实验一 WEKA分类

一、实验目的

使用数据挖掘中的分类算法，对数据集进行分类训练并测试。应用不同的分类算法，比较他们之间的不同。与此同时了解Weka平台的基本功能与使用方法。

二、实验环境

实验采用Weka 3.8 平台，数据使用给定的鸢尾花xls格式的数据集iris.xls。 对于iris数据集，它包含了150个实例（每个分类包含50个实例），共有sepal length、sepal width、petal length、petal width和class五种属性。期中前四种属性为数值类型，class属性为分类属性，表示实例所对应的的类别。该数据集中的全部实例共可分为三类：Iris Setosa、Iris Versicolour和Iris Virginica。

实验所需的训练集和测试集均为iris.arff。

三、实验要求

1、将iris.xls转化为iris.arff。

2、应用iris数据集，分别采用KNN、C4.5决策树分类器和朴素贝叶斯分类器进行测试和评价，分别在训练数据上训练出分类模型，并对三个模型进行全面评价比较，得到一个最好的分类模型以及该模型所有设置的最优参数。

3、使用这些参数以及训练集和校验集数据一起构造出一个最优分类器，并利用该分类器对测试数据进行预测。

四、实验过程及结果

1. 实验总结