《人工智能导论》大作业

（二选一）

一、 入侵检测及数据投毒分析

1、目标

首先构建分类器来完成入侵流量检测的分类任务，区分不同网络包记录；之后生成坏数据样本对训练好的模型进行投毒攻击，破坏模型性能。包括：

- ✓ 二分类任务：数据标签只有正常/非正常两类
- ✓ 多分类任务：在二分类任务的基础上，非正常类包含许多不同的小类
- ✓ 投毒攻击：利用不正确或有偏斜的数据来影响模型，本任务具体过程为利用坏数据来训练模型，使得模型性能下降

2、模型

可选用**决策树**、**逻辑回归**、**随机森林**、支持向量机、深度神经网络

3、数据集

NSL-KDD，数据集中包含数据特征、标签等

4、评分标准：满分100分

（1）基本项：90分

- ✓ 对数据进行预处理，构建训练集、测试集（6分）
- ✓ 实现**决策树**、**逻辑回归**、**随机森林**3种模型（每种模型各10分，共30分）
- ✓ 训练并测试3种模型，实现二分类任务（每种模型各10分，共30分）
- ✓ 生成坏数据样本，实现对3种模型的投毒攻击并测试对模型性能的影响（每种模型各8分，共24分）

（2）加分项：10分

✓ 完成训练并测试**支持向量机或深度神经网络**，实现任意一种加5分

✓ 投毒攻击实施前后模型边界的可视化分析，加5分

5、实现要求

✓ **模型设计及其训练测试过程均需自己实现**

✓ 可借鉴部分现有库函数来实现

✓ 全部通过调用库函数实现，酌情扣除一定分数

✓ 对于雷同作业，经调查核实后，扣除一定分数

二、 手写数字识别及数据投毒分析

1、目标

首先构建分类器完成手写数字分类任务；之后生成坏数据样本对训练好的模型进行投毒攻击，破坏模型性能。包括：

✓ 多分类任务：包含10个数字，分成10类

✓ 投毒攻击：利用不正确或有偏斜的数据来影响模型，本任务具体过程为利用坏数据来训练模型，使得模型性能下降

2、模型

可选用**决策树、朴素贝叶斯算法、KNN**、支持向量机、多层感知机、卷积神经网络

3、数据集

MNIST，数据集中包含数据图片、标签等

4、评分标准：满分100分

（1）基本项：90分

- ✓ 对数据进行预处理，构建训练集、测试集（6分）
- ✓ 实现**决策树、朴素贝叶斯算法、KNN**3种模型（每种模型各10分，共30分）
- ✓ 训练并测试3种模型，实现手写数字识别任务（每种模型各10分，共30分）
- ✓ 生成坏数据样本，实现对3种模型的投毒攻击并测试对模型性能的影响（每种模型各8分，共24分）

（2）加分项（选择实现其一）：10分

- ✓ 实现**支持向量机、多层感知机或卷积神经网络**，完成训练及测试，加10分
- ✓ 利用FGSM/PGD算法生成对抗样本，可视化分析投毒攻击实施前后模型边界的变化，加10分

5、实现要求

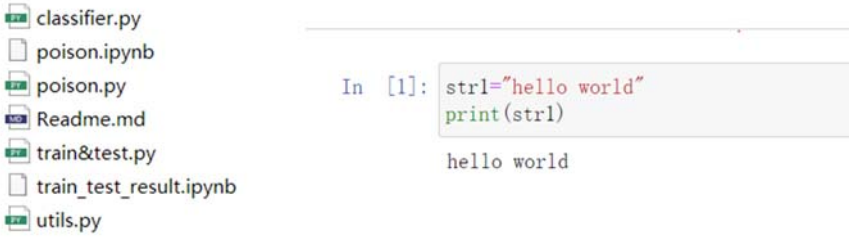
- ✓ **模型设计及其训练测试过程均需自己实现**
- ✓ 可借鉴部分现有库函数来实现
- ✓ 全部通过调用库函数实现，酌情扣除一定分数
- ✓ 对于雷同作业，经调查核实后，扣除一定分数

作业分组及提交

（1）最多5人一组，并由小组协商推选组长，组长负责任务统筹和确定成员贡献比例；

（2）提交的大作业文档大纲模板见“人工智能导论大作业模板.docx”，大作业命名：“人工智能导论大作业-组号.docx”；

（3）提交完整代码，包含的代码文件及运行结果示例如下：



（4）请于第 18 周结束（6 月 16 日）之前上传大作业电子版至 Canvas 上所建立的相关目录；打印版请交给课代表，由课代表统一收齐后交教师办公室。