

西安电子科技大学

考试时间 120 分钟

试 题

题号	一	二	三	四	总分
分数					

1. 考试形式：闭卷 ☒ 开卷 ☐

2. 考试日期： 年 月 日 (答题内容请写在装订线外)

一、单选题（每小题 3 分，共 30 分）

1. 在关系数据模型中，用来唯一标识一个元组的最小属性集合，称为 （ ）。

- (A) 码
- (B) 关系属性
- (C) 主键
- (D) 域

2. 卷积神经网络是一个多层神经网络，每层由多个二维平面组成，而每个平面由多个独立神经元组成，如图 1 所示，其中 C3 层表示 （ ）。

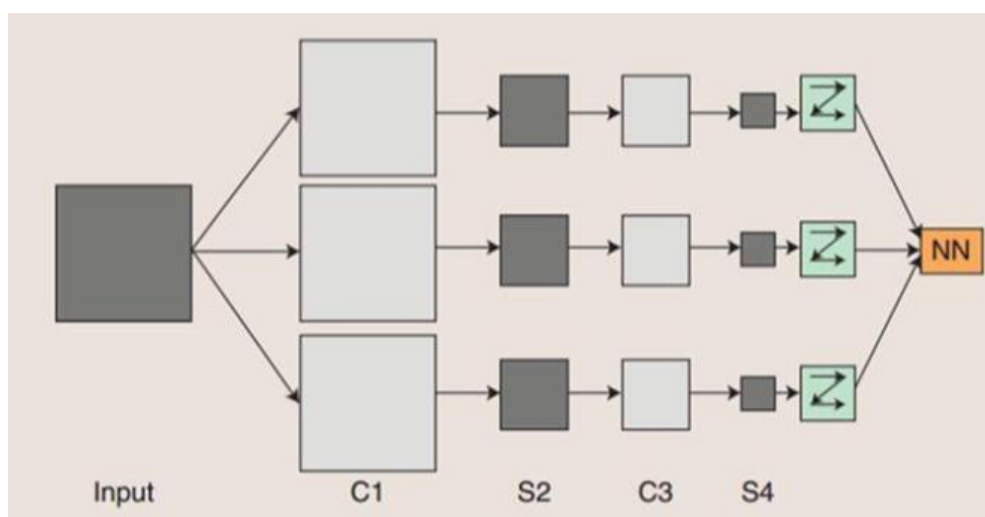


图 1. 特征映射图

- (A) 映射层
- (B) 卷积层

- (C) 全连接层
- (D) 采样层
3. 关于词袋模型的描述正确的是 ()。
- (A) 是较为复杂的词向量表示方法
- (B) 使用一组无序的单词来表达一段文字或一个文档
- (C) 考虑文本的语法
- (D) 文档中每个单词的出现都是关联的
4. () 用于进行身份鉴别, 证明用户身份, 即“你是谁”。
- (A) 授权管理基础设施 (PMI)
- (B) 公钥基础设施 (PKI)
- (C) 信任源点 (SOA)
- (D) 资源管理 (RM)
5. 以数据为中心的智慧城市有诸多特点, 以下不属于的是 ()。
- (A) 绿色宜居
- (B) 多学科交叉
- (C) 数据驱动
- (D) 区域特性
6. 以下不属于工业大数据特征的是 ()。
- (A) 多模态
- (B) 价值密度高
- (C) 强关联
- (D) 高通量
7. 关于流式数据, 描述错误的是 ()。
- (A) 流式数据是一组顺序、大量、快速、连续到达的数据序列
- (B) 流式数据常用于网络监控、传感器网络、航空航天、气象测控等领域
- (C) 流式数据可以在不额外付出代价的情况下, 被再次取出处理
- (D) 流式数据到达次序独立, 不受应用系统所控制
8. () 包括对数据生命周期整个过程实现完全意义的管理, 包括数据生产、加工、采集、使用、传输、存储、销毁等环节。
- (A) 数据管理权

(B) 数据控制权

(C) 数据主权

(D) 数据财产权

9. 标识两个数据对象语义相关的过程叫 ()。

(A) 模式识别

(B) 语义翻译

(C) 数据映射

(D) 模式匹配

10. Hadoop 分布式文件系统采用的是 () 的主从结构

(A) 多主服务器

(B) 单一从服务器

(C) 单一主服务器

(D) 多主服务器和若干数据服务器

二、判断题（每小题 2 分，共 20 分）

1. 根据数据采集方式的不同，数据采集分为基于拉的方法和基于推的方法。()

2. 关系数据模型基本数据结构就是关系，一个关系对应一个二维表。()

3. 事务由一系列的数据库操作构成，必须满足的特性有：原子性、隔离性和持久性。()

4. NoSQL 数据库中 MongoDB 是一款分布式文档数据库，它为大数据量、高度并发访问、弱一致性要求的应用而设计。()

5. K-means 算法是一种简单、快速的经典算法，是一种有监督学习的算法。()

6. 使用平行坐标进行高位数据可视化时，每个坐标轴表示一个维度，每条折线代表一个数据记录。()

7. 开放数据是能被任何人、在任何时间和任何地点进行自由利用、再利用和分发的电子数据。()

8. 城市大数据不仅具备大数据的固有特性，还具备时空多维性、多元性、异构性、多尺度与多粒度等特点。()

9. 数据安全更关注于数据存储和传输过程的内容防护及隐私保护。()

10. XML 语言是一种典型的结构化数据。()

三、填空题（每小题 2 分，共 20 分）

1. 按照数据是否有强的结构模式，可将数据划分为：_____、
_____和_____。
2. 1EB = _____ GB。
3. 按照不同的用途，结构化查询语言通常被分成三个子集（子语言），分别为
_____、_____和_____。
4. 影响力最大化问题的目标函数具有 _____、
_____和_____性质。
5. 数据安全与传统信息安全共性问题包括 _____、
_____和_____。
6. IBM 给出了大数据“4V 特性”，具体包括 _____、
_____和_____等方面。
7. 根据开放数据的来源不同可以将开放数据划分为 _____、
_____和_____。
8. 缺失值填充是针对不完整数据进行缺失数据的填充，常用的填充方法包括
_____、_____、_____和
_____。
9. 关系数据模型的数据操作一般分为两类： _____ 和
_____。
10. 距离度量应满足的四个性质的： _____、 _____、
_____和_____。

四、简答题（共 30 分）

1. 请简要介绍一下数据、信息、知识之间的区别。（6 分）

2. 描述数据处理的一般过程。(8 分)

3. 简单描述机器学习中的有监督学习和无监督学习，并分别列举常见的有监督学习和无监督学习算法。(8 分)

4. 你注册过哪些在线社会网络平台？请你谈谈你使用这些在线社会网络平台的心得。
(8 分)