Agenda

• WEKA: A Machine Learning Toolkit

The Explorer

- **o** Classification and Regression
- Clustering
- Association Rules
- Attribute Selection
- **o** Data Visualization
- The Experimenter
- The Knowledge Flow GUI
- Conclusions

WEKA

- Machine learning/data mining software written in Java (distributed under the GNU Public License)
- Used for research, education, and applications

Main features:

- Comprehensive set of data pre-processing tools, learning algorithms and evaluation methods
- o Graphical user interfaces (incl. data visualization)
- Environment for comparing learning algorithms





Explorer: pre-processing the data

- Data can be imported from a file in various formats: ARFF, CSV, C4.5, binary
- Data can also be read from a URL or from an SQL database (using JDBC)
- Pre-processing tools in WEKA are called "filters"
- WEKA contains filters for:
 - Discretization, normalization, resampling, attribute selection, transforming and combining attributes, ...





WEKA with "flat" files

@relation heart-disease-simplified

@attribute age numeric
@attribute sex { female, male}
@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}
@attribute cholesterol numeric
@attribute exercise_induced_angina { no, yes}
@attribute class { present, not_present}

@data

63,male,typ_angina,233,no,not_present 67,male,asympt,286,yes,present 67,male,asympt,229,yes,present 38,female,non_anginal,?,no,not_present

Flat file in ARFF format





WEKA only deals with "flat" files

numeric attribute

-nominal attribute

@relation heart-disease-simplified

@attribute age numeric

@attribute sex { female, male}

@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}

@attribute cholesterol numeric

@attribute exercise_induced_angina { no, yes}

@attribute class { present, not_present}

@data 63,male,typ_angina,233,no,not_present 67,male,asympt,286,yes,present 67,male,asympt,229,yes,present 38,female,non_anginal,?,no,not_present







WEKA with "flat" files (Sparse format)

@relation example
@attribute att1 numeric
@attribute att2 numeric
@attribute att3 numeric
@attribute att4 numeric
@attribute class {"class A", "class B"}

@data 0, X, 0, Y, "class A" 0, 0, W, 0, "class B"

-Considering only the non 0 values

-Represent each values with: POSITION VALUE information

-Each couple (POSITION - VALUE) is separated with a comma

-Useful for documents representation in ARFF format

@relation example
@attribute att1 numeric
@attribute att2 numeric
@attribute att3 numeric
@attribute att4 numeric
@attribute class {"class A", "class B"}

@data {1 X, 3 Y, 4 "class A"} {2 W, 4 "class B"}







Explorer: Pre-processing

Data can be imported from a file in various formats:
 O ARFF, CSV, C4.5, binary

- Data can also be read from a URL or from an SQL database (using JDBC)
- Pre-processing tools in WEKA are called "filters"
- WEKA contains filters for:
 - Discretization, normalization, resampling, attribute selection, transforming and combining attributes, ...





| 000 | Weka Knowledge Explorer | | | | |
|---|--------------------------------|--|----------------------------|--|--|
| · | Preprocess Classify Cluster As | ssociate Select attributes Visualize | | | |
| Open file | Open URL Oper | n DB Undo | Save | | |
| Filter | | | | | |
| Choose None | | | Apply | | |
| Current relation Relation: None Instances: None | Attributes: None | Selected attribute Name: None Missing: None Distinct: None | Type: None Unique: None | | |
| Attributes | | | | | |
| Status | | | Visualize All | | |
| Status Welcome to the Wek | a Knowledge Explorer | \subset | Log 💉 x 0 | | |

| 000 | Weka Knowledge Explorer | | | | |
|---|--------------------------------|--|----------------------------|--|--|
| (| Preprocess Classify Cluster As | ssociate Select attributes Visualiz | ze | | |
| Open file | Open URL Oper | n DB Undo | Save | | |
| Choose None | | | Apply | | |
| Current relation Relation: None Instances: None | Attributes: None | Selected attribute Name: None Missing: None Distinct: None | Type: None Unique: None | | |
| Attributes | | | | | |
| | | | | | |
| | | | | | |
| | | | Visualize All | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| Status | | | | | |
| Welcome to the Wek | a Knowledge Explorer | | | | |

