

Hand on Weka: Just a Taste

2012/11/20

Petra Kralj Novak
Petra.Kralj.Novak@ijs.si

Data Mining Tools

- Weka <http://www.cs.waikato.ac.nz/ml/weka/>
- Orange <http://orange.biolab.si/>
- Knime <http://www.knime.org/>
- Taverna <http://www.taverna.org.uk/>
- Rapid Miner <http://rapid-i.com/content/view/181/196/>
- CloudFlows <http://cloudflows.org/>

Weka (Waikato Environment for Knowledge Analysis)

- Collection of machine learning algorithms for data mining tasks
- The algorithms
 - Can be applied directly to a dataset
 - Can be called from Java code (library)
- Weka contains tools for
 - Data pre-processing
 - Classification
 - Regression
 - Clustering
 - Association rules
 - Visualization
- Weka is open source software issued under the GNU General Public License

Practice with Weka

1. Build a decision tree with the ID3 algorithm on the lenses dataset, evaluate on a separate test set

Weka: Install

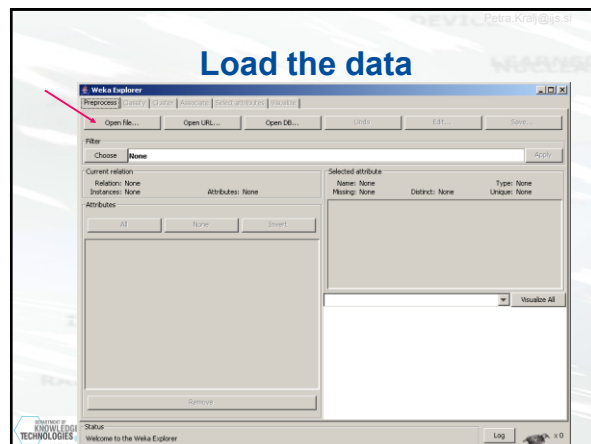
<http://www.cs.waikato.ac.nz/ml/weka/>

Weka: Run Explorer

Choose Explorer

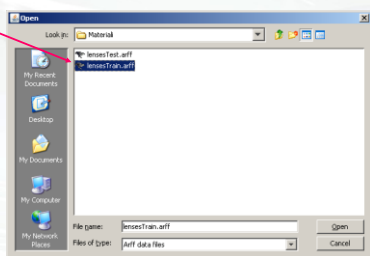
Exercise: Lenses dataset

- In the Weka data mining tool, induce a decision tree for the lenses dataset with the ID3 algorithm.
- Data:
 - lensesTrain.arff
 - lensesTest.arff
- Compare the outcome with the manually obtained results.



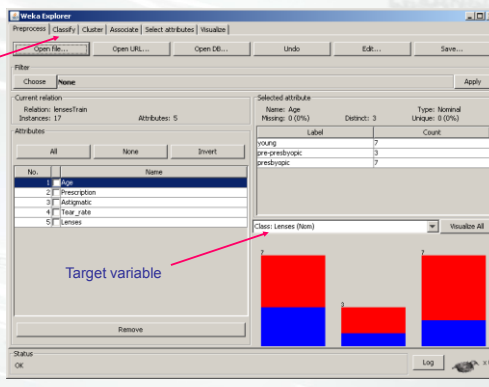
Load the data - 2

lensesTrain.arff



The data are loaded

Choose "Classify"



Choose algorithm

