

Descriptive induction

15/11/2007

Voting dataset

Iris dataset

Voting dataset

- 435 instances
- 16 attributes
 - 16 nominal attributes
 - 0 numeric attributes
- No target variable
- No missing values

The screenshot shows the Weka Explorer window for the 'voting' dataset. The 'Attributes' list includes 16 nominal attributes, with 'handicapped-infants' selected. The 'Selected attribute' panel shows the distribution: 'n' (236) and 'y' (187). A bar chart at the bottom right visualizes this distribution with two bars, one for 'n' (236) and one for 'y' (187), each split into blue and red segments.

Label	Count
n	236
y	187

Association rules

1

2

Weka Explorer

Preprocess | Classify | Cluster | **Associate** | Select attributes | Visualize

Associator

Choose **Apriori** -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0

Start Stop

Result list (right-click for o
16:25:34 - Apriori

Associator output

```
Size of set of large itemsets L(4): 1

Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> party=democrat 219   conf: (1)
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 198 ==> party=democrat 198   conf: (1)
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> party=democrat 210   conf: (1)
4. physician-fee-freeze=n education-spending=n 202 ==> party=democrat 201   conf: (1)
5. physician-fee-freeze=n 247 ==> party=democrat 245   conf: (0.99)
6. el-salvador-aid=n party=democrat 200 ==> aid-to-nicaraguan-contras=y 197   conf: (0.99)
7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204   conf: (0.98)
8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y party=democrat 203 ==> physician-fee-freeze=n 203   conf: (0.98)
9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> party=democrat 197   conf: (0.97)
10. aid-to-nicaraguan-contras=y party=democrat 218 ==> physician-fee-freeze=n 210   conf: (0.96)
```

Status
OK

Log x 0

Iris dataset

- 150 instances
- 4 attributes
 - 0 nominal attributes
 - 4 numeric attributes
- Nominal target variable
 - 3 values:
 - Iris-setosa (30%)
 - Iris-versicolor (30%)
 - Iris-virginica (30%)
- No missing values

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Ope... | Ope... | Ope... | Undo | Edit... | Sav...

Filter: Choose **None** Apply

Current relation: Relation: iris Instances: 150 Attributes: 5

Selected attribute: Name: iris Type: N... Missing: ... Distinct: Unique: 0...

Label	Count
Iris-setosa	50
Iris-versicolor	50
Iris-virginica	50

Class: iris (Nom) Visualize All

50 50 50

Status: OK Log x 0

Clustering

1



2



Weka Explorer

Preprocess | **Cluster** | Associate | Select attributes | Visualize

Clusterer

- weka
 - clusters
 - Cobweb
 - EM
 - FarthestFirst
 - MakeDensityBasedClusterer
 - SimpleKMeans**

Clusterer output

Number of iterations: 6
Within cluster sum of squared errors: 6.99811400482676

Cluster centroids:

weka.gui.GenericObjectEditor

weka.clusterers.SimpleKMeans

About

Cluster data using the k means algorithm More

numClusters

seed

Open... Save... OK Cancel

0	61	(41%)
1	50	(33%)
2	39	(26%)

17:12:13 - SimpleKMeans
17:12:22 - SimpleKMeans

Status: OK Log x 0

3



Clustering visualization

The image shows the Weka Clusterer Visualize interface. On the left, the main window displays the 'iris' dataset and a 'Cluster:' button. A context menu is open over the 'Cluster:' button, with 'Visualize cluster assignments' highlighted. The main window also shows 'Store clusters for visualization' checked, 'Ignore attributes', 'Start', and 'Stop' buttons. Below these are 'Result list' entries and a 'Status' bar.

The 'Weka Clusterer Visualize' window is open, showing the following settings:

- X: sepal length (Num)
- Y: petal width (Num)
- Colour: Cluster (Nom)
- Select Instance

Buttons for 'Reset', 'Clear', and 'Save' are visible. A 'Jitter' slider is also present.

The plot area shows a scatter plot of 'iris-weka.filters.unsupervised.attribute.Remove-R5_clustered'. The X-axis ranges from 4.3 to 7.9, and the Y-axis ranges from 0.1 to 2.5. Data points are colored according to their cluster assignment: cluster0 (red), cluster1 (blue), and cluster2 (green). A legend on the right shows the color mapping for 'X' and 'Y' axes.

At the bottom, the 'Class colour' section displays the legend: `cluster0 cluster1 cluster2`.