

## 实验 2：聚类与分类实验指南

### 一、实验目的

1. 通过实验了解聚类与分类的应用场景；
2. 熟练掌握至少两种聚类算法的原理，并了解不同算法的优劣势；
3. 熟练掌握至少一种分类算法的原理。

### 二、实验步骤

#### 1. Python 环境安装

##### 建议安装 python3

##### 1. 安装原生 Python 环境

直接在官网下载对应平台（Linux、Windows、Mac OS）最新版本的 python，参考文档和教程安装。

Python 官网：<https://www.python.org>

Python 文档地址：<https://docs.python.org/3/>

##### 2. 安装 Anaconda（推荐！）

Anaconda 提供了很多基础依赖库，并支持自己创建多个不同虚拟环境，且通过 conda install 命令能自动安装相关依赖库，十分方便。

Anaconda 下载地址 <https://www.anaconda.com/distribution/> 下载后直接根据提示安装即可。

检验安装成功：在命令行输入 `python --version` 能够看到安装的对版本 python。  
或者输入 `conda` 可以看到 Anaconda 相应的提示即可。

#### 2. 聚类算法实现

##### 1. clustering 文件夹下的 todo.py 文件部分：

- 在 kmeans 函数下实现 k-means 或 k-medoids 算法
- 在 spectral 函数下实现谱聚类算法

##### 2. clustering 文件夹下的 clustering.py 文件部分

- 调试 knn\_graph 的参数，优化谱聚类的效果
- 运行 clustering.py 文件，保存两种算法的聚类结果，比较分析其不同

#### 3. 分类算法实现

##### 1. classification 文件夹下 todo.py 文件部分：

- 在 func 函数下实现线性回归（LR）、支持向量机（SVM）和感知机（Perceptron）三种分类算法的其中一种

##### 2. 在 classification 文件夹下 classification.py 文件部分：

- 选择划分训练集和测试集的比例

- 编写计算训练和测试错误率的代码
- 保存分类结果, 报告训练和测试错误率

加分项: 实现多个分类算法, 比较分析其异同

### 三、提交内容

1. 实验报告: 包含上述第二、第三项实验内容。
2. 实验代码: 第二、第三项实验的完整代码。

实验 3 提交方式: “学在浙大” 平台