

深圳大学期末考试试卷

开/闭卷 开 A/B 卷
课程编号 01 课程名称 机器学习 学分 2.5
命题人(签字) 审题人(签字) 2024 年 11 月 29 日

题号	一	二	三	四	五	六	七	八	九	十	基本题 总分	附加题
得分												
评卷人												

一、可选题目

1、泰坦尼克号生存计划

问题描述：泰坦尼克号的沉没是历史上最著名的船难之一。1912 年 4 月 15 日，在其首航期间，被广泛认为“永不沉没”的 RMS 泰坦尼克号在与冰山相撞后沉没。不幸的是，船上的救生艇数量不足，导致 2224 名乘客和船员中有 1502 人遇难。虽然幸存下来在某种程度上带有一定的运气成分，但似乎某些群体的人更有可能幸存。在本题中，我们希望你建立一个预测模型，回答以下问题：“什么样的人更有可能幸存？”通过乘客数据（如姓名、年龄、性别、社会经济阶层等）进行分析。

训练数据 (train.csv) 用于构建你的机器学习模型。对于训练集，我们提供了每位乘客的结果（也称为“真实值”）。你的模型将基于乘客的性别、舱位等级等“特征”构建。你还可以通过特征工程创建新的特征。

测试集 (test.csv) 用于评估你的模型在未见过的数据上的表现。对于测试集，我们没有提供每位乘客的真实值。你的任务是预测这些结果。对于测试集中的每位乘客，使用你训练的模型预测他们是否在泰坦尼克号沉没中幸存。并根据人类直观经验，讨论所预测结果是否合理，并分析模型作出本次预测的主要因素。

任务：

- 将 train.csv 手动划分为训练集和验证集（推荐比例 8:2 或 7:3），训练集用于机器学习模型训练，验证集用于检测所训练模型效果，并根据验证集验证效果选择合适停止条件、超参数等变量，在测试集 (test.csv) 用于评估你的模型在未见过的数据上的表现。
- 采用图、表等方式分析数据集特征。
- 选用一种机器学习模型对什么样的人更有可能幸存进行建模预测：要求理论与实践相结合，并采用正确的模型评估方法，基于多种评价指标分析预测结果。
- 根据模型预测结果，分析影响幸存可能性的重要因素。
- 扩展：用多种机器学习模型进行预测建模并对预测性能进行对比分析。

要求把算法原理、评价标准和实验过程描述清楚，实验结果结合图表多样化呈现，并对算法模型进行合理分析。

最终成绩由课程报告、实验结果及分析和上交代码综合评定。

2、手写数字识别

问题描述：MNIST (“Modified National Institute of Standards and Technology”) 数据集是计算机视觉领域的事实标准 “Hello World” 数据集。自 1999 年发布以来，这个经典的手写数字图像数据集一直作为分类算法基准测试的基础。随着新的机器学习技术不断涌现，MNIST 仍然是研究人员和学习者的可靠资源。

在本题中，你的目标是从成千上万的手写数字图像数据集中正确识别数字。

数据文件：train-minst.csv 和 test-minst.csv 包含灰度手绘数字的图像数据，范围从 0 到 9。

- 每张图像为 28 像素高，28 像素宽，总计 784 像素。
- 每个像素都有一个对应的像素值，用于表示该像素的亮度或暗度。数值越高表示越暗，像素值是介于 0 到 255 之间的整数。

训练数据集：train-minst.csv 文件包含 785 列：

- 第一列名为 “label”，表示用户绘制的数字。
- 其余列包含与图像相关的像素值。

像素列命名规则：

训练集中每个像素列的名称为 pixelx，其中 x 是 0 到 783 之间的整数。要定位图像上的这个像素，假设将 x 分解为 $x = i * 28 + j$ ，其中 i 和 j 分别是 0 到 27 之间的整数。那么 pixelx 就位于 28x28 矩阵的第 i 行、第 j 列（从 0 开始索引）。

任务：

- 将 train-minst.csv 手动划分为训练集和验证集（推荐比例 8:2 或 7:3），训练集用于机器学习模型训练，验证集用于检测所训练模型效果，并根据验证集验证效果选择合适停止条件、超参数等变量，在测试集 (test-minst.csv) 用于评估你的模型在未见过的数据上的表现。
- 采用图、表等方式分析数据集特征。
- 选用一种机器学习模型对手写字图片进行分类：要求理论分析与代码实践相结合，并采用正确的模型评估方法，基于多种评价指标分析预测结果。
- 根据模型预测结果，分析影响分类准确率的重要因素。
- 扩展：用多种机器学习模型进行预测建模并对预测性能进行对比分析

要求把算法原理、评价标准和实验过程描述清楚，实验结果结合图表多样化呈现，并对算法模型进行合理分析。

最终成绩由课程报告、实验结果及分析和上交代码综合评定。

二、课程报告要求

请从给出的题目中选择一个完成课程报告。课程报告需使用模板《课程报告专用答题纸》撰写（模板下载网址：<https://jwb.szu.edu.cn/info/1074/1077.htm>），其主要内容需包括以下几个部分：1. 标题；2. 背景介绍；3. 方法描述；4 实验内容和实验结果；5. 分析与结论。报告的排版格式可参考本科生毕业论文撰写规范及要求 (<https://jwb.szu.edu.cn/info/1358/2010.htm>)。

三、课程报告提交

提交电子版课程报告和程序代码文件。其中，课程报告提交 word 版本文件，命名为“学号-姓名-课程报告”。程序代码须可直接运行，有必要的注释，代码文件名须具有较好的可读性。所有文件整理到以“学号-姓名-课程报告”命名的文件夹中，并将文件夹压

缩打包提交到 Blackboard 上相应的位置。

四、截止时间

课程报告及相关文件提交截止时间为：2025 年 01 月 05 日 23:59，逾期成绩无效。

五、评分标准：

- 1、课程报告撰写规范，结构完整，叙述详细，语言表达准确，排版合理；程序代码清晰，可读性高；20%
- 2、查阅资料的情况，所选题目及所构造算法模型相关的背景介绍；10%
- 3、结合机器学习相关知识，运用图、表、数值等多样化形式对数据集进行分析和处理；20%
- 3、运用恰当的机器学习算法建立模型，并给出详细的理论分析与代码实践；30%
- 4、结合实际问题的，运用图、表、数值等多样化形式对实验结果以及算法模型进行分析；20%
- 5、如果发现抄袭迹象，课程报告成绩无效。