| 😑 🖯 🖯 Weka Knowle | Weka Knowledge Explorer | | | | |
|--|---|---------------------------------|--|--|--|
| Preprocess Classify Cluster As | sociate Select attributes Visualiz | e | | | |
| Open file Open URL Open | DB Undo (| Save | | | |
| Choose None | | Apply | | | |
| Current relation | Selected attribute | | | | |
| Relation: iris Instances: 150 Attributes: 5 | Name: sepallength Missing: 0 (0%) Distinct: 35 | Type: Numeric Unique: 9 (6%) | | | |
| Attributes | Statistic | Value | | | |
| No. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class | Maximum 7.9 Mean 5.843 StdDev 0.828 | | | | |
| Status OK | Colour: class (Nom) | Visualize All | | | |



| 😑 🖯 Weka Knowledge Explorer | | | | |
|---|--|---------------------------------|--|--|
| Preprocess Classify Cluster | Associate Select attributes Visu | alize | | |
| Open file Open URL | Open DB Undo | Save | | |
| Filter | | | | |
| Choose None | | Apply | | |
| Current relation | Selected attribute | | | |
| Relation: iris Instances: 150 Attributes: 5 | Name: class Missing: 0 (0%) Distinct: 3 | Type: Nominal Unique: 0 (0%) | | |
| Attributes | Label | Count | | |
| No. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class | Iris-versicolor 50 Iris-virginica 50 | | | |
| | Colour: class (Nom) | Visualize All | | |
| | 50 50 | 50 | | |
| Status | | | | |
| ОК | | | | |

| 🖯 🖯 Weka Knowledge Explorer | | | | |
|---|--|--|--|--|
| Preprocess Classify | Cluster Associate Select attributes Visualize | | | |
| Open file Open URL | Open DB Undo Save | | | |
| Filter | | | | |
| Choose None | Apply | | | |
| Current relation | Selected attribute | | | |
| Relation: iris Instances: 150 Attributes: 5 | Name: class Type: Nominal Missing: 0 (0%) Distinct: 3 Unique: 0 (0%) | | | |
| Attributes | Label Count | | | |
| No. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class | Iris-versicolor 50 Iris-virginica 50 Colour: class (Nom) Visualize All | | | |
| | | | | |
| Status | | | | |
| OK | | | | |



| 000 | Weka Knowl | edge Explorer | |
|--|-------------------------------|---|---|
| | Preprocess Classify Cluster A | ssociate Select attributes | Visualize |
| Open file | Open URL Ope | n DB Undo | Save |
| Choose None | | | Apply |
| Current relation Relation: iris Instances: 150 | Attributes: 5 | Selected attribute Name: petallength Missing: 0 (0%) Dist | Type: Numeric inct: 43 Unique: 10 (7%) |
| Attributes | | Statistic | Value |
| No. 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class | Name | Maximum Mean StdDev | 6.9 3.759 1.764 |
| | | Colour: class (Nom) | Visualize All |
| Status OK | | | Log 💉 X O |













