

实验一 数据可视化技术

南理工泰州科技学院

叶志鹏

2023 年 3 月 12 日

1 实验目标

- 回顾 Python 基础知识
- 掌握 Python 数据读写
- 掌握 Python 面向对象的概念
- 掌握 Python 模块与包的概念

2 实验要求

- 自学并巩固 Python 基础知识。
- 学术诚信，抄袭零分。
- 3 月 26 日晚 12 点之前完成，请参考相关超时惩罚机制，特殊情况除外。
- 电子报告,格式工整(代码不建议大面积截图),pdf 提交。可参考 Word,Markdown, Latex 编辑器等，可参考模板。
- 报告内容，包括清晰的描述实验步骤的结果以及结合理论课内容解释说明（为什么会有这样的结果，what, how, why）。贴上重要代与图片截图。

- 将代码，实验报告，图片结果，数据等一并打包成压缩包 (.zip)，文件夹与实验报告命名为班级_姓名_学号，通过学习通上传。

3 Python 相关库

请熟悉下列 Python 相关库，将有利于完成实验内容。

Numpy, Matplotlib 等

官网文档：Numpy: <https://numpy.org/>; Matplotlib: <https://matplotlib.org/>。提示：允许调库，但是请了解各个库的使用说明，不然很容易出错。如果你可以正确的自己实现库里面的算法，会有相应加分！

4 任务

4.1 Iris 数据集的数据分析 50'

本部分将对 Iris 数据集进行可视化并分析，请完成以下几个任务，数据集 iris.data 在压缩包中。

- 读取 iris.data 数据集到内存中，并存储为列表命名为 iris_list，里面的元素可以是自定义对象，也可以是 2 维列表，并输出 iris.data 有多少个样本个数，以及有多少种类别。10'
- 按照第一列属性（sepal length 花萼长度）将上步操作得到的列表升序排序并打印结果。5'
- 实现一个 Python 函数，能够实现对 iris_list 的有放回随机抽样，函数参数为抽样列表 data，抽样个数 number，并测试打印结果。10'

```
1 def sampling_with_replacement(data, number):  
2     # todo
```

- 实现一个 Python 函数，能够实现对 iris_list 的无放回随机抽样，函数参数为抽样列表 data，抽样个数 number，并测试打印结果。10’

```
1     def sampling_without_replacement(data, number):
2         # todo
```

- 统计 iris 各列属性均值，方差，标准差，中位数并打印输出。15 ‘

4.2 实现 Vector 类 50’

- 实现 Vector 类，并完成 __init__(self, data) 方法。10’

```
1     class Vector:
2         def __init__(self, data):
3             # todo
```

- 实现 __len__(self) 特殊方法，能够通过 len(vector) 获取到向量的维度。10’
- 实现 __str__(self) 特殊方法，能够通过 print(vector) 获取到向量的元素。10’
- 实现向量的加法运算，如 $\text{vec3} = \text{vec1} + \text{vec2}$ 。10’
- 实现向量的内积运算, $\text{scale} = \vec{x_1} * \vec{x_2} = x_1^T * x_2$ 。10’