

暨南大学本科实验报告专用纸

课程名称 数据挖掘 成绩评定
实验项目名称 Weka 软件环境的安装和使用 指导教师 韩旭明
实验项目编号 0806011601 实验项目类型 验证性 实验地点 DF116
学生姓名 陈宇 学号 2020101642
学院 信息科学技术学院 系 计算机 专业 软件工程
实验时间 2023 年 3 月 17 日 上午~3 月 17 日 上午 温度 °C 湿度

实验目的

1. 配置实验环境，安装 Weka 数据挖掘软件；
2. 了解 Weka 平台的基本功能与使用方法；
3. 初步了解 Weka 数据挖掘软件的数据预处理、分类、聚类、关联规则的使用方法。

实验原理

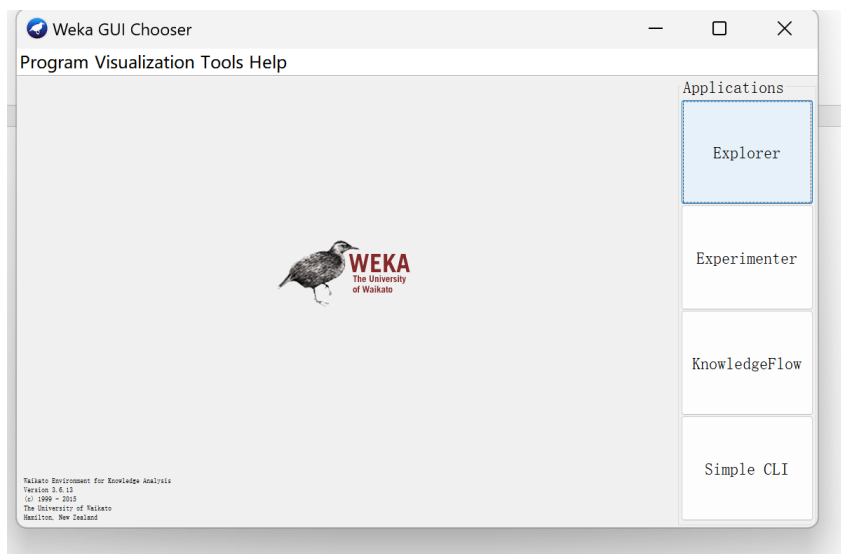
Windows 操作系统+Weka 数据挖掘软件+JRE (Java 运行环境)

实验内容和步骤

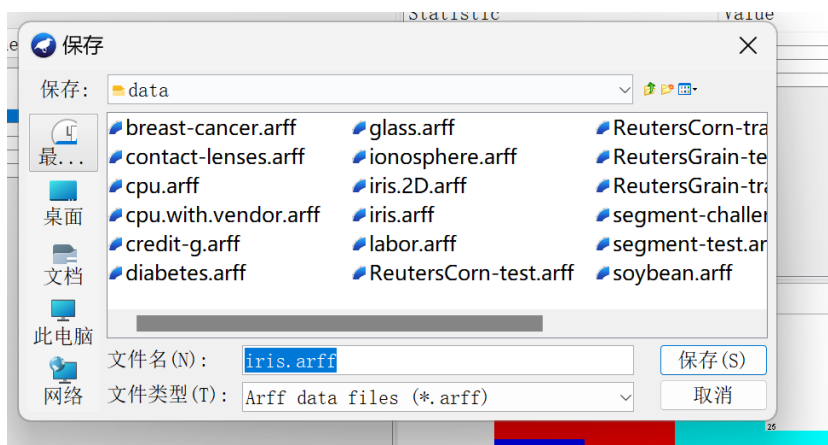
1. 实验环境的安装；
2. 掌握 Weka 的数据类型和数据文件 (arff)；
3. Weka 平台的基本功能与使用方法