| 000 | | Weka Knowledge | Explorer | |
|-------------------|---------------------------------------|-------------------------------------|-------------------------------------|------------------------|
| | Preprocess Classify | Cluster Associa | ate Select attributes Visualize | |
| 0 | pen file Open URL | Open DB | . Undo | Save |
| Filter | | | | |
| Choos | e Discretize -B 10 -R first-last | 😑 🖯 🖯 🛛 wek | a.gui.GenericObjectEditor | Apply |
| Current re | elation | weka.filters.unsuper | vised.attribute.Discretize | |
| Relati Instanc | on: iris :es: 150 Attributes: | About An instance filter that of | discretizes a range of numeric More | : Numeric : 10 (7%) |
| Attributes | | aunoutes in the datase | | e |
| No. | Name 1 sepallength 2 sepalwidth | attributeIndices | first-last | |
| | 3 petallength | bins | 10 | |
| 1 | 4 petalwidth 5 class | findNumBins | False 🛟 |) |
| | | invertSelection | False 🛟 | Visualize All |
| | | makeBinary | False 🛟 |) |
| | | useEqualFrequency | False 🛟 |) |
| | | Open | Save OK Cancel |) |
| | | 11 | 2 0 0 3 4 3.95 | 10 2 4 6.9 |
| Status | | | <u> </u> | |
| OK | | | Log | 🗩 🛷 x u |

| 000 |) | Weka Knowledge I | Explorer | |
|------------|--------------------------------|---|--|----------------|
| | Preprocess Classify | Cluster Associa | te Select attributes Visualize | |
| 0 | pen file Open URL | Open DB | Undo | Save |
| Filter | Discretize -R 10 -R first-last | \varTheta 🖯 🔵 🛛 weka | a.gui.GenericObjectEditor | Annly |
| Current re | elation | weka.filters.unsuper | vised.attribute.Discretize | : Numeric |
| Instand | ces: 150 Attributes: ! | An instance filter that of attributes in the datase | liscretizes a range of numeric More More | : 10 (7%) e |
| No. | Name 1 sepallength | attributeIndices | first-last |] |
| | 2 sepaiwidth 3 petallength | bins | 10 | |
| | 4 petalwidth 5 class | findNumBins | False |) |
| | | invertSelection | False | Visualize All |
| | | makeBinary | False |) |
| | | useEqualFrequency | False ‡ |) |
| | | open | Save OK Cancel |) |
| | | | 2 0 0 3 4 | 10 2 4 |
| Status | | | 3.95 | 6.9 |
| ОК | | | Log | 💭 🛷 × 0 |

| 000 | | Weka Knowledge | Explorer | |
|-------------------|---------------------------------------|-------------------------------------|-------------------------------------|------------------------|
| | Preprocess Classify | Cluster Associa | ate Select attributes Visualize | |
| 0 | pen file Open URL | Open DB | . Undo | Save |
| Filter | | | | 3 |
| Choos | e Discretize -B 10 -R first-last | 🖯 🖯 🖯 wek | a.gui.GenericObjectEditor | Apply |
| Current re | elation | weka.filters.unsuper | vised.attribute.Discretize | |
| Relati Instanc | on: iris :es: 150 Attributes: | About An instance filter that of | discretizes a range of numeric More | : Numeric : 10 (7%) |
| Attributes | | aunoutes in the datase | | e |
| No. | Name 1 sepallength 2 sepalwidth | attributeIndices | first-last | |
| | 3 petallength | bins | 10 | |
| | 4 petalwidth 5 class | findNumBins | False 🛟 |) |
| | | invertSelection | False 🗧 | Visualize All |
| | | makeBinary | False 🛟 |) |
| | | useEqualFrequency | True 🗘 |) |
| | | Open | Save OK Cancel |) |
| | | 11 | 2 0 0 3 4 3.95 | 10 2 4 6.9 |
| Status | | | | |
| ОК | | | Log |) 📣 × 0 |

| 000 | Weka Knowledge | Explorer | |
|--|---|---------------------------------|-----------------------------|
| Preprocess Classify | Cluster Associa | ate Select attributes Visualize | |
| Open file Open URL | Open DB | Undo | Save |
| Filter | | | |
| Choose Discretize -B 10 -R first-last | \varTheta 🖯 🖯 🛛 wek | a.gui.GenericObjectEditor | Apply |
| Current relation Relation: iris Instances: 150 Attributes: Attributes | About An instance filter that of attributes in the datase | vised.attribute.Discretize | : Numeric : 10 (7%) e |
| No. Name 1 sepallength | attributeIndices | first-last | 1 |
| 3 petallength | bins | 10 | |
| 4 petalwidth 5 class | findNumBins | False | |
| | invertSelection | False | Visualize All |
| | makeBinary | False | |
| | useEqualFrequency | True | |
| | Open | Save OK Cancel |) |
| | | 2 0 0 3 4 3.95 | 10 2 4 6.9 |
| Status OK | | Log |) 💉 🔊 |





