**复旦大学研究生课程考试试卷**

（2022 - 2023 学年第 1 学期）

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 授课学期： | | 2022-2023学年第1学期 | | | 考试日期： | | 2022-12-21 | |
| 课程名称： | | 机器学习与神经网络导论 | | | 课程代码： | | MATH620165 | |
| 开课院系： | | 数学科学学院 | | | 试卷类型： | | ☑A卷 □B卷 | |
| 考试形式： | | ☑开卷 □闭卷 □课程论文 □其他 | | | | | | |
| 姓名： |  | | 学号： |  | | 成绩： | |  |

（共6题，每题20分，1-5题选4题完成，6-10题选2题完成）

一、 （15分） 叙述并推导Resnet型神经网络的BP监督学习算法。分别推导以下神经网络层的随机梯度公式包含：卷积层（以Relu函数作为激发函数）、Max Pooling层和Normalization层、残差层、全连接层（以Softmax作为激发函数）；以交叉熵加上参数L2的regularization作为惩罚函数。

二、（15分）简述分类（Classification）与聚类(Clustering)的区别与联系。在做分类问题时，如果训练数据量较少，如何处理来提高分类算法的可靠性。

三、（15分） 什么是Fisher Information Matrix? 并列举三个Fisher Information Matrix在理论分析和算法设计的应用。

四、（15分）在统计检测(Detection)问题中，什么是第一类错误和第二类错误？试叙述并证明Newman-Pearson引理。

五、（15分）关于潜在顾客的购买可能的估计问题。每个顾客的信息包括：年龄（Age）、收入情况(Incoming)、是否是学生(Is Student?)、信用等级(Credit Rating)。是通过下表（表一）中的数据建立Decision Tree，并判断如下用户的购买可能。新用户的信息如下：

年龄(50), 收入（Medium）,非学生，信用记录 (excellent).

表一

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| User id | Age | Incoming | Student | Credit Rating | Buying |
| 1 | <=30 | high | no | fair | no |
| 2 | <=30 | high | no | excellent | no |
| 3 | [31,40] | high | no | fair | yes |
| 4 | >40 | medium | no | fair | yes |
| 5 | >40 | low | yes | fair | yes |
| 6 | >40 | low | yes | excellent | no |
| 7 | [31,40] | low | yes | excellent | yes |
| 8 | <=30 | medium | no | fair | no |
| 9 | <=30 | low | yes | fair | yes |
| 10 | >40 | medium | yes | fair | yes |
| 11 | <=30 | medium | yes | excellent | yes |
| 12 | [31,40] | medium | no | excellent | yes |
| 13 | [31,40] | high | yes | fair | yes |

六、 （20分） 多神经元对刺激方向的响应模型如下。假设每个神经元的条件发放率表示如下

其中a=1，…,N是神经元编号，s是受到外部刺激的方向角度（[0,180°]），是神经元a的偏好方向，是神经元调谐函数的参数。假设

* 神经元动作电位发放之间是统计独立的；
* 神经元动作电位发放是一个泊松过程。

假设贝叶斯损失为平方差：

是估计方向角度。

通过这N个神经元的动作电位发放时间点，建立对外部刺激方向s的最小方差的无偏(MVUB)估计，并计算其平方误差。

七、（20分）利用Hopfield神经网络实现一个TSP问题的求解，并进行数值验证。

八、（20分）利用Reinforcement Learning算法求解迷宫最短路径搜索问题。迷宫图请见附件(maze.jpg；黄色是出发位置)。

要求：任意给出中点位置，规划出最短路径或者告知不能到达。最好给出交互式的界面。

九、（20分）建立系统用于识别垃圾邮件和非垃圾邮件，参考数据库：<http://www.aueb.gr/users/ion/data/enron-spam/>

十、（20分） 深证Ｂ股指数有大约50多家样本股的价格加权平均而得，请通过一段时间的历史数据，挑出尽量少的样本股精确估计该指数。

数据来源：http://www.szse.cn