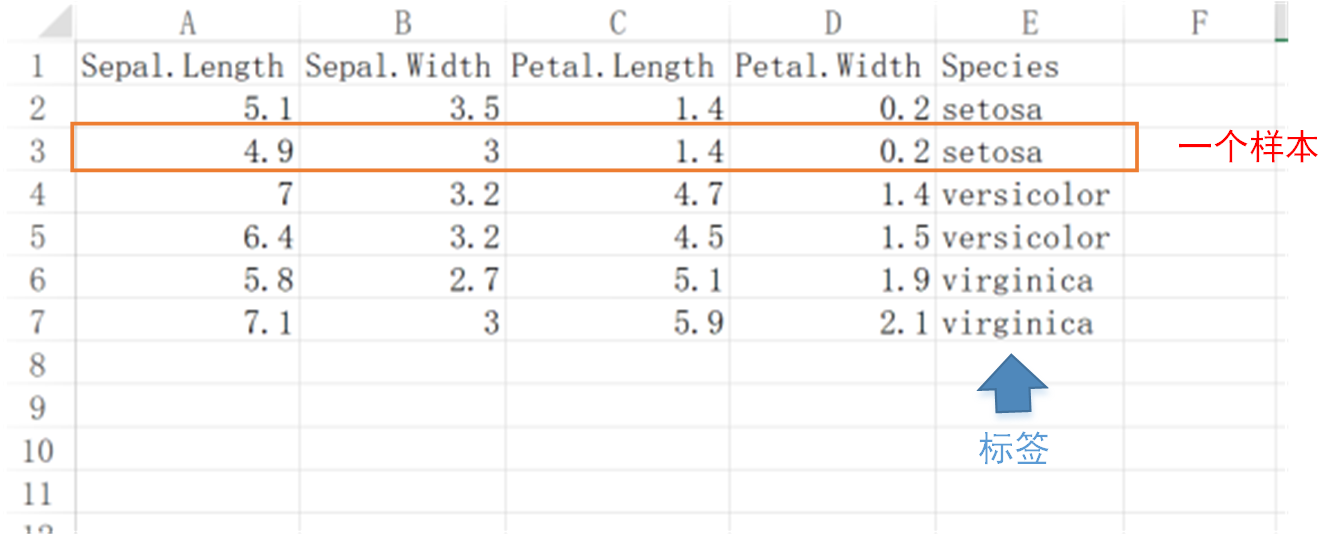
《大数据概论》程序设计要求

北方工业大学人工智能系张远 请勿上传网络传播

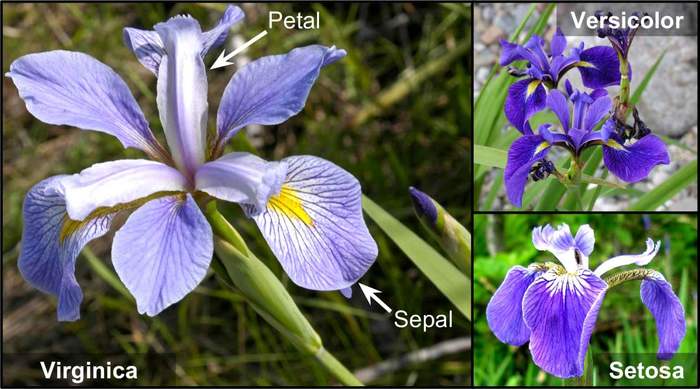
**1. 数据介绍**

数据为鸢尾花样本，分为training\_data.csv和test\_data.csv。training\_data是训练数据，包含80个样本。test\_data是测试数据，包含20个样本。

**2. 数据格式**



鸢尾花数据中，每一行代表一个样本。样本维度是4，第一维度是花萼长度(Sepal Length)，第二维度是花萼宽度(Sepal Width)，第三维度是花瓣长度(Petal Length)，第四维度是花瓣宽度(Petal Width)。这四个维度分别对应csv文件的列A B C D。最后列E是样本标签，即花的种类，包含两类setosa, versicolor。下图是实际的鸢尾花示意图。



**3. 提交作业要求**

选择一种机器学习算法，使用Python语言设计一个分类器，利用training\_data对分类器进行训练，利用test\_data对分类器进行测试。最终提交一个MS Word文档电子版，命名方式是：学号\_姓名\_班级\_年月日.doc，例如：10067\_张三\_电19-2\_20200102.doc 。这个文档应该包含如下内容：

* 完整的Python实现分类器代码，包括训练部分和测试部分，请使用小五号字
* 所使用的算法的文字说明，要与分类器对应，字数不限，请使用小五号字

将Word文档以附件形式发送到邮箱: [zhangyuanbuaa@qq.com](mailto:zhangyuanbuaa@qq.com) , 邮件题目和Word文档名字一致即可。

**4. 如何运行参考程序？**

文件夹里给出了一个感知机算法的参考程序perceptron.py。

首先安装python-3.8.3-amd64.exe，这是python语言环境，直接安装到默认路径即可。然后，打开windows命令行，进入程序所在目录，执行python perceptron.py