

北京邮电大学 2023—2024 学年第一学期
《神经网络与深度学习》课程实验作业（一）

实验内容：深度学习基础

注意事项：

- ① 本次实验包含两道题，共计 20 分；
- ② 部分问题根据学号不同需进行不同的配置，若完成内容与实验要求不对应，则该项记为 0 分；
- ③ 所有实验结果需以实验报告的形式进行提交，文件命名格式：实验一_姓名_学号.pdf；
- ④ 实验报告中可插入代码片段，完整代码无需放在实验报告中，以压缩包的形式添加即可，压缩包命名格式：实验一代码_姓名_学号.zip；
- ⑤ 作业提交截止时间：2023 年 10 月 23 日 24: 00 前。

1. 机器学习：回归问题 (10 分)

皮马印第安人糖尿病数据集由年龄大于等于 21 岁的皮马印第安女性的已有诊断信息组成，包含若干医学预测变量和一个目标变量 **Outcome**，共九个字段。其中预测变量包括患者的怀孕次数、BMI、胰岛素水平、年龄等。

请运用回归模型分析糖尿病数据集中自变量和因变量之间的关系，对某人是否患糖尿病进行预测。

- (1) 使用特征工程方法对数据集中的特征变量进行处理并在实验报告中作出说明；(2 分)
- (2) 使用逻辑回归模型完成实验并绘制逻辑回归散点图；(2 分)
- (3) 使用 `train_test_split` 函数对数据集进行拆分，并在拆分得到的测试集上测试模型准确率。学号为奇数的同学请设置 `train_size` 为 0.7、0.75、0.8，学号为偶数的同学请设置 `train_size` 为 0.75、0.8、0.85；(2 分)
- (4) 试分析多个特征值（自变量）与病情（因变量）的关系(2 分)，依据与病情的关联性对自变量进行排序并进行可视化展示(2 分)。(共计 4 分)

2. 多层感知机：分类问题 (10 分)

CIFAR-10 数据集是深度学习中常用的数据集，其包含 60000 张 32×32 色图像，分为

10 个类，每类 6000 张。有 50000 张训练图片和 10000 张测试图片。

请基于该数据集搭建包含三个及以上的全连接层的多层感知机网络，以解决 10 分类问题。

- (1) 输出网络结构；(1 分)
- (2) 使用 `tensorboard` 对训练过程中的 `loss` 和 `accuracy` 进行可视化展示；(2 分)
- (3) 保存训练模型为模型文件，并使用训练好的模型对测试集中的图片进行预测，输出预测结果与预测概率；(2 分)
- (4) 画出训练集和验证集的混淆矩阵；(2 分)
- (5) 分析网络参数（例如网络深度、不同的激活函数、神经元数量等）对预测结果的影响；(2 分)
- (6) 在损失函数为交叉熵的情况下，对比网络最后一层是否使用 `softmax` 的性能差异并分析其产生的原因。(1 分)