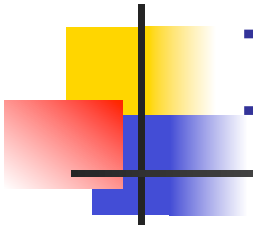
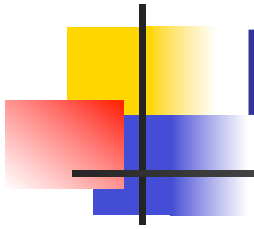
WEKA

By Joshua Hirtz



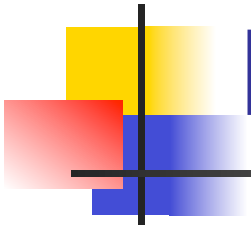
Introduction

- History
- Objectives
- Features
- Dataset
- Examples
- Problems
- SPSS Comparison
- Overview



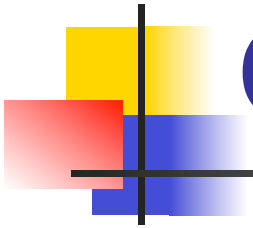
History

- Waikato Environment for Knowledge Analysis a.k.a. (WEKA)
- Created in New Zealand by the University of New Zealand's Computer Science Department



History (Cont.)

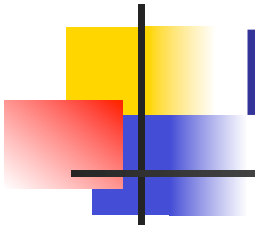
- Current versions
  - The book version is currently locked in 3-4 so that is may stay constant with the book
  - The developer version is currently in 3-5



Objectives

- Our objectives are to
  - make ML techniques generally available;
  - apply them to practical problems that matter to New Zealand industry;
  - develop new machine learning algorithms and give them to the world;
  - contribute to a theoretical framework for the field.

Taken from http://
www.cs.waikato.ac.nz/~ml/
index.html



Features

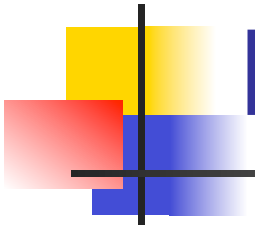
- CLI – offers a simple *Weka shell* with separated commandline and output.
- Explorer – an easy to use graphical user interface that harnesses the power of the Weka software.

Taken from http://
weka.sourceforge.net/wekadoc/
index.php/

Features (Cont.)

- Experimenter – enables the user to create, run, modify, and analyse experiments in a more convenient manner than is possible when processing the schemes individually.
- Knowledge flow - an alternative to the Explorer as a graphical front end to WEKA's core algorithms.

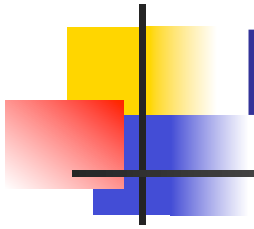
Taken from http://weka.sourceforge.net/wekadoc/index.php/



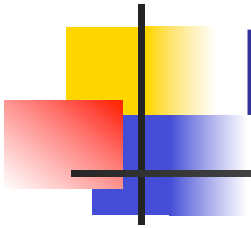
Dataset

- Uses the .arff extension
- @RELATION *name* – denotes the name of the file
- @ATTRIBUTE *name type*– denotes the name of the attribute
  - *type* consists of numeric, nominal, string, and date

Dataset (Cont.)

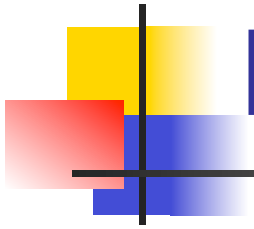


- @DATA – denotes the beginning of the data
- *data,data,data* – data is then entered with attributes separated by commas and different instances separated by lines



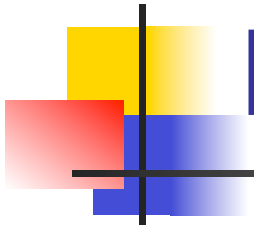
Dataset (Example)

```
@RELATION iris

@ATTRIBUTE sepallength  REAL
@ATTRIBUTE sepalwidth   REAL
@ATTRIBUTE petallength  REAL
@ATTRIBUTE petalwidth   REAL
@ATTRIBUTE class        {Iris-setosa,Iris-versicolor,Iris-virginica}

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
```



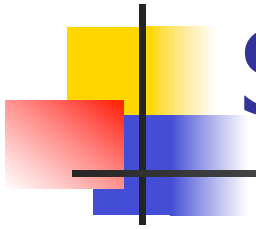
Examples

- A quick example with the Explorer and KnowledgeFlow to show how they work.



