29 Case Studies: Java Machine Learning Software Available on the Web

Chapter Objectives This chapter provides a number of sources for open source and free machine learning software on the web.

Chapter Contents 29.1 Java Machine Learning Software

29.1 Java Machine Learning Software

There are many popular java-based open-source machine learning software packages available on the internet. Several important and widely used ones are described below.

Weka Weka is a Java-based open-source software distributed under the GNU General Public License. It was developed at the University of Waikato in Hamilton, New Zealand in 1993.

Weka is a very popular machine learning software that is widely used for data-mining problems. The main algorithms implemented in Weka focus on pattern classification, regression and clustering. Tools for data preprocessing and data visualization are also provided. These algorithms can either be directly applied to a dataset or be called from other Java code. Weka algorithms can also be used as building blocks for implementing new machine learning techniques.

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

ABLE ABLE is a freely-available Java-toolkit for agent-based machine learning problems developed by the IBM T. J. Watson Research Center in Yorktown Heights, NY.

The ABLE framework provides a library of packages, classes and interfaces for implementing machine learning techniques like neural networks, Bayesian classifiers and decision trees. It also provides a Rule Language for rule-based inference using Boolean and fuzzy logic. The packages and classes can be extended for developing custom algorithms. Support is also provided for reading and writing text and database data, data transformation and scaling and invocation of user-defined external functions.

http://www.alphaworks.ibm.com/tech/able

JOONE JOONE (Java Object-Oriented Neural Engine) is a free java framework for implementing, training and testing machine learning algorithms using artificial neural networks (ANN). The software includes algorithms for feed-forward neural networks, recursive neural networks, time-delay neural networks, standard and resilient back propagation, Kohonen self-

organizing maps, Principal Component Analysis (PCA), and modular neural networks.

JOONE components can be plugged into other software packages and can be extended to design more sophisticated algorithms. It comes with a GUI editor to visually create and test any neural network and a Distributed Test Environment to train many neural networks in parallel and select the best one for a given problem.

http://www.jooneworld.com/

LIBSVM LIBSVM (Library for Support Vector Machines) is an integrated software solution for classification, regression and distribution-estimation using support vector machines (SVM) developed at the National Taiwan University. The source code is freely available with both C++ and Java versions.

The main features of the LIBSVM software are different SVM formulations, efficient multi-class classification, cross-validation for model selection, probability estimates, weighted SVM for unbalanced data and automatic model selection. It also includes a GUI and interfaces for other languages like Python, R, MATLAB, Perl, Ruby and Common Lisp.

http://www.csie.ntu.edu.tw/~cjlin/libsvm/