

西安电子科技大学

考试时间 120 分钟

试 题

题号	一	二	总分
分数			

1. 考试形式：闭卷☒ 开卷☐
2. 考试日期：2022 年 月 日
3. 所有试题答案写在答题纸上, 答案写在试卷上无效。

一、简答题（共 36 分）

1. 模型评估与选择（8 分）。
 - (1) 简述过拟合产生的原因。
 - (2) 简述克服神经网络分类器过拟合的方法。
 - (3) 简述 k 折交叉验证法。
2. 支持向量机（SVM）（6 分）。
 - (1) 描述 SVM 中将原始问题转换成对偶问题求解的优势。（至少列举 3 点）
 - (2) 简述 SVM 中核函数的作用。
3. 神经网络（6 分）。
 - (1) 简述池化层在卷积神经网络中的作用。（至少列举 3 点）
 - (2) 简述卷积神经网络中 1×1 卷积的作用。（至少列举 3 点）
4. 决策树（6 分）。
 - (1) 分别列举出三种决策树 ID3、C4.5 和 CART 的最优划分属性度量。
 - (2) 简述预剪枝和后剪枝处理的优势和缺点。
5. 聚类（10 分）。
 - (1) 简述 K-means 聚类方法的流程。
 - (2) 简述直接密度可达、密度可达、密度相连的定义。
 - (3) 写出 DBSCAN 使用基于中心的密度将样本点划分为三类点的名称。

二、计算题（共 64 分）

1. 相似性计算（4 分）。
 - (1) 给定向量 $x=(1,0,0,0,0,0,0,0,0,0)$, $y=(0,0,0,0,0,0,1,0,0,1)$, 计算简单匹配系数 (SMC) 和 Jaccard 系数 (J)。
 - (2) 给定向量 $x=(1,1,0,1,0,1)$, $y=(1,1,1,0,0,1)$, 计算余弦值。
 - (3) 给定向量 $x=(0,1,0,1)$, $y=(1,0,1,0)$, 计算皮尔逊系数。

2. 神经网络的基本参数和复杂度计算（20 分）。

神经网络由卷积层和全连接层组成。输入图像大小为 $18 \times 18 \times 3$ ，第一个卷积层有 3 个 3×3 卷积核，第二个卷积层有 5 个 3×3 卷积核，假设每一个卷积核滑动步长 $s_1=1$ ，边界填充 $p_1=0$ ；前两个卷积层后均有最大池化，池化核大小为 2×2 ，其步长 $s_2=2$ ，边界填充 $p_2=0$ 。第二个池化层与全连接输入层（该层有 100 个神经元）的连接通过 3×3 卷积实现。全连接层仅有一个含 100 个神经元隐含层，输出层有 10 个神经元。求解以下问题并给出计算步骤。

- （1）分别计算第一个和第二个池化层输出的特征图大小。
- （2）计算整个网络需要估计的网络参数量。
- （3）如果一次乘法和一次加法均表示一次浮点运算，计算第 1 次卷积的浮点计算量，并且计算全连接输入层到第 1 个隐含层的浮点计算量。
- （4）如果将一个‘乘-加’组合视为一次浮点运算，计算第 1 次卷积的浮点计算量，并且计算全连接输入层到第 1 个隐含层的浮点计算量。

3. 频繁模式挖掘（18 分）。

- （1）对表 I 中所示的事务数据集 D，计算用 Apriori 算法生成频繁项集的过程和最终所有频繁项集的结果（设最小支持度 $\text{sup}=50\%$ ）。
- （2）计算并写出由频繁项集推出所有强关联规则的过程（设最小置信度 $\text{conf}=70\%$ ）。

表 I 事务数据库 D

TID	项集
1	{1, 3, 4}
2	{2, 3, 5}
3	{1, 2, 3, 5}
4	{2, 5}

4. 序列模式挖掘（6 分）。

考虑下面各频繁 3-子序列：

$\langle \{1\}\{2\}\{3\} \rangle$ 、 $\langle \{1\}\{2, 5\} \rangle$ 、 $\langle \{1\}\{5\}\{3\} \rangle$ 、 $\langle \{2\}\{3\}\{4\} \rangle$ 、 $\langle \{2, 5\}\{3\} \rangle$ 、 $\langle \{3\}\{4\}\{5\} \rangle$ 、 $\langle \{5\}\{3, 4\} \rangle$ 。

- （1）列出利用序列模式发现的类 Apriori 算法的候选生成步骤产生的所有候选 4-序列。
- （2）列出利用序列模式发现的类 Apriori 算法的候选剪枝步骤剪掉的所有候选 4-序列和剪枝后的候选序列。

5. 凝聚层次聚类（16 分）。

给定一组样本点 **p1, p2, p3, p4, p5, p6** 之间的距离矩阵（如表 II 所示），利用单链方式（最短距离）进行层次聚类，写出详细的聚类计算过程，画出层次树状图。

表 II 6 个样本点之间的距离矩阵

	p1	p2	p3	p4	P5	P6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.148	0.20	0.14	0.25
p3	0.22	0.148	0.00	0.151	0.28	0.11
p4	0.37	0.20	0.151	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
P6	0.23	0.25	0.11	0.22	0.39	0.00