

模式识别课程实验说明

2023年春季

实验整体说明

- 基本要求

- 从给出的5个子课题题目中任选3个（其中神经网络为必选）进行算法设计与实验，形成代码、结果文件与实验报告，并进行提交。

- 提交时间、地址与格式

- 纸质版报告请于2023年6月9日时进行提交，电子档于2023年6月9日24时整前发送至邮箱 seu_pr_2023@163.com。
- 提交的电子档请命名为“学号-姓名”（如“09118666-张三”），该电子档需包含3个子课题的压缩文件（命名方式与压缩文件包含的具体内容请参考子课题说明文件）与一个实验报告pdf文件。

实验整体说明

- 实验文档与数据

- 实验详细文档、报告模板、及数据下载链接

<https://pan.baidu.com/s/1y6LBr18NAH-FmuGqR8rSGQ>

提取码 z5re

- 注册华为云账号，并将华为云账号ID及学生姓名提供给助教，助教发放代金券

<https://auth.huaweicloud.com/authui/login.html?service=https%3A%2F%2Fconsole.huaweicloud.com%2Fconsole%2F#/login>

实验整体说明

- 华为**MindSpore**平台使用

- 华为云账号ID及学生信息登记

<https://docs.qq.com/sheet/DRExRd1IYbnVlcnND?tab=BB08J2&u=a71ef9e1b72d4d3ea9255f5a085f25c1>

- MindSpore使用入门，请参见

<https://mindspore.cn/resources/courses/list?id=47>

- MindSpore官网，请参见<https://www.mindspore.cn/>

- MindSpore官方模型脚本仓库，请参见

<https://gitee.com/mindspore/models>

子课题要求

- 算法与实验要求

- 根据子课题的数据与任务说明，完成指定的课题任务，形成相应的代码文件、测试结果文件等。
- 不限制代码语言（推荐Python）
- 不得使用集成度较高、函数调用式的代码库（如Python环境下的sklearn、PyTorch、Tensorflow、Opencv等），可采用（Numpy、Pandas、Matplotlib等基础或辅助代码集成库）。
- 详情请参考子课题说明文档。

子课题要求

- 格式要求

- 需提供完整的**代码文件**、**测试集分类结果文件**，将以上内容打包压缩，压缩文件命名格式：**学号-姓名-xxx任务实验**。实验报告和代码注释应尽量详细。具体提交内容与格式请参照不同的子课题说明。
- 尽量以**相对路径**的形式索引数据集，便于我们对代码进行复现。
- 代码和报告若有雷同或者抄袭，一律按**0分**处理。

子课题1-数据降维与分类任务

- 概述

- 利用两种降维技术，对葡萄酒数据进行降维并分类。

- 任务说明

- 对红白葡萄酒数据分别进行PCA和LDA降维处理，在降维后的数据集上完成基于logistic回归分类器的训练和测试。要求比较应用降维技术前后，分类器准确率的变化。（对于降维后的数据，可以尝试利用可视化方法展示结果。）

子课题1-数据降维与分类任务

- **任务说明**

- 基于MindSpore平台提供的官方模型库，对相同的数据集进行训练，并与自己独立实现的算法对比结果（包括但不限于准确率、算法迭代收敛次数等指标），并分析结果中出现差异的可能原因，给出使用MindSpore的心得和建议。
- （加分项）使用MindSpore平台提供的相似任务数据集（例如，其他的分类任务数据集）测试自己独立实现的算法并与MindSpore平台上的官方实现算法进行对比，并进一步分析差异及其成因。

子课题2-KNN分类任务

- 概述

- 利用KNN算法，对 Iris 鸢尾花数据集中的测试集进行分类。

- 任务说明

- 利用欧式距离作为KNN算法的度量函数，对测试集进行分类。实验报告中，要求在验证集上分析近邻数 k 对KNN算法分类精度的影响。
- 利用马氏距离作为KNN算法的度量函数，对测试集进行分类。马氏距离是一种可学习的度量函数，详情请参考课题说明文档。注意：实验报告中请对优化过程的梯度计算公式进行推导。

子课题2-KNN分类任务

- 任务说明

- 基于MindSpore平台提供的官方模型库，对相同的数据集进行训练，并与自己独立实现的算法对比结果（包括但不限于准确率、算法迭代收敛次数等指标），并分析结果中出现差异的可能原因，给出使用MindSpore的心得和建议。
- （加分项）使用MindSpore平台提供的相似任务数据集（例如，其他的分类任务数据集）测试自己独立实现的算法并与MindSpore平台上的官方实现算法进行对比，并进一步分析差异及其成因。

