**一、组员：蔡焕斑**

**二、题目**

基于**实验一**中清洗后的数据练习统计和视化操作，100个同学（样本），每个同学有11门课程的成绩（11维的向量）；那么构成了一个100x11的数据矩阵。以你擅长的语言C/C++/Java/Python/Matlab，编程计算：

1. 请以课程1成绩为x轴，体能成绩为y轴，画出散点图。
2. 以5分为间隔，画出课程1的成绩直方图。
3. 对每门成绩进行z-score归一化，得到归一化的数据矩阵。
4. 计算出100x100的相关矩阵，并可视化出混淆矩阵。（为避免歧义，这里“协相关矩阵”进一步细化更正为100x100的相关矩阵，100为学生样本数目，视实际情况而定）
5. 根据相关矩阵，找到距离每个样本最近的三个样本，得到100x3的矩阵（每一行为对应三个样本的ID）输出到txt文件中，以\t,\n间隔。

#### 提示：

计算部分不能调用库函数；画图/可视化显示可可视化工具或API实现。

**三、作业说明**

①文件说明

压缩包中包含10个文件

final\_data.xlsx为实验一处理后得到的数据

README.docx为说明文件

其它文件为各个小题的代码、截图或结果，看文件名即可清楚

②涉及的技术

将excel文件的数据存入列表

将列表中的数据导出到excel文件

**四、难题与解决**

一开始打算完全使用java来完成实验二，但在完成第1小题绘制出散点图之后，我发现没办法使用java来完成第2小题(画出的图达不到想要的效果)，后面的题目(第2-4小题)就转向使用python来完成。因为对python的使用不是特别熟悉，所以后面的几个小题就参考了肖小恩同学的实验代码，这里附上她的代码地址：<https://github.com/tiger-xiao31/Machine_Learning--experiment/tree/main>，表示感谢

这个实验的难题就在于我对python的一些函数不是很了解，需要花费一些时间去查看相关函数的作用，以及函数需要传递的参数的类型，实验过程中一直报错，不停地进行调试，最后终于完成了。整个实验我完成了第1-4小题，第5小题和第4小题的混淆矩阵可视化存在一些问题，没能顺利完成。

**五、总结**

整个实验下来，我学会了使用java的jfree包进行相关的绘图操作。同时，实验涉及到对python语言的使用，加深了我对python的认识与使用。

通过本次实验，我学到了Z-score归一化的公式及方法，学会了如何调用函数去画散点图和直方图，同时知道了如何计算相关矩阵、相关系数。在计算的过程我也对公式有了更深入的理解、对公式的运用能力有了增强。总的来说，收获颇多。