

数据仓库与数据挖掘模拟三

共 5 道题 · 内容提取自网页存档

1 简答题 20 分

1.何为聚类分析？（5分） 2.请写聚类分析的经常应用的五大方面。（5分）（共10分）

2 简答题 20 分

1. 何为有监督学习？ 何为无监督学习？（5分） 2.什么是决策树？ 请详细写出使用决策树进行分类的算法步骤。（15分） 3.针对下面数据集，请使用ID3算法求出根节点，并写出求解步骤。（10分）（共30分）

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
30...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

3

简答题 20 分

1.对关联规则 $X \& Y \rightarrow Z$ ，请说明其支持度s和置信度c的含义。请分别举例写出一条单维布尔型和三维布尔型关联规则，并说明其实际含义。（10分） 2.何为频繁项集？假定 L_{k-1} 是频繁K-1项集，请写出生成候选项集CK的算法语句。（10分）（共20分）

4

简答题 20 分

1.请简述分离数据库和数据仓库的主要原因（5分） 2.请简述数据仓库的多维数据模型（5分） 3.请简述多维数据库模式。（10分）（共20分）

5

简答题 20 分

1.请简述为什么数据仓库的功能是支持管理层进行科学决策，而不是事务处理？（16分） 2.请简述为什么把数据仓库与数据挖掘集成到一个系统中将能够更有效地提高系统的决策支持能力？（4分）（共20分）