

# Decision trees with Weka

Lab 1.1

# Using Command Line Interface

# CLI. ID3 decision tree

Build decision tree with Id3 algorithm (no numeric values, splits are based on information gain):

```
java weka.classifiers.trees.Id3 -t data/weather.nominal.arff
```

Build decision tree with J48 algorithm (allows numeric values, uses gain ratio, pruning):

```
java weka.classifiers.trees.J48 -t data/weather.arff
```

Save model to disk:

```
java weka.classifiers.trees.Id3 -t data/weather.nominal.arff -d  
data/Id3.model
```

# CLI: J48 algorithm

Run decision tree induction algorithm J48 (allows numeric values, uses gain ratio, pruning):

```
java weka.classifiers.trees.J48 -t data/weather.arff
```

Performs 10-fold cross validation  
Estimates error rate on this 10 sets

**To be fully explained  
in Lecture 7**



# Classify a new record

- From the original arff file create a new file test1.arff where in data section put the records to classify, and put ? for the class value:

```
@relation weather.symbolic

@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,hot,high,FALSE,?
sunny,hot,high,TRUE,?
```

# CLI: classify records

```
java weka.classifiers.trees.Id3 -T data/test1.arff -l  
models/Id3.model -p 0
```

Using GUI

# Weka explorer

- Load file weather.arff
- Explore attributes

The screenshot shows the Weka Explorer application window. The 'Open file...' button is circled in red. The interface displays the 'weather.symbolic' relation with 14 instances and 5 attributes. The 'outlook' attribute is selected, showing a table with 3 distinct values: sunny (5), overcast (4), and rainy (5). A stacked bar chart visualizes the distribution of the 'play' class (Nom) for each outlook value.

No.	Label	Count
1	sunny	5
2	overcast	4
3	rainy	5

Class: play (Nom)

Visualize All

5 4 5

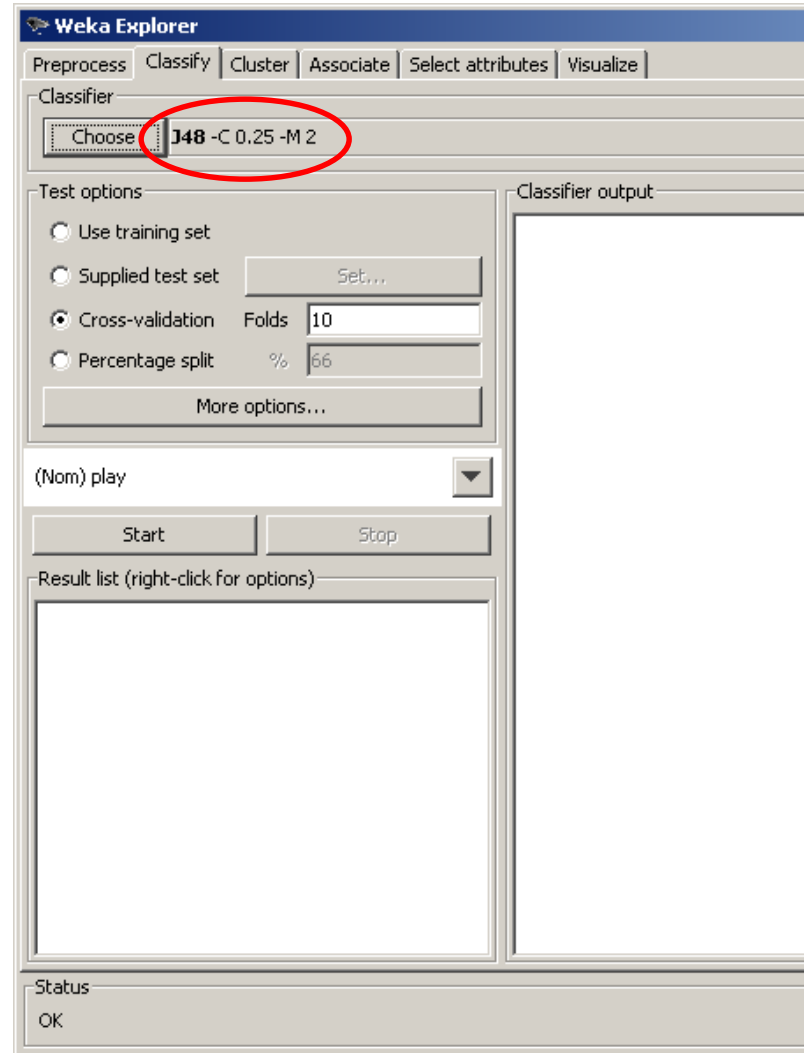
Remove

Status: OK Log x 0



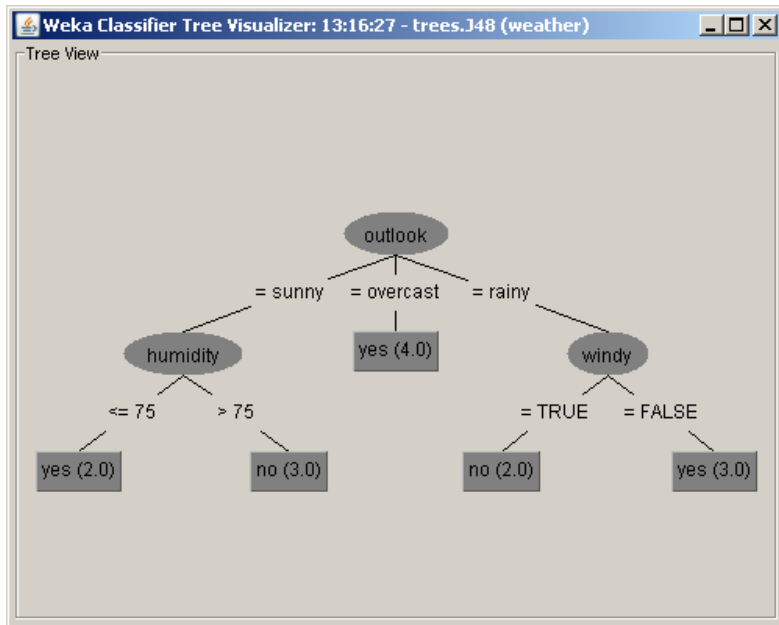
# Weka explorer

- Run decision tree classifiers:
  - Id3 (simple, based on information gain): no numeric attributes – weather.nominal
  - J48 (C4.5 – gain ratio, missing values, tree pruning)
  - SimpleCart (numeric attributes)



# Weka explorer

- Compare error rates
- Right-click on model line – visualize tree
- Can save the model and load it



select attributes | Visualize

Classifier output

```
outlook = sunny
| humidity <= 75: yes (2.0)
| humidity > 75: no (3.0)
outlook = overcast: yes (4.0)
outlook = rainy
| windy = TRUE: no (2.0)
| windy = FALSE: yes (3.0)
```

Number of Leaves : 5

Size of the tree : 8

Time taken to build model: 0 seconds

=== Stratified cross-validation ===  
=== Summary ===

Correctly Classified Instances	9	64.2857 %
Incorrectly Classified Instances	5	35.7143 %
Kappa statistic	0.186	
Mean absolute error	0.2857	

Log