

第四次编程

主讲老师：江 瑞

学生：姓名（学号）

1 题目介绍

鸢尾花卉数据集中的每一个样本包含四个属性的特征：花萼长度（sepal length）、花萼宽度（sepal width）、花瓣长度（petal length）以及花瓣宽度（petal width）。该数据集上的任务为预测鸢尾花卉属于山鸢尾（Setosa）、杂色鸢尾（Versicolour）、维吉尼亚鸢尾（Viginica）三个种类中的哪一类。

在上一次的编程作业中，我们只使用了花瓣长度特征，搭建了一元 Logistic 回归模型，解决了数据集中山鸢尾和维吉尼亚鸢尾的二分类问题。

在本次编程作业中，我们将尝试使用花卉的原始多元特征，在完整数据集上求解三分类问题。

2 编程要求

请在下列模型中**任选其一**，使用原始四维特征作为模型输入，求解鸢尾花卉三分类问题，完成本次练习：

1. 搭建 Softmax 回归模型，以最小化交叉熵损失为优化准则，使用随机梯度下降求解本问题；
2. 搭建全连接前馈神经网络，同样使用交叉熵损失作为损失函数，使用随机梯度下降求解本问题。

我们推荐大家使用[PyTorch](#)机器学习框架完成本次作业。你也可以使用上次作业中我们已经使用过的[scikit-learn](#)，或其他主流机器学习框架，例如[PyTorch](#)、[TensorFlow](#)、[Keras](#)。请你自行选定一种框架进行学习，并使用该框架搭建相应的模型，完成本次编程练习。请你挑选合适的评价指标评价模型优劣。

3 提交说明

为方便起见，脚本 `code.py` 中已经为你实现了数据加载、特征提取和数据集划分的相关代码，你可以根据自己的需要使用、改进。

请提交完成的代码文件及实验报告。

本次作业截止日期：**2022 年 12 月 7 日（周三）晚 12 点（两周后）**