| \varTheta 🖯 🖯 Weka Knowle | Weka Knowledge Explorer | | | | |
|---|--|---------------------------------|--|--|--|
| Preprocess Classify Cluster As | ssociate Select attributes Visualize | | | | |
| Open file Open URL Open | n DB Undo | Save | | | |
| Filter | | | | | |
| Choose Discretize -F -B 10 -R first-last | | Apply | | | |
| Current relation | Selected attribute | | | | |
| Relation: iris-weka.filters.unsupervised.attribute.Disc Instances: 150 Attributes: 5 | Name: petallength Missing: 0 (0%) Distinct: 10 | Type: Nominal Unique: 0 (0%) | | | |
| Attributes No. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class 5 class | Label '(-inf-1.45]' 23 '(1.45-1.55]' 14 '(1.55-1.8]' 11 '(1.8-3.95]' 13 '(3.95-4.35]' 14 '(4.35-4.65]' 15 '(4.65-5.05]' 18 Colour: class (Nom) | Count | | | |
| Status OK | | Log 💉 X O | | | |

Explorer: Building "Classifiers"

 Classifiers in WEKA are models for predicting nominal or numeric quantities

- Implemented learning schemes include:
 - Decision trees and lists, instance-based classifiers, support vector machines, multi-layer perceptrons, logistic regression, Bayes' nets, ...
- "Meta"-classifiers include:
 - Bagging, boosting, stacking, error-correcting output codes, locally weighted learning, ...





| 000 | 🖯 🖯 🧶 Weka Knowledge Explorer | | | | | | |
|------------------------------|-------------------------------|-----------|----------------|-----------|-------------------|-----------|--|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | |
| Classifier | | | | | | | |
| Choose ZeroR | | | | | | | |
| Test options | | Clas | ssifier output | | | | |
| O Use training set | | | | | | | |
| Supplied test set | set | | | | | | |
| Cross-validatior | Folds 10 | | | | | | |
| O Percentage split | % 66 | | | | | | |
| More op | otions | \supset | | | | | |
| (Nom) class | | • | | | | | |
| Start | Stop | | | | | | |
| Result list (right-click for | options) — | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| Status | | | | | | | |

| 0 | ĸ | |
|---|----|--|
| U | n, | |



| 🖯 🖯 😌 Weka Knowledge Explorer | | | | | | | |
|-------------------------------|------------|----------|----------------|-----------|-------------------|-----------|---|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | |
| Classifier | | | | | | |] |
| Choose ZeroR | | | | | | | |
| Test options | | Cla | ssifier output | | | | |
| Use training set | | | | | | | |
| O Supplied test set | t Set | | | | | | |
| Cross-validation | n Folds 10 | | | | | | |
| O Percentage split | % 66 | | | | | | |
| More op | otions | | | | | | |
| | | | | | | | |
| (Nom) class | | + | | | | | |
| Start | Stop | | | | | | |
| Result list (right-click for | options) | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| 1 | | | | | | | |
| Status | | | | | | | |