实验过程

1. 实验环境的安装：安装并打开 weka 软件，如下图所示。



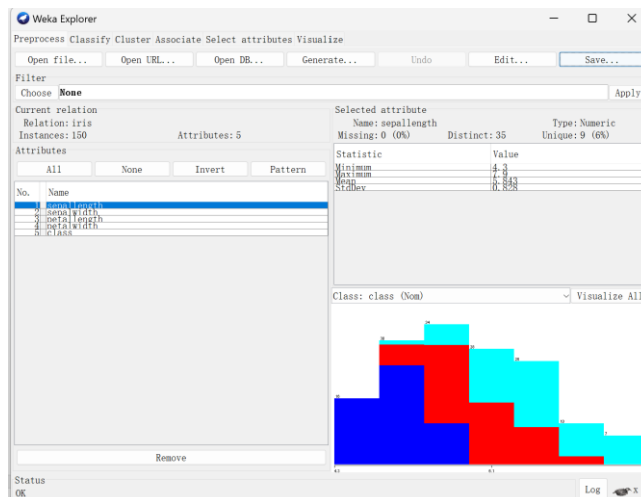
- 2.掌握 Weka 的数据类型和数据文件（arff）（将.csv 文件转换为.arff 文件）：
尝试将 csv 文件转换为 arff 文件如下图。



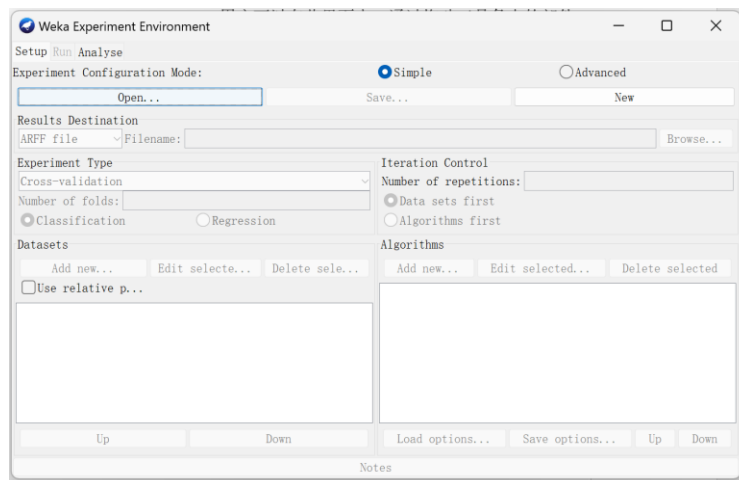
使用界面上的 save button 可以将 csv 文件存为 arff 文件。

3.Weka 平台的基本功能与使用方法：

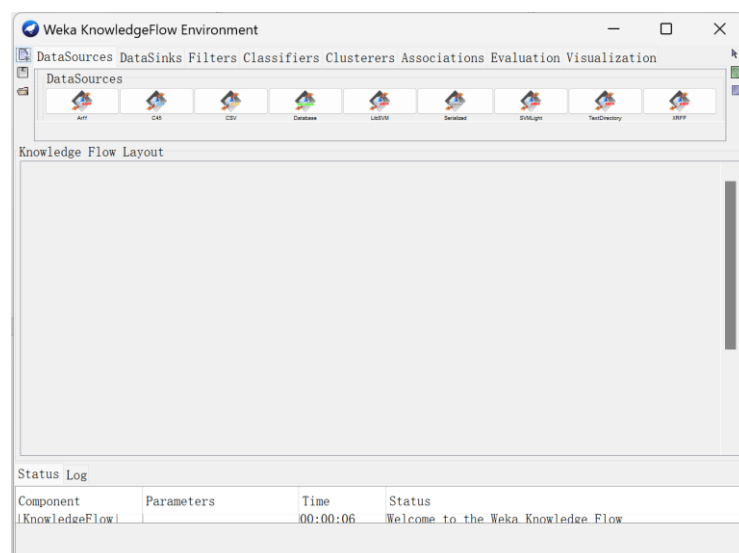
- 1) **explorer**: 数据挖掘常用界面，可以加载数据集，对数据进行预处理，选择 Weka 提供的各种数据挖掘算法和设置参数，执行数据挖掘，获得挖掘结果，并在整个过程进行可视化查看。



