

实验模式选择说明

在 11 月 5 日、11 月 6 日的实验课上，TA 将会向大家提供 Clementine、RExcel、Weka 的安装程序（Clementine 所需空间 1.03GB，RExcel 为 65.6MB），并向大家讲解决策树分类模型的代码实现注意事项等。请每位同学选择如下两种实验模式之一：

第一种实验模式为代码实现类。本门课将会讲授十余类人工智能/机器学习模型（代码实现的难度系数以*作为参考，*越多，通常会相对更难），包括 **k-最近邻**（kNN，可用于回归或分类，*）、**朴素贝叶斯**（NB，可用于回归或分类，*）、**决策树**（分为 ID3、C4.5、CART 三类，用于分类，**）、**感知机学习**（PLA，用于分类，**）、**神经网络**（NN，可用于回归或分类，****）、**逻辑回归**（LR，用于分类，***）、**支持向量机**（SVM，用于分类，****）、**关联规则挖掘**（Apriori，**）、**k-均值**（k-Means，用于聚类，*）、**DBSCAN**（用于聚类，**）。

【代码实现类的优秀标准】：实现了五类及以上的模型（如 kNN、NB、ID3、C4.5、CART），且与传统方法相比，有两类及以上的模型达到了更好的效果（如在 Lab 2 的 1000 篇测试文本上，实现的 kNN 模型比文档中的方法效果更好）。如果有同学想实现课堂讲授内容之外的其他模型，或者有评分疑问，请随时发邮件至 raoyangh@mail.sysu.edu.cn。

第二种实验模式为工具或代码应用类。基于 Clementine、RExcel、Weka 及相关的开源工具，以回归、分类、关联规则挖掘、聚类中的某类问题为研究目标，撰写实验报告。

【工具或代码应用类的优秀标准】：采用两种及以上的公开数据集，对选定的一类问题（如基于决策树的分类问题）分析全面，包含对这类问题的深入探讨（如决策树的多种剪枝方式及其实验结果的对比）。具体参考 Classification by Clementine.pdf、Classification by RExcel.pdf、Classification by Weka.pdf 这三篇优秀实验报告模板中的任意一篇。

注：如有同学对上述安排有任何疑问，或者想增加新的实验模式，请于 11 月 4 日晚上 10 点之前发邮件至 raoyangh@mail.sysu.edu.cn。谢谢！