ОК



| $\Theta \Theta \Theta$ | Weka Knowledge Explorer | | | | |
|------------------------|--|---------|--|--|--|
| Preprocess Classify | Cluster Associate Select attributes Visual | ize | | | |
| Classifier | | | | | |
| 📁 weka | | | | | |
| 📃 🔻 📁 classifiers | | | | | |
| 📔 🕨 📁 bayes | ifier output | | | | |
| functions | | | | | |
| ► 📁 lazy | | | | | |
| 🕨 📁 meta | | | | | |
| ▶ 📁 misc | | | | | |
| V v trees | | | | | |
| ▶ j adtree | | | | | |
| DecisionStump | | | | | |
| | | | | | |
| ¥ 📁 j48 | | | | | |
| / htt | | | | | |
| m5 | | | | | |
| RandomForest | | | | | |
| RandomTree | | | | | |
| REPTree | | | | | |
| UserClassifier | | | | | |
| rules | | | | | |
| | | | | | |
| | | | | | |
| OK | | Log x 0 | | | |

| 00 | Weka Knowledge Explorer | | | | | | | |
|------------------------------|-------------------------|----------|----------------|-----------|-------------------|-----------|-----|-------|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | | |
| Classifier | | | | | | | | |
| Choose J48 -C | 0.25 -M 2 | | | | | | | |
| Test options | | Cla | ssifier output | | | | | |
| O Use training set | | | | | | | | 11111 |
| Supplied test set | t Set | | | | | | | |
| Cross-validation | n Folds 10 | | | | | | | 11111 |
| Percentage split | % 66 | | | | | | | |
| More op | otions | | | | | | | |
| | | | | | | | | |
| (Nom) class | | • | | | | | | |
| Start | Stop | | | | | | | |
| Result list (right-click for | options) | | | | | | | 11111 |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| Status | | | | | | 6 | 100 | ≥ x 0 |
| UK | | | | | | 0 | | r ~ • |

| $\Theta \Theta \Theta$ | Weka Knowledge Explorer | | | | | | |
|------------------------------|-------------------------|----------|----------------|-----------|-------------------|-----------|-----------|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | |
| Classifier | | | | | | | |
| Choose J48 -C | 0.25 -M 2 | | | | | | |
| Test options | | Clas | ssifier output | | | | |
| Use training set | | | | | | | |
| Supplied test set | t Set | | | | | | |
| Cross-validation | n Folds 10 | | | | | | |
| O Percentage split | % 66 | | | | | | |
| More op | otions | | | | | | |
| (Nom) class | | • | | | | | |
| Start | Stop | | | | | | |
| Result list (right-click for | options) | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| Status | | 111 | | | | | |
| OK | | | | | | Lo | g 💉 🗙 x 0 |

| 000 | Weka Knowledge Explorer | | | | | |
|------------------------------|-------------------------|------------------------|--------------|-------------------|-----------|--|
| | Preprocess | Classify Cluster | Associate | Select attributes | Visualize | |
| Classifier | | | | | | |
| Choose J48 -C (| 0.25 -M 2 | 😑 🖯 🖯 weka.gui. | .GenericObje | ctEditor | | |
| Test options | | weka.classifiers.trees | s.j48.J48 | | | |
| O Use training set | | binarySplits | False | • | | |
| Supplied test set | O Supplied test set Set | | 0.25 | | | |
| Cross-validation | Folds 10 | minNumObi | 0.25 | | | |
| O Percentage split % 66 | | numEolds | 2 | | | |
| More op | cions | | 5 | | | |
| (Nom) class | | reducedErrorPruning | False | • | | |
| | | | False | • | | |
| Start Stop | | subtreeRaising | True | • | | |
| Result list (light-click lot | opuons) | unpruned | False | ; | | |
| | | useLaplace | False | ÷ | | |
| | | Open Save. |) Ок | Cancel | | |
| | | | | | | |
| Status | | | | | C | |
| OK | | | | | 0 | |
| 000 | Weka Knowledge Explorer | | | | | | | | | |
|---------------------------------|-------------------------|-------------------------|----------------------------------|--------|--|--|--|--|--|--|
| F | Preprocess | Classify Cluster | Associate Select attributes Visu | Jalize | | | | | | |
| Classifier | | | | | | | | | | |
| Choose J48 -C 0.2 | 25 –M 2 | 🖯 🖯 🖯 weka.gui.(| GenericObjectEditor | | | | | | | |
| Test options | | weka.classifiers.trees. | .j48.J48 | | | | | | | |
| O Use training set | | binarySplits | False | | | | | | | |
| Supplied test set | Set | confidenceEactor | 0.25 | | | | | | | |
| Cross-validation | Folds 10 | connuenceractor | 0.25 | | | | | | | |
| Percentage split | % 66 | minNumObj | 2 | | | | | | | |
| More optio | ons | numFolds | 3 | | | | | | | |
| (Nom) class | | reducedErrorPruning | False | | | | | | | |
| | - | saveInstanceData | False 🗧 | | | | | | | |
| Result list (right-click for or | Stop | subtreeRaising | True | | | | | | | |
| | | unpruned | False 🛟 | | | | | | | |
| | | useLaplace | False | | | | | | | |
| | | Open Save | OK Cancel | | | | | | | |
| | | | | | | | | | | |
| Status | | | | | | | | | | |
| ОК | | | | | | | | | | |

