

2023 年《机器学习与模式识别》实验及大作业

一、编程实践（50 分）

实验要求如下：（推荐使用 Scikit-learn 和 numpy 工具包。）

- 1) 请同学选择一种常用的手写数字数据集（比如：SK-Learn 自带玩具数据集或 MNIST 数据集等）。要求：从分类算法中，选择至少一种算法来进行编程实践。最好有两种以上的进行对比分析。比如：选择逻辑回归，感知器，贝叶斯进行手写数字的分类。完成两个及以上算法进行对比实验有较多的加分。
- 2) 对该算法的超参数进行基本的调优和选择。如果能够使用 GridSearchCV 和 K-Fold 工具进行参数寻优将有加分。
- 3) 能正确使用训练集和测试集，对分类的结果进行适当的分析（比如：精度，过拟合或欠拟合的情况等）。如果能够使用模型评价工具，比如：confusion matrix，分类器报告。
- 4) 在实验的总结中，以要点的形式列举出自己的主要工作内容（哪些部分是自己所参考的代码，哪些部分是自己做出修改的代码）
- 5) 该实验报告主要包括：实验内容的简述，实验代码（含必要的注释），结果分析及总结四个部分。

二、扩展实验部分（30 分）

1) 要求 1：自己拍摄手写数字的图片（可以是只包含单个数字的图像）。根据前面训练得到的分类器，自行设计数据预处理流程，达到分类器所要求的输入图像大小和灰度等级的要求，生成图像的特征（8 分）。应用上一题中自己训练好的分类器完成至少 3 种手写数字的识别实验（6 分）。给出实验代码和结果分析。

2) 要求 2：引入 PCA，研究特征选择过程对不同的分类器算法的影响（5 分），并且分享在识别自己拍摄的手写数字图片中的经验。（5 分）

3) 要求 3：以 PCA 为基础，在二维特征空间进行十种手写数字的特征的可视化，并进行分析（6 分）。

说明：MNIST 数据集中的图像都是 28x28 的 256 阶灰度图，即灰度值 0 表示白色（背景），255 表示黑色（前景）。SK-Learn 自带玩具数据集中的图像是 8x8 的 17 阶灰度图，即灰度值 0 表示黑色（背景），16 表示白色（前景）。

三、强化学习的综述（20 分）

请参考周志华《机器学习》书中所介绍的强化学习算法。（综述字数至少 800 字）综述的范围：经典强化学习算法（Q-learning, Sarsa, DQN 等）。报告内容主要包括：算法的基本思路，基本的推理过程。（特别要求：报告中的公式必须使用公式编辑器输入。）

说明：算法的综述各占 15 分，排版格式的规范性占 5 分。

注意：需要大家提交大作业的源码，不能提供可执行的源码的部分进行相应的扣分。

本科课程论文评阅表

学号		班级		姓名	
课程名称	机器学习与模式识别		论文题目	算法实验及智能算法综述的大作业	

评阅点	评分标准	最高分	得分
编程实现 (80分)	正确、有新意、对教学中所学内容有较强的扩展或提升	80	
	正确、有一定新意、对教学中所给的示例做出一定变化	65	
	基本正确, 程序基本使用教学中给出的示例, 有调参, 有对比实验	50	
	错误较明显, 或程序与教学示例相同	40	
	基本完全错误, 程序无法执行	20	
	完全错误, 或没有完成本题	0	
算法综述 (15分)	论证严谨、思路清晰、逻辑性强、有较强综述能力, 引文准确	15	
	论证较严谨、思路较清晰、符合逻辑、论述正确, 引文准确	10	
	思路较清晰、论述正确、引文较恰当	8	
	有一定的条理性	5	
	紊乱、大段抄袭他人文章	0	
规范化 (5分)	格式完全符合规范, 字数完全符合要求	5	
	格式比较规范, 字数偏少	4	
	格式基本符合规范, 但有个别地方不合规, 字数较少	3	
	格式规范性尚可, 但不足之处较多, 字数太少	2	
	格式不规范或字数严重不足	0	
总分		100	

教师签名: _____

时间: 年 月 日

注: 1、此表的论点、论据、论证部分可以由学生根据自己论文内容填写, 教师参考“课程论文评阅标准”进行评阅。

2、此表按人填写, 与学生课程论文一并存放。

3、以文字和数字填写内容。