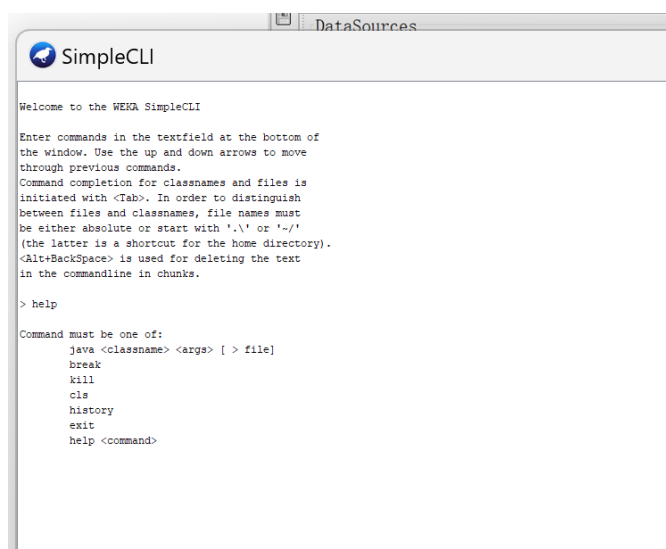
- 2) **Experimenter**: 用户可以在此界面中同时使用多个算法对一组数据进行分析, 并对各种算法性能比较, 从中选择最佳算法。



- 3) **KnowledgeFlow**: 用户可以在此界面中, 通过拖拽工具条中的部件, 将其放置在画布中。大型数据集可以被分批处理和读取。

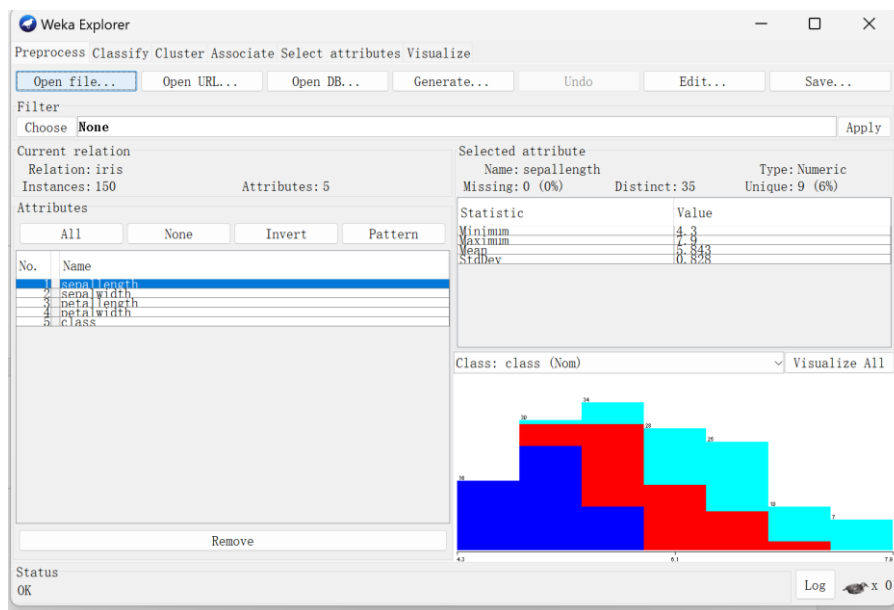


- 4) **SimpleCLI**: 其他三个界面的功能可以通过使用文本命令的方法来运行。



4. 数据预处理、分类、聚类、关联规则的使用方法。

1) 数据预处理：针对鸢尾花数据集



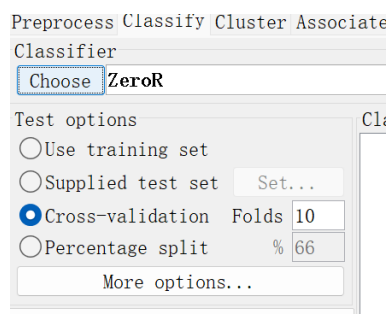
由图可知：数据实例有 150 个，属性数为 5。将 petal 离散化：

Weka Viewer interface showing the Iris dataset. The 'Relation: iris-weka.filters.supervised.attribute.Discretize-Rf...' is displayed. The table shows the following columns: No., sepal.length, sepal.width, petal.length, petal.width, and class. The 'petal' attribute is highlighted in blue.

