Experiment - 4

Aim: Demonstrate performing classification on data sets

→ Load each dataset into Weka and run Id3, J48 classification algorithm. Study the classifier output. Compute entropy values, kappa statistic.

→ Extract if-then rules from the decision tree generated by the classifier, observe the confusion matrix.

→ Load each dataset into Weka and perform Naive-bayes and K-NN classifiers for each dataset, and classification and k-Nearest Neighbour classification. Interpret the results obtained.

→ Plot Roc Curves.

→ Compare classification results of ID3, J48, Naive-Bayes and k-NN classifiers for each dataset and deduce which classifier is performing best and poor for each dataset and justify.

Objectives:

The ultimate objective of classification is to relate a variable of interest with observed variables. The actual variable of interest is meant to be of "Qualitative" type. The algorithm required for performing the classification is known as the classifier.

Zero R :-

→ ZeroR is the simplest classification method which relies on the largest and ignores all predictors

→ Zero R classifier simply predicts the majority category.

→ Although there is no predictability power in ZeroR it is useful for determining a baseline performance as a benchmark for other classification methods.

oneR :

→ This method is used in the Sequential learning Algorithm for learning the rules.

→ It returns a single rule that covers at least some examples.

→ However, what makes it really powerful is its ability to create relations among the attributes given. Hence covering a larger hypothesis space.

Explorer :

It is an environment for exploring data.

Simple CLI :

It provides a simple Command-line Interface and allows direct execution of Weka Commands.

Experimenter:

It is an environment for performing experiment and conducting statistical tests between learning schemes.

Knowledge Flow:

It is a Java-Beans based interface to setting up and running machine learning experiments.

Preprocess:

It is the first step in machine learning to preprocess the data. It is used to select the data file preprocessing and make it fit for applying the various machine learning Algorithms.

Classify:

The classify tab provides you several machine learning algorithms for the classification of your data. Such as linear regression, Logistic Regression.

Test options:

Before you run the classification algorithm, you need to set test options. Set test options in the test options box. The test options that available Now are:-

1) use training set: evaluates the classifier on how well it predicts the class of the instances it was trained on.

2) Supplied test set: Evaluates the classifier on how well it predicts the class of a set of instances loaded from a file. clicking on the "set..." button brings up a dialog allowing you to choose the file to test on.

3) Cross Validation:
evaluates the classifier by cross-validation, using the number of folds that are entered in the 'Folds' text field.

4) percentage split:
evaluates the classifier on how well it predicts a certain percentage of the data, which is held out for testing the amount of data held out depends on the value entered in the '%' field.

Steps Required:

1. open Weka you can see 5 tabs on the right side of the application. These are explorer, experimentor, knowledge flow, Work bench, Simple CLI.
2. click on 'explorer'.
3. you can see classify tab click on the classify button.

4. you can observe choose test options etc.
5. In test option you can see cross-validation folds. Set it as 10.
6. Right click on choose option, Then Select the ZeroR algorithm or one R algorithm.
7. Click Start button.
8. ZeroR algorithm or one R algorithm will execute and it gives the Output.

Output:

ZeroR