**实验二 Weka API的熟悉与应用**

**实验要求：**

1. Instances类
2. 读入bank.arff文件，并输出该数据集的样例数，属性数，类别数(注：最后一列为类别属性)
3. filters的相关类

以读入的bank数据集为基础，分别实现以下几个要求：

1. 利用Filter类和Discretize类将**所有numeric属性**（即age和income）离散化，将属性按值平均离散为10段，并且输出离散化的前20个样例；

报告时用截图

1. 利用Filter类和Normalize类将**所有numeric属性**（即age和income）归一化，归一的区间为[-1,1]，并且输出归一化的前20个样例；

要用到参数列表

1. 利用Filter类和Remove类删掉**第2和第3个属性**(即sex和region)，并且输出删除第2和第3个属性的前20个样例；
2. classifier的相关类和Evaluation类
3. 分别采用NaiveBayes，SMO，J48，1NN (KNN中k=1)分类器对bank数据集进行分类，在采用**10折交叉验证**的情况下，输出分类正确率，观察是否与Weka Explorer面板中调用这4种分类器得到的分类正确率相同(Weka Explorer中同样采用10折交叉验证) 对比一下是否一样；
4. 以U\_segmentation\_train数据集为训练集，以U\_segmentation\_test数据集为测试集，分别采用NaiveBayes，SMO，J48，1NN分类器，输出分类正确率，观察是否与Weka Explorer面板中调用这4种分类器得到的分类正确率相同(Weka Explorer中同样采用U\_segmentation\_test为测试集)；
5. 选做
6. 读sample.txt手动生成arff文件 先转成cvs格式 然后用weka的save as；或者手写
7. 采用KNN分类器对bank数据集进行分类，验证方式为10折交叉验证，在K分别选1，5，10，30的情况下，输出分类正确率，对这4种输出结果进行**比较和分析**。