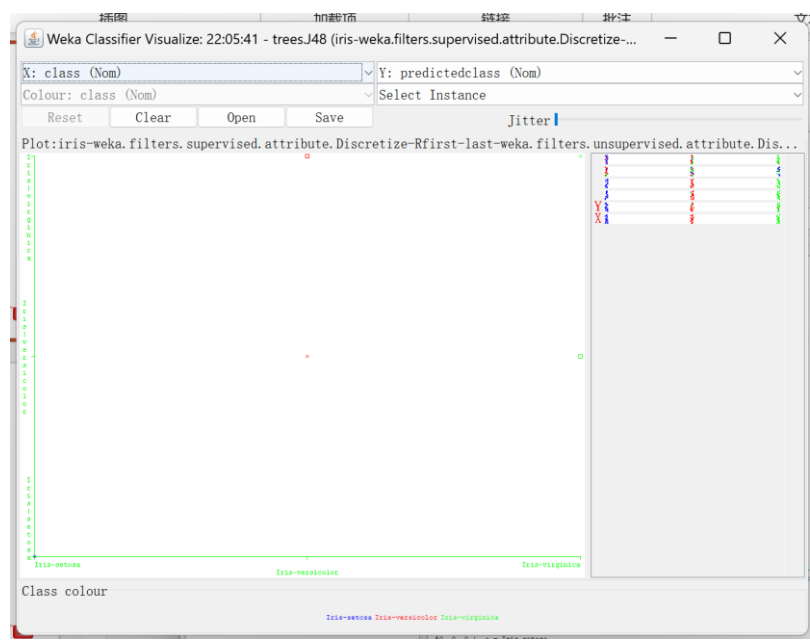
No.	sepal.length	sepal.width	petal.length	petal.width	class
1	(-inf, 2]	(3, 3.5]	(-inf, 2]	(-inf, 1]	r1
2	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
3	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
4	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
5	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
6	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
7	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
8	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
9	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
10	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
11	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
12	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
13	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
14	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
15	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
16	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
17	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
18	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
19	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
20	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
21	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
22	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
23	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
24	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
25	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
26	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
27	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
28	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
29	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
30	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
31	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
32	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
33	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
34	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
35	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
36	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
37	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
38	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
39	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
40	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
41	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
42	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
43	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
44	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
45	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
46	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
47	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
48	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
49	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
50	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
51	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
52	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
53	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
54	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
55	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
56	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
57	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
58	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
59	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
60	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
61	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
62	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
63	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
64	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
65	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
66	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
67	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
68	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
69	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
70	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
71	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
72	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
73	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
74	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
75	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
76	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
77	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
78	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
79	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
80	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
81	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
82	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
83	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
84	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
85	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
86	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
87	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
88	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
89	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
90	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
91	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
92	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
93	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
94	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
95	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
96	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
97	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
98	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
99	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
100	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
101	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
102	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
103	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
104	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
105	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
106	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
107	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
108	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
109	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
110	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
111	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
112	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
113	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
114	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
115	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
116	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
117	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
118	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
119	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
120	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
121	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
122	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
123	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
124	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
125	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
126	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
127	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
128	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
129	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
130	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
131	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
132	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
133	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
134	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
135	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
136	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
137	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
138	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
139	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
140	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
141	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
142	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
143	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
144	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
145	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
146	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
147	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
148	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
149	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1
150	(-inf, 2]	(2, 2.5]	(-inf, 2]	(-inf, 1]	r1

2) 分类：

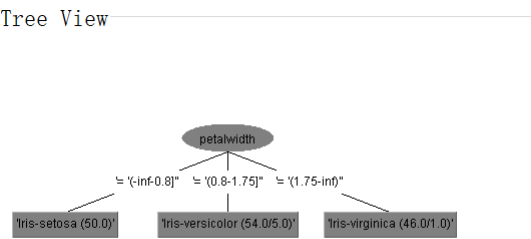
选择分类器如下图所示：



分类错误可视化:



树可视化:

