不可避免的。过失误差：又称粗大误差一般数值较大，可直接剔除。

4. 数据分析挖掘的过程一般包含哪几个步骤？

答：数据分析挖掘的过程：(1)任务目标的确定；(2)目标数据集的提取；(3)数据预处理；(4)建立适当的数据分析与挖掘模型；(5)模型的解释与评估；(6)知识的应用。

5. 多源异构数据统计分析常用方法有哪些？

答：多源异构数据统计分析常用方法：回归分析、主成分分析、分类算法、聚类算法、关联规则算法、时间序列分析等。

6. 数据挖掘常用算法有哪些？

答：数据挖掘常用算法：主成分分析、分类算法、聚类算法等也可以看作数据挖掘算法的一种，此外还有神经网络算法、机器学习算法、模式识别算法等。

7. 机器学习方法的三要素，依据监督学习的程度可以怎么分类？

答：机器学习方法都是由模型、策略和算法(优化算法)构成的，即机器学习方法由三要素构成，可以简单地表示为：方法=模型+学习准则+优化算法。大部分机器学习算法可以分成监督学习、半监督学习和无监督学习。