| 00 | Weka Knowledge Explorer | | | | | | | |
|------------------------------|-------------------------|----------|----------------|-----------|-------------------|-----------|-----|-------|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | | |
| Classifier | | | | | | | | |
| Choose J48 -C | 0.25 -M 2 | | | | | | | |
| Test options | | Cla | ssifier output | | | | | |
| O Use training set | | | | | | | | 10000 |
| Supplied test set | t Set | | | | | | | |
| Cross-validation | n Folds 10 | | | | | | | 20000 |
| Percentage split | % 66 | | | | | | | |
| More op | otions | | | | | | | |
| | | | | | | | | |
| (Nom) class | | • | | | | | | 2000 |
| Start | Stop | | | | | | | |
| Result list (right-click for | options) | | | | | | | 1000 |
| | | | | | | | | 2000 |
| | | | | | | | | 22222 |
| | | | | | | | | |
| | | | | | | | | 2000 |
| | | | | | | | | 11111 |
| | | | | | | | | 20000 |
| | | | | | | | | |
| Status | | | | | | C | Log | x 0 |
| UK | | | | | | | | |

| 00 | Weka Knowledge Explorer | | | | | | | |
|------------------------------|-------------------------|----------|----------------|-----------|-------------------|-----------|-----|--|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | | |
| Classifier | | | | | | | | |
| Choose J48 -C | 0.25 -M 2 | | | | | | | |
| Test options | | Cla | ssifier output | | | | | |
| O Use training set | | | | | | | | |
| Supplied test set | t Set | | | | | | | |
| Cross-validation | n Folds 10 | | | | | | | |
| Percentage split | % 66 | | | | | | | |
| More op | otions | | | | | | | |
| | | | | | | | | |
| (Nom) class | | + | | | | | | |
| Start | Stop | | | | | | | |
| Result list (right-click for | options) | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| Status | | | | | | 6 | | |
| UK | | | | | | 0 | Log | |

| 000 | Weka Knowledge Explorer | | | | | | | |
|------------------------------|-------------------------|----------|----------------|-----------|-------------------|-----------|---------|--|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | | |
| Classifier | | | | | | | | |
| Choose J48 -C | 0.25 -M 2 | | | | | | | |
| Test options | | Cla | ssifier output | | | | | |
| Use training set | | | | | | | | |
| Supplied test set | set | | | | | | | |
| O Cross-validation | Folds 10 | | | | | | | |
| Percentage split | % 66 | | | | | | | |
| More op | tions | | | | | | | |
| | | | | | | | | |
| (Nom) class | | • | | | | | | |
| Start | Stop | | | | | | | |
| Result list (right-click for | options) | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| Chabus | | 1 | | | | | | |
| OK | | | | | | C | Log x 0 | |