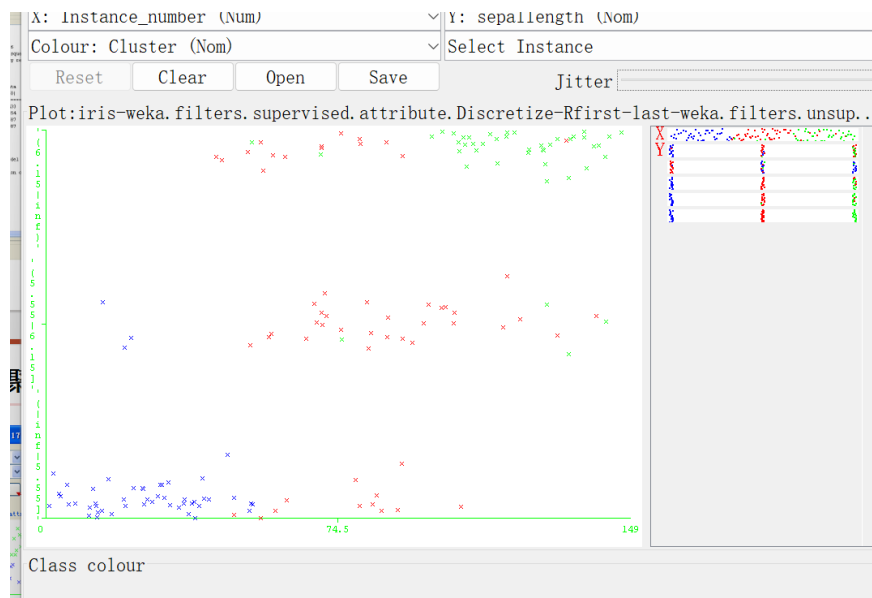


3) 聚类:

删除原有的标签类:

All		None		Invert		Pattern	
No.	Name						
1	sepal.length						
2	sepal.width						
3	petal.length						
4	petal.width						
5	class						

簇数设置为 3，可视化结果:



4) 关联分析:

对 weather-nominal 数据集进行关联分析:

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 conf:(1)
2. humidity=normal windy=FALSE 4 ==> play=yes 4 conf:(1)
3. outlook=sunny humidity=high 3 ==> play=no 3 conf:(1)
4. outlook=rainy windy=FALSE 3 ==> play=yes 3 conf:(1)
5. outlook=sunny humidity=normal 2 ==> play=yes 2 conf:(1)
6. outlook=sunny temperature=hot 2 ==> play=no 2 conf:(1)
7. outlook=overcast temperature=hot 2 ==> play=yes 2 conf:(1)
8. outlook=overcast humidity=high 2 ==> play=yes 2 conf:(1)
9. outlook=overcast humidity=normal 2 ==> play=yes 2 conf:(1)
10. outlook=overcast windy=TRUE 2 ==> play=yes 2 conf:(1)
11. outlook=overcast windy=FALSE 2 ==> play=yes 2 conf:(1)
12. outlook=rainy windy=TRUE 2 ==> play=no 2 conf:(1)
13. temperature=mild humidity=normal 2 ==> play=yes 2 conf:(1)
14. temperature=cool windy=FALSE 2 ==> play=yes 2 conf:(1)
15. outlook=sunny temperature=hot humidity=high 2 ==> play=no 2 conf:(1)
16. outlook=sunny humidity=high windy=FALSE 2 ==> play=no 2 conf:(1)
17. outlook=overcast temperature=hot windy=FALSE 2 ==> play=yes 2 conf:(1)
18. outlook=rainy temperature=mild windy=FALSE 2 ==> play=yes 2 conf:(1)
19. outlook=rainy humidity=normal windy=FALSE 2 ==> play=yes 2 conf:(1)
20. temperature=cool humidity=normal windy=FALSE 2 ==> play=yes 2 conf:(1)
21. humidity=normal 7 ==> play=yes 6 conf:(0.86)

实验总结

通过此次实验，在我自己的机器上配置好了实验环境；我也学习了 weka 数据挖掘软件的基本使用方法、基本功能，其中包括：数据预处理、分类、聚类、关联规则的使用方法。为接下来深入学习数据挖掘算法打下了基础，能让我更好的通过在 weka 软件上的实践，了解各种算法、各种功能。