

实验6

实验名称：K近邻分类

实验内容：用K近邻算法对给定的数据进行分类。

K近邻分类

实验内容：利用K近邻算法进行数据分类

- 利用已给的训练数据，对测试数据的每一条记录进行分类。
- K近邻分类原则：
 - ① 给定一个测试样本
 - ② 计算它到每个训练样本的距离
 - ③ 取离测试样本最近的K个训练样本
 - ④ “投票法”选出在这K个样本中出现最多的类别，作为测试样本的预测类别

Iris flower data set



1-Setosa



2-Versicolour



3-Virginica

Iris也称鸢尾花卉数据集，是一类多重变量分析的数据集。

每个数据包含4个属性。

可通过花萼长度，花萼宽度，花瓣长度，花瓣宽度4个属性预测鸢尾花卉属于（Setosa, Versicolour, Virginica）三个种类中的哪一类。

- 在训练样本中，给出每朵花的花萼长度，花萼宽度，花瓣长度，花瓣宽度，以及花的类别（1代表Setosa，2代表Versicolour，3代表Virginica；每个种类40个训练样本，总计 $40 \times 3 = 120$ 个训练样本）。
- 对于每一条测试样本（共31个测试样本），根据花萼长度，花萼宽度，花瓣长度，花瓣宽度，利用**K近邻算法**确定花的类别。
- 计算测试样本的识别率。

测试样本的识别率=正确识别的测试样本个数/测试样本总数

花萼长度	花萼宽度	花瓣长度	花瓣宽度	类别
5.1	3.5	1.4	0.2	1
4.9	3	1.4	0.2	1

实验数据说明

- 训练数据：iris-data-train.csv或iris-data-train.txt（两个文件内容一样）
- 测试数据：iris-data-test.csv或iris-data-test.txt（两个文件内容一样，最后一列“类别”仅用于比较输出结果和样本的真实类别是否一致）

实验报告内容:

- 1) 实验目的
- 2) 实验内容
- 3) 算法流程图（或伪代码）
- 4) 实验运行过程截图、实验结果
- 5) 实验过程中遇到的问题
- 6) 实验心得体会。

另外需要提交完整的源代码（需有注释说明，单独作为文件，不要放到实验报告word文档里面）。

注意事项:

- 1、可以相互讨论，但必须独立完成代码和实验报告，若与网上或者其他同学雷同，按不及格处理。
- 2、实验报告请使用老师提供的实验模板。实验报告命名：完整学号_姓名_AI_project6.doc;
例： 0304120101_张三_AI_project7.doc;

每个同学建立个人文件夹放实验报告和源代码，文件夹名“完整学号_姓名_AI_project6”。

- 3、报告提交时间：请在**第16周周三下午5点**前把报告发给课代表。