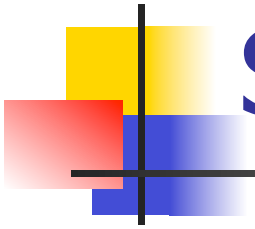
Problems

- Large datasets cause problems
- Data needs to be in main data for traditional algorithms.



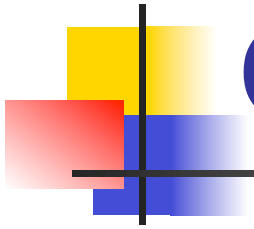
SPSS Comparison

- WEKA
  - GPU – General Public License
  - Problems with large datasets
  - Comes in a book and developer version
- SPSS – Clementine
  - Expensive
  - Created to handle large datasets
  - Comes in various versions to cover various environments
    - Base, Server, Batch, etc.



SPSS Comparison (Conclusion)

- WEKA is a cheaper solution for smaller datasets, however it lacks seems to lack the power, customer support, and system flexibility of SPSS Clementine.



Overview

- History
- Objectives
- Features
- Dataset
- Examples
- Problems
- SPSS Comparison

References

- Collaborated with John Aleshunas.
- *Weka Machine Learning Project*. (N.A.). Retrieved May 6, 2008, from http://www.cs.waikato.ac.nz/~ml/index.html
- *WEKA (Machine Learning)*. (May 3, 2008). Retrieved May 6, 2008, from http://en.wikipedia.org/wiki/WEKA
- *Frank, Eibe. (N.A.). Machine Learning with WEKA. Retrieved May 6, 2008, from* http://www.cs.waikato.ac.nz/ml/weka/

References (Cont.)

- *Pfahringer, Bernhard. (N.A.). Machine Learning with WEKA. Retrieved May 6, 2008, from* http://www.cs.waikato.ac.nz/ml/weka/
- *N.A. (N.A.). WEKA: Machine Learning and Data Mining as ClickandPlay. Retrieved May 6, 2008, from* http://www.google.com/search?q=cache:MeH2vRZYZ5EJ:www.informatik.uni-freiburg.de/~mlpult/slides/WEKA-1201.pdf+weka+pros+and+cons&hl=en&ct=clnk&cd=7&gl=us
- *Jakob, Michal. (N.A.). WEKA: Machine Learning & Softcomputing. Retrieved May 6, 2008, from* http://www.google.com/search?q=cache:dSWjfexxIDcJ:cyber.felk.cvut.cz/gerstner/teaching/ppdm/weka_lecture.ppt+weka+pros+and+cons&hl=en&ct=clnk&cd=6&gl=us

References (Cont.)

- *Scope Creep*. (April 28, 2008). Retrieved May 6, 2008, from http://en.wikipedia.org/wiki/Functionality_creep
- *Assessing Student Proficiency in a Reading Tutor that Listens*. (N.A.). Retrieved May 6, 2008, from http://www.cs.cmu.edu/~listen/pdfs/UM2003_paper_test_prediction.pdf
- *en:SimpleCLI (3.5.6)*. (June 4, 2007). Retrieved May 6, 2008, from http://weka.sourceforge.net/wekadoc/index.php/en:Simple_CLI_%283.5.6%29

References

- *en:Explorer (3.5.6)*. (June 4, 2007). Retrieved May 6, 2008, from http://weka.sourceforge.net/wekadoc/index.php/en:Explorer_%283.5.6%29
- *en:Experimenter – Standard Experiments (3.5.6)*. (February 25, 2008). Retrieved May 6, 2008, from http://weka.sourceforge.net/wekadoc/index.php/en:Experimenter_-_Standard_Experiments_%283.5.6%29
- *en:KnowledgeFlow (3.5.7)*. (February 21, 2008). Retrieved May 6, 2008, from http://weka.sourceforge.net/wekadoc/index.php/en:Knowledge_Flow_%283.5.7%29