| 000 | Weka Knowledge Explorer | | | | | | | | |
|------------------------------|-------------------------|----------|----------------|-----------|-------------------|-----------|-----|-------|--|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | | | |
| Classifier | | | | | | | | | |
| Choose J48 -C | 0.25 -M 2 | | | | | | | | |
| Test options | | Cla | ssifier output | | | | | | |
| Use training set | | | | | | | | 11111 | |
| O Supplied test set | t Set | | | | | | | | |
| O Cross-validation | n Folds 10 | | | | | | | | |
| Percentage split | % 66 | | | | | | | | |
| More op | ptions | | | | | | | 11111 | |
| | | | | | | | | 10000 | |
| (Nom) class | | | | | | | | | |
| Start | Stop | | | | | | | 10000 | |
| Result list (right-click for | r options) | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | 11111 | |
| | | | | | | | | | |
| Status | | J L | | | | | | | |
| ОК | | | | | | C | Log | x 0 | |

| $\Theta \Theta \Theta$ | 🖯 🖯 Weka Knowledge Explorer | | | | | | | | | |
|-------------------------------------|-----------------------------|----------|----------------|---------------|--------------------|-----------|--|--|--|--|
| Prep | rocess | Classify | Cluster | Associate | Select attributes | Visualize | | | | |
| Classifier | | | | | | | | | | |
| Choose J48 -C 0.25 -M 2 | | | | | | | | | | |
| Test options | | Cla | ssifier output | l | | | | | | |
| O Use training set | | | | | | | | | | |
| O Supplied test set | Set | | | | | | | | | |
| Cross-validation Fold | ds 10 | | 00 | 🖯 Classifie | er evaluation opt | | | | | |
| • Percentage split | % 66 | | 🗹 Out | put model | | | | | | |
| More options. | •• | | 🗹 Out | put per-clas | s stats | | | | | |
| (Nom) class | | • | 🗌 Out | put entropy | evaluation measu | ures | | | | |
| Start | Stop | | 🗹 Out | put confusio | on matrix | | | | | |
| Result list (right-click for option | s) ——— | | 🗹 Stor | re prediction | s for visualizatio | n | | | | |
| | | | 🗌 Out | put text pre | dictions on test s | et | | | | |
| | | | Cos | t–sensitive | evaluation Set |) | | | | |
| | | | Randon | n seed for X | /al / % Split | 1 | | | | |
| | | | \square | | ОК | | | | | |
| -Status | | 비밀리 | | | | | | | | |

ОК

| Log | 100 | x 0 |
|-----|-----|-----|
|-----|-----|-----|



| 000 | Weka Knowledge Explorer | | | | | | | |
|------------------------------|-------------------------|----------|----------------|-----------|-------------------|-----------|---------|--|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | | |
| Classifier | | | | | | | | |
| Choose J48 -C | 0.25 -M 2 | | | | | | | |
| Test options | | Cla | ssifier output | | | | | |
| Use training set | | | | | | | | |
| Supplied test set | Set | | | | | | | |
| O Cross-validation | Folds 10 | | | | | | | |
| Percentage split | % 66 | | | | | | | |
| More op | tions | | | | | | | |
| | | | | | | | | |
| (Nom) class | | • | | | | | | |
| Start | Stop | | | | | | | |
| Result list (right-click for | options) | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| Chabus | | 1 | | | | | | |
| OK | | | | | | C | Log x 0 | |

