|  |
| --- |
| **Practical 9 POSTLAB****Use WEKA to implement Association Mining land clustering** |
| **K-means, Agglomerative and Divisive Clustering** |

# **How do you load Weka?**

**Below are the steps to get started with weka**

* **Download weka from Weka Download Page And Install**
* **Start Weka by clicking the weka icon**
* **Click on the ‘Explorer’ button to launch the weka Explorer**
* **Click the ‘Open File’ to load the dataset**
* **Select the Required Algorithm and click ‘Start’ to run the algorithm**
* **Right click on the output and select the visualize option for visualization**

**-----------------------------------x--------------------------------------x------------------------------------------------**

# **What options are available on main panel?**

**In total there are 5 options available on the main menu displayed after launching weka**

* **Explorer – This option provides an environment for exploring data**
* **Experimenter – It is an environment for performing experiments and conducting statistical tests between learning schemes**
* **KnowledgeFlow – It is a Java-Beans based interface for setting up and running machine learning experiments**
* **Workbench – It is an environment that combines all of the GUI interfaces into a single interface**
* **Simple CLI – This option provides a simple command-line interface and allows direct execution of Weka commands**

**-----------------------------------x--------------------------------------x------------------------------------------------**

# **What is the purpose of the the following in Weka?**

**The Explorer This option provides an environment for exploring data. Clicking on this option opens a new window displaying the following actions to be performed**

1. **Pre-process**
2. **Classify**
3. **Cluster**
4. **Associate**
5. **Select attributes**
6. **Visualize**

**The Knowledge  Flow interface**

**The KnowledgeFlow presents a "data-flow" inspired interface to Weka. The user can select Weka components from a tool bar, place them on a layout canvas and connect them together in order to form a "knowledge flow" for processing and analyzing data. At present, all of Weka's classifiers and filters are available in the KnowledgeFlow along with some extra tools.**

**The KnowledgeFlow can handle data either incrementally or in batches. Of course learning from data incrementally requires a classifier that can be updated on an instance by instance basis. Currently in Weka there are five classifiers that can handle data incrementally: NaiveBayesUpdateable, IB1, IBk, LWR.**

**The Experimenter**

**The Experimenter enables you to set up large-scale experiments, start them running, leave them, and come back when they have finished and analyze the performance statistics that have been collected.They automate the experimental process. The statistics can be stored in ARFF format. It allows users to distribute the computing load across multiple machines using Java RMI**

**Command Line Interface**

**One can use the command line interface of Weka either through a command prompt or through the SimpleCLI mode. Weka consists of a hierarchical package system. For example here J48 program is part of the treespackage which further resides in the classifier package. Finally the wekapackage contains the classifiers package**

# **Describe the arff file format**

**ARFF stands for Attribute-Relation File Format. It is an ASCII text file that describes a list of instances sharing a set of attributes. ARFF files have two distinct sections. The first section is the Header information, which is followed by the Data information.**

**Header looks like the following**

**% 1. Title: Iris Plants Database**

**%**

**% 2. Sources:**

**% (a) Creator: R.A. Fisher**

**% (b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)**

**% (c) Date: July, 1988**

**%**

**@RELATION iris**

**@ATTRIBUTE sepallength NUMERIC**

**@ATTRIBUTE sepalwidth NUMERIC**

**@ATTRIBUTE petallength NUMERIC**

**@ATTRIBUTE petalwidth NUMERIC**

**@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}**

**Data looks like the following**

**@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**

**5.4,3.9,1.7,0.4,Iris-setosa**

**4.6,3.4,1.4,0.3,Iris-setosa**

**5.0,3.4,1.5,0.2,Iris-setosa**

**4.4,2.9,1.4,0.2,Iris-setosa**

**4.9,3.1,1.5,0.1,Iris-setosa**

**Lines that begin with a % are comments. The @RELATION, @ATTRIBUTE and @DATA declarations are case insensitive.**

# **Press the Explorer button on the main panel and load the weather dataset and answer the following questions**

1. **How many instances are there in the dataset?**

**There are a total of 14 instances in the dataset**

1. **State the names of the attributes along with their types and values.**

**Outlook – Type: Nominal, values: sunny, overcast, rainy**

**Temperature – Type: Numeric**

**Humidity – Type: Numeric**

**Windy – Type: Nominal, values: True, False**

**Play – Type: Nominal, values: Yes, No**

1. **In the histogram on the bottom-right, which attributes are plotted on the X,Y-axes? How do you change the attributes plotted on the X,Y-axes?**

**The histogram is based on the attribute selected in the attributes list. For example, the default selected attribute here is ‘outlook’. You can change this by selecting other attribute from the attribute list.**

1. **What happens with the Visualize All button is pressed?**

**When the Visualize All button is pressed the graph of all the attributes with respect to the class attribute is displayed.**

1. **How will you view the instances in the dataset? How will you save the changes?**

**There is a ‘save’ button made available for saving purpose on the window.**

# **What is the purpose of the following in the Explorer Panel?**

**Preprocess Panel**

**The preprocess panel is the start point for knowledge exploration. From this panel you can load datasets, browse the characteristics of attributes and apply any combination of Weka's unsupervised filters to the data.**

1. **What are the main sections of the Preprocess panel?**
2. **Open file 5. Edit**
3. **Open URL 6. Save**
4. **Open DB 7. Filter**
5. **Generate**
6. **What are the primary sources of data in Weka?**

**Classify Panel**

**The classifier panel allows you to configure and execute any of the weka classifiers on the current dataset. You can choose to perform a cross validation or test on a separate dataset. Classification errors can be visualized in a pop-up data visualization tool. If the classifier produces a decision tree it can be displayed graphically in a pop-up tree visualizer.**

### **Cluster Panel**

**From the cluster panel you can configure and execute any of the wekaclusterers on the current dataset. Clusters can be visualized in a pop-up data visualization tool.**

### **Associate Panel**

**From the associate panel you can mine the current dataset for association rules using the weka associators.**

### **Select Attributes Panel**

### **This panel allows you to configure and apply any combination of weka attribute evaluator and search method to select the most pertinent attributes in the dataset. If an attribute selection scheme transforms the data then the transformed data can be visualized in a pop-up data visualization tool.**

### **Visualize Panel**

### **This panel displays a scatter plot matrix for the current dataset. The size of the individual cells and the size of the points they display can be adjusted using the slider controls at the bottom of the panel. The number of cells in the matrix can be changed by pressing the "Select Attributes" button and then choosing those attributes to displayed.This panel allows you to visualize the current dataset in one and two dimensions.This panel can also be popped up in a separate window from the classifier panel and the cluster panel to allow you to visualize predictions made by classifiers/clusterers.**