**重庆邮电大学研究生考卷**

学号 姓名 考试方式 开 卷 班级

考试课程名称 人工智能与大数据 考试时间：2023年1月5日

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 题号 | 一 | 二 | 三 | 四 | 五 | 总分 |
| 得分 |  |  |  |  |  |  |

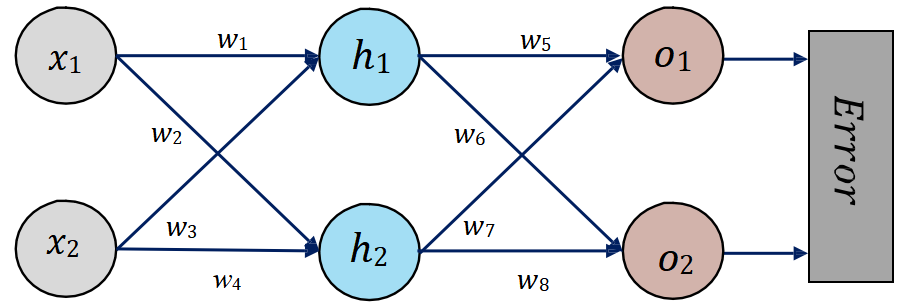
1. 表1是一个由15个样本组成的贷款申请训练数据。数据包括贷款申请人的4个特征(属性)：
2. 第1个特征是年龄，有3个可能值：青年、中年、老年；
3. 第2个特征是有工作，有2个可能值：是、否；
4. 第3个特征是有自己的房子，有2个可能值：是、否；
5. 第4个特征是信贷情况，有3个可能值：非常好、好、一般。

表的最后一列是类别，是否同意贷款，取2个值：是、否。对表1所给的训练数据集，根据信息增益准则选择最优特征，并写出计算过程。(本题共20分)

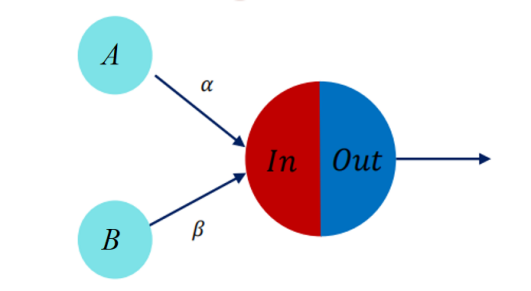
**表1 贷款申请样本数据表**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **年龄** | **有工作** | **有自己的房子** | **信贷情况** | **类别** |
| 1 | 青年 | 否 | 否 | 一般 | 否 |
| 2 | 青年 | 否 | 否 | 好 | 否 |
| 3 | 青年 | 是 | 否 | 好 | 是 |
| 4 | 青年 | 是 | 是 | 一般 | 是 |
| 5 | 青年 | 否 | 否 | 一般 | 否 |
| 6 | 中年 | 否 | 否 | 一般 | 否 |
| 7 | 中年 | 否 | 否 | 好 | 否 |
| 8 | 中年 | 是 | 是 | 好 | 是 |
| 9 | 中年 | 否 | 是 | 非常好 | 是 |
| 10 | 中年 | 否 | 是 | 非常好 | 是 |
| 11 | 老年 | 否 | 是 | 非常好 | 是 |
| 12 | 老年 | 否 | 是 | 好 | 是 |
| 13 | 老年 | 是 | 否 | 好 | 是 |
| 14 | 老年 | 是 | 否 | 非常好 | 是 |
| 15 | 老年 | 否 | 否 | 一般 | 否 |

1. 如图1、图2所示，神经元*h*1、*h*2、*o*1、*o*2的激活函数是sigmoid函数，。计算的梯度、的梯度，并写出计算过程(本题共20分)



**图1 神经网络的结构**



**图2 神经元的结构**

1. 如表2所示，假设有5笔交易数据，编号是T1～T5。求四个关联规则A=>B、B=>A、B=>C和C=>B的支持度和置信度，填入表3中，并写出计算过程。(本题共20分)

**表2 五比交易数据**

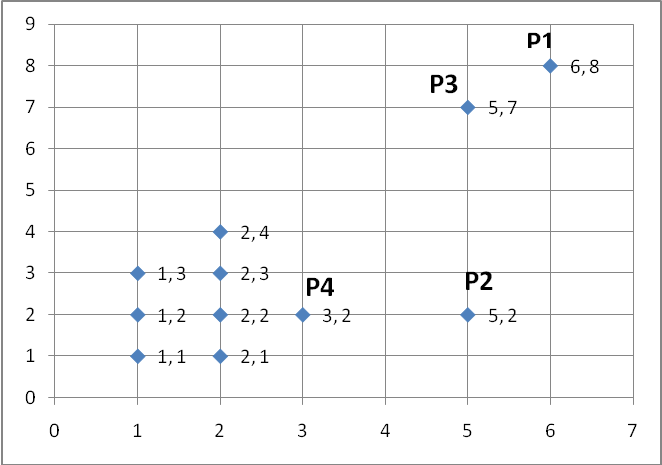
|  |  |
| --- | --- |
| 交易数据编号 | 数据项 |
| T1 | A、B、C、D |
| T2 | A、B |
| T3 | C、D、E |
| T4 | B、C |
| T5 | B、C、E |

**表3 四个关联规则的支持度和置信度**

|  |  |  |
| --- | --- | --- |
| 关联规则 | 支持度 | 置信度 |
| A=>B |  |  |
| B=>A |  |  |
| B=>C |  |  |
| C=>B |  |  |

1. 有一种基于距离的离群点检测算法被称为*k*最近邻算法。假设有一个样本*p*，用一种距离的度量方法（例如欧氏距离）来计算得到离*p*最近的*k*个样本，这*k*个样本可以组成集合*N*(*p*,*k*)。*p*的离群因子可以定义为：

其中，*distance*(*p*,*y*)表示样本*p*和*y*的距离度量，表示集合*N*(*p*,*k*)的大小，即所包含样本的个数。离群因子越大，越有可能是离群点。如图3所示的二维数据集中，当*k*=2时，使用欧式距离，判断P1、P2哪个点更可能是离群点？写出判断过程。(本题共20分)



**图3 数据集**

1. 在图4中，机器人寻路问题被简化为2×2的网格：

|  |  |
| --- | --- |
| *S*3 | *S*4 |
| *S*1 | *S*2 |

**图4 网格**

假设有位于*S*1位置的机器人拟从这一初始位置向*S*4这一目标位置移动。机器人每次只能向上或者向右移动一个方格，到达目标位置*S*4则会获得奖励且游戏终止，机器人在移动过程中如果越出方格(*Sd*)则会被惩罚且被损坏，并且游戏终止。奖励值定义如下：当*Sn*+1=*S*4时奖励值为1；当*Sn*+1=*Sd*时惩罚值为-1；其他情况下奖励值为0。若折扣因子= 0.99，智能体在*S*1、*S*2、*S*3的策略都初始化为“上”，终止状态*S*4、*Sd*的价值函数定义为0，列出求解状态*S*1、*S*2、*S*3价值函数的贝尔曼方程组，并解之。(本题共20分)