
Introduction to Weka

Neo Qiang Hao
Computer Science &
Learning, Design and Technology

Data Mining & Machine Learning

1. Data Collection
2. Data Cleaning
3. Data Processing ----- Weka
4. Data Analysis ----- Weka
5. Sharing Data and Results ----- Weka

Weka

Free

Friendly GUI

Rich collection of supervised & unsupervised methods

Memory hungry



Weka



Weka

Data: Attribute-Relation File Format (arff)

.csv or .txt Files can be easily converted to .arff. files

R:

```
write.arff(x, file, eol = "\n", relation =  
deparse(substitute(x)))
```

Python:

```
..... ..
```



Weka

Data: Attribute-Relation File Format (arff)

```
weather.nominal.arff x
1  @relation weather.symbolic
2
3  @attribute outlook {sunny, overcast, rainy}
4  @attribute temperature {hot, mild, cool}
5  @attribute humidity {high, normal}
6  @attribute windy {TRUE, FALSE}
7  @attribute play {yes, no}
8
9  @data
10 sunny,hot,high,FALSE,no
11 sunny,hot,high,TRUE,no
12 overcast,hot,high,FALSE,yes
13 rainy,mild,high,FALSE,yes
14 rainy,cool,normal,FALSE,yes
15 rainy,cool,normal,TRUE,no
16 overcast,cool,normal,TRUE,yes
17 sunny,mild,high,FALSE,no
18 sunny,cool,normal,FALSE,yes
19 rainy,mild,normal,FALSE,yes
20 sunny,mild,normal,TRUE,yes
21 overcast,mild,high,TRUE,yes
22 overcast,hot,normal,FALSE,yes
23 rainy,mild,high,TRUE,no
24
```



Weka

1. Transformation
2. Attribute Selection
3. Supervised Learning Methods
4. Unsupervised Learning Methods



Weka

1. Transformation

Example 1: Transformation of documents

```
>>> StringTOWordVector
```

Example 2: Discretization

```
>>> Discretize
```



Weka

1. Transformation
2. Attribute Selection

Reduce noise. Improve accuracy.



Weka

1. Transformation

2. Attribute Selection

Example 1: Ranker Method

```
>>> GainRatioAttributeEval
```

Example 2: Wrapper Method

```
>>> CfsSubsetEval
```



Weka

1. Transformation
2. Attribute Selection
3. Supervised Learning Methods

Bayes
Logistic
Multilayer
Perceptron

Support Vector
Machine
K nearest neighbors
Boosting

