

# 聚类实验一

一、实验目的：输入数据集，完成数据的聚类。

二、实验方法：K-MEANS、EM 聚类

三、实验要求：

1.数据集采用 Iris 数据集，这个数据集共有 150 个数据，每个类别有 50 个数据。它的属性信息如下：

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
  - Iris Setosa
  - Iris Versicolour
  - Iris Virginica

属性信息中的 class 是数据的类别标签，因此这个数据集经常用于分类问题。在这个聚类实验中把 class 属性去掉，每个数据只有前 4 个属性，请根据这 4 个属性将 150 个数据进行聚类。

在 Iris 数据集存放在 Iris 文件夹中，iris.names 是文件属性的说明，iris.data 是 150 个数据。

2. 聚类的评价指标，可以选择第一章绪论中讲到的指标进行评价聚类结果。

3. 实验方法采用 K-MEANS 和 EM 聚类。

4.提交实验报告和原代码，请打包成一个文件提交。实验报告中至少包括具体的实验方法、实验步骤、实验结果、实验分析、以及实验心得。