子课题3-朴素贝叶斯分类实验

- 概述

- 利用朴素贝叶斯算法，对MNIST数据集种的测试集进行分类。

- 任务说明

- 在课程学习中同学们已经学习了贝叶斯分类理论并掌握了其基本原理，即利用贝叶斯公式 $p(\omega_j|x) = \frac{p(x|\omega_j)p(\omega_j)}{p(x)}$ ，对 $p(\omega_j|x)$ 作出预测，由于 $p(x)$ 为一固定值，所以一般不在计算过程中求得 $p(x)$ 的具体值。在实际运用中，为了方便计算，通常假设数据特征之间相互独立，即 $p(x|\omega_j) = p(x_1|\omega_j) \cdot p(x_2|\omega_j) \dots \cdot p(x_d|\omega_j)$ ， $x \in \mathbb{R}^d$ ，这便是著名的朴素贝叶斯算法。

子课题3-朴素贝叶斯分类实验

- 任务说明

- MNIST数据集本身以二进制形式保存，所以首先需要选择合适的编程语言编写读写二进制数据的程序完成对图片、标记信息的初步提取工作。读取了图片信息后，发现每个像素点的值在 $[0,1]$ 区间中，这是图像压缩后的结果，所以可以先将像素值乘以255再取整，得到每一个点的灰度值。将图像二值化，得到可以用于分类的 28×28 个特征向量以及对应的标签数据，之后便可以交由贝叶斯分类器进行学习。
- 基于MindSpore平台提供的官方模型库，对相同的数据集进行训练，并与自己独立实现的算法对比结果（包括但不限于准确率、算法迭代收敛次数等指标），并分析结果中出现差异的可能原因，给出使用MindSpore的心得和建议。

子课题3-朴素贝叶斯分类实验

- 任务说明

- （加分项）使用MindSpore平台提供的相似任务数据集（例如，其他的分类任务数据集）测试自己独立实现的算法并与MindSpore平台上的官方实现算法进行对比，并进一步分析差异及其成因。

子课题4-隐马尔科夫模型实验

- 概述

- 利用隐马尔科夫模型，对语料库的中文语句进行分词。
 -

- 任务说明

- 中文信息处理是自然语言处理的分支。和大部分西方语言不同，书面汉语的词语之间没有明显的空格标记，句子是以字串的形式出现。因此对中文信息进行处理的第一步就是进行分词，将字串（**character string**）转变成词串（**word string**）。
- 依据字在词语中位置，为每个字赋予不同的状态（如句子开始、句子中间、句子结尾、单字成词等），将输入的中文句子转化为状态序列。

子课题4-隐马尔科夫模型实验

• 任务说明

- 在训练集上统计语料信息，训练隐马尔可夫模型，对测试集中的中文句子进行分词测试，并选取部分实验结果进行分析。
- 基于MindSpore平台提供的官方模型库，对相同的数据集进行训练，并与自己独立实现的算法对比结果（包括但不限于准确率、算法迭代收敛次数等指标），并分析结果中出现差异的可能原因，给出使用MindSpore的心得和建议。
- （加分项）使用MindSpore平台提供的相似任务数据集（例如，其他的分类任务数据集）测试自己独立实现的算法并与MindSpore平台上的官方实现算法进行对比，并进一步分析差异及其成因。

子课题5-神经网络图像分类任务

- 概述

- 利用神经网络算法，对MNIST数据集中的测试集进行分类。

- 实验内容

- 基于神经网络模型及BP算法，根据训练集中的数据对你设计的神经网络模型进行训练，随后对给定的打乱的测试集中的数据进行分类。
 - 基于MindSpore平台提供的官方模型库，对相同的数据集进行训练，并与自己独立实现的算法对比结果（包括但不限于准确率、算法迭代收敛次数等指标），并分析结果中出现差异的可能原因。

子课题5-神经网络图像分类任务

- 实验内容

- （加分项）使用MindSpore平台提供的相似任务数据集（例如，其他的分类任务数据集）测试自己独立实现的算法并与MindSpore平台上的官方实现算法进行对比，并进一步分析差异及其成因。

实验报告要求

- **内容要求**

- 实验报告包含3个子课题的实验内容与一个总的心得体会（谈一谈你对这门课或者实验的感悟与体会）
- 每个子课题实验内容分为3个主要部分：
 1. **问题描述：**描述该实验的问题并进行分析。
 2. **实现步骤与流程：**介绍算法实现的方法、步骤与流程（如：实验思路、数学模型、关键难点、算法步骤分解）。
 3. **实验结果与分析：**评估与分析你的算法，进行必要的实验图表绘制。根据算法谈一谈可以改进的方法。

实验报告要求

- 格式要求

- 报告正文为宋体小四号字，行距20磅。
- 电子版请转换为pdf版本进行提交。
- 纸质版请在封面的签名和时间处进行手写签字。
- 电子版命名方式为“学号-姓名”。
- 具体格式请参考给定的报告模板。