

# 计算机科学与技术学院大数据分析与实践课程实验报告

实验题目：手机数据采集与分析实践		学号：202322130197
日期：	班级：数据	姓名：崔嘉铭
Email：cjm13969665900@gmail.com		

## 实验目的：

本实验旨在利用智能手机传感器数据，对人体日常行为进行分析与可视化。通过对加速度与角速度等传感器数据的处理，探索不同人体行为在数据层面的差异特征，理解手机传感器数据与真实行为模式之间的对应关系，从而加深对现实世界数据分析流程的认识。

## 实验软件和硬件环境：

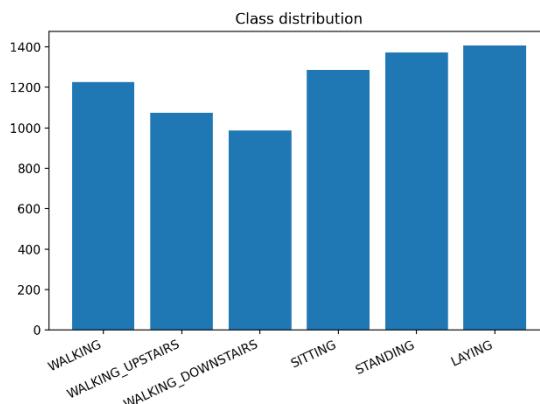
实验在 Windows 操作系统下完成，使用 Python 作为主要分析语言。  
数据处理与分析主要依赖 numpy、pandas 等科学计算库，可视化部分使用 matplotlib 库完成，行为分类模型使用 scikit-learn 实现。  
数据来源为 UCI Machine Learning Repository 提供的 Human Activity Recognition Using Smartphones 数据集

## 实验原理和方法：

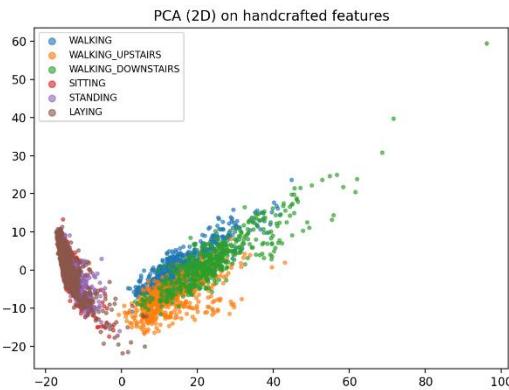
本实验使用 UCI HAR 数据集中已经提取好的时域与频域特征作为分析对象。首先通过统计不同活动类别的样本分布情况，了解数据的整体结构；随后采用主成分分析（PCA）方法对高维特征进行降维，以观察不同人体行为在低维空间中的分布差异；最后使用逻辑回归模型对行为类别进行分类，以定量方式验证不同活动在特征层面的可区分性。

## 实验步骤：(不要求罗列完整源代码)

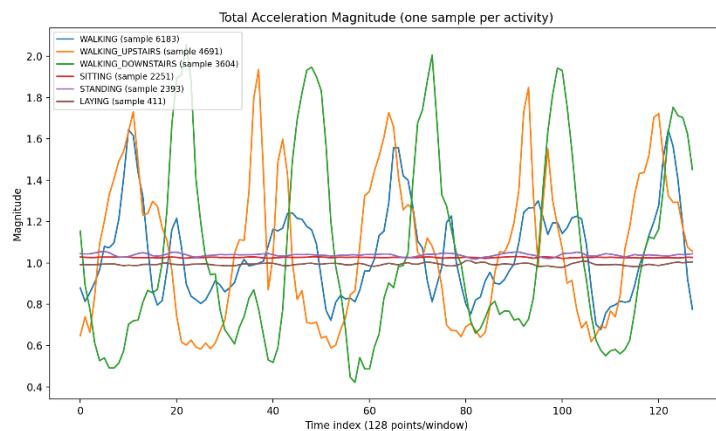
实验首先读取 UCI HAR 数据集中训练集与测试集的特征数据和对应标签，对数据维度和样本数量进行检查，确认数据完整性。



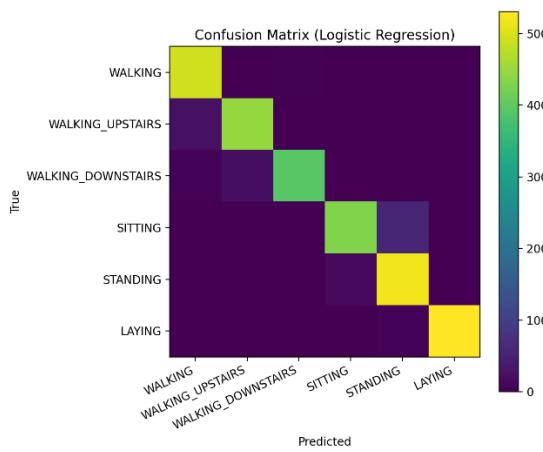
从图中可以看出，不同行为类别在样本数量上分布较为均衡，为后续分析与建模提供了良好的数据基础。



通过 PCA 将高维特征映射至二维空间，可以观察到动态行为（如 Walking、Upstairs、Downstairs）与静态行为（如 Sitting、Standing、Laying）在分布上存在明显差异，说明传感器特征能够有效反映不同人体行为模式。



从加速度幅值随时间变化的曲线可以看出，动态行为具有明显的周期性和较大的波动幅度，而静态行为的曲线相对平稳，这与现实中的人体运动情况相一致。



使用逻辑回归模型对行为进行分类，测试集上取得了较高的分类准确率，混淆矩阵显示大多数行为类别能够被正确识别，进一步验证了传感器特征在区分人体行为方面的有效性。

### 结论分析与体会：

通过对手机传感器数据的分析与可视化，可以发现不同人体行为在加速度等特征上具有明显差异。动态行为表现出较强的波动性和周期性，而静态行为相对平稳。实验结果表明，手机传感器数据能够较好地反映人体行为特征，简单的机器学习模型即可实现有效区分。本实验使我对真实世界传感器数据的分析流程有了更加直观和深入的理解。

就实验过程中遇到和出现的问题，你是如何解决和处理的，自拟 1—3 道问答题：

**Q：为什么需要对高维特征进行 PCA 降维？**

**A：**高维特征不便于直接可视化，通过 PCA 可以在保留主要信息的前提下将数据映射到低维空间，更直观地观察不同行为的分布差异。