

Please go to the website : www.cs.waikato.ac.nz/ml/weka/ and download the bookversion of Weka. You might want to get the latest cersion, but the possible option might be different from the ones used here in this exercise sheet. Start Weka as described on the webpage.

Klick on the Explorer Button to start the Explorer.

Data Pre-Processing

- Open the *labor* dataset (labor.arff)
- Observe attribute 4
- Choose the "Replace Missing Values" filter from Filters→Attribute→Unsupervised
- Apply it and consider attribute k again
- Undo your changes, choose Filters→Unsupervised→Attribute→Normalize
- Apply it, have a look at the attributes
- Load the iris dataset
- Apply Filters→Unsupervised→Attribute→Discretize with Number of Bins = 3, have a look at the attributes
- Undo
- Apply Filters→Supervised→Attribute→Discretize, have a look at the attributes

Comparing Evaluation Methods

- Open the ... dataset
- Choose the J48 classifier
- Choose evaluation method "On Training Data" and run the classifier
- Observe the results
- Repeat for evaluation method "Percentage Split: 66%"
- Repeat for evaluation method "Cross-Validation Folds: 10"
- Choose the Random Tree classifier with evaluation method "On Training Data"

Comparing Classifier Settings

- Open the *balance – scale* dataset
- Choose the J48 classifier
- Run it with standard settings on the dataset, evaluation method 10-fold cross-validation and observe the results
- Repeat with pruning turned off
- Repeat with minimum number of instances per leave set to 1 (minNumObj=1)
- Repeat with reduced error pruning instead of standard method

Comparing Classifiers

- Open the *Voting Record* dataset (voting.arff)
- In the classification tab, choose the perceptron network classifier from Functions
- Let it run and observe the learned model
- Repeat the process for the Naive Bayes classifier
- Repeat the process for the J48 classifier
- Repeat the process for the PART classifier