| Preprocess Classify Cluster Associate Select attributes Visualize Classifier Choose J48 - C 0.25 - M 2 Test options Use training set Supplied test set Set Cross-validation Folds 10 Percentage split % 66 More options (Norm) class Start Stop Result list (right-click or options) | 000 | Weka Knowledge Explorer | | | | | | | |
|---|--|-------------------------|----------|----------------|-----------|-------------------|-----------|-----|-----|
| Classifier Choose J48 - C 0.25 - M 2 Test options Use training set Supplied test set Set Cross-validation Folds 10 e Percentage split % 66 More options (Norm) class Start Stop Result list (right-click er options) | | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | | |
| Choose J48 - C 0.25 - M 2 Test options Classifier output Use training set Supplied test set Supplied test set Set Cross-validation Folds 10 Percentage split % 66 More options (Nom) class Start Stop Result list (right-click ter options) (Nom) | Classifier | | | | | | | | |
| Test options Classifier output Use training set Supplied test set Supplied test set Set Cross-validation Folds 10 Percentage split % 66 More options (Nom) class Start Stop Result list (right-click or options) (Ist (right-click or options)) | Choose J48 -C | 0.25 -M 2 | | | | | | | |
| Use training set Supplied test set Cross-validation Folds 10 Percentage split % 66 More options (Nom) class Start Start Stop Result list (right-click ar options) | Test options | | Cla | ssifier output | | | | | |
| Supplied test set Cross-validation Folds 10 Percentage split % More options (Nom) class Start Stop Result list (right-click ar options) | 🔘 Use training set | | | | | | | | |
| Cross-validation Folds 10 Percentage split % 66 More options (Nom) class Start Stop Result list (right-click er options) | O Supplied test set | t Set | | | | | | | |
| More options (Nom) class Start Stop Result list (right-click for options) | Cross-validation Percentage split | Folds 10 % 66 | | | | | | | |
| (Nom) class Start Stop Result list (right-click/br options) | More op | otions | | | | | | | |
| Start Stop Result list (right-click for options) | (Nom) class | | • | | | | | | |
| - Status | Start Result list (right-click for | Stop | | | | | | | |
| OK Log x C | OK | | | | | | C | Log | x 0 |



-Status OK



× 0



-Status OK



× 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Log 🔨 🔨 x O

| | | | | | | | 1 | _ | | |
|---------------------------------------|--|--|---|--|--------------------------------|--------------------------------|---|---|--|--|
| Classifier | | | | | | | | | | |
| Choose J48 -C 0.25 -M 2 | | | | | | | | | | |
| Test options | Classifi | er output | | | | | | _ | | |
| O Use training set | Time | taken to b | wild model: | 0.24 seconds | | | | ŕ | | |
| O Supplied test set Set | 5 | valuation | on test enli | + === | | | | l | | |
| O Cross-validation Folds 10 | === St | mmary === | | | | | | l | | |
| • Percentage split % 66 | Corre | tly Class | ified Instan | ces | 49 | | 96.0784 % | | | |
| More options | Kappa Mean a | Correctly Classified Instances 49 96.0784 % Incorrectly Classified Instances 2 3.9216 % Kappa statistic 0.9408 0.0396 Mean absolute error 0.1579 | | | | | | | | |
| (Nom) class | Relat: Root | ive absolu celative s | te error quared error | | 8.8979 33.4091 | 8 8 | | | | |
| Start Stop | === De | etailed Ac | curacy By Cl | ass === | 51 | | | 1 | | |
| Result list (right-click for options) | | | | | | | | | | |
| 11:49:05 – trees.j48.J48 | TP Rat 1 0.88 === Co a 1 15 0 1 0 2 | ce FP Ra 0 0.0 20 0 0 0 0 a 0 b 2 15 c | lte Precisi 1 163 0.90 1 latrix === classified = Iris-setos = Iris-versi = Iris-virgi | on Recall 1 5 1 0.882 as a color nica | F-Measur 1 0.95 0.938 | e Clas Iris Iris Iris | s -setosa -versicolor -virginica | | | |

Status

OK

| Weka | Knowledge | Explorer |
|------|-----------|----------|
| | | |

| 00 | | | Weka Knowledge Explorer | | | | | | | |
|------------------------------|------------|----------|-------------------------|---------------|---------------|------------|------------|-----------------------------|--|--|
| | Preprocess | Classify | Cluster | Associate | Select attrib | outes | Visualize | | | |
| Classifier | | | | | | | | | | |
| Choose J48 -C | 0.25 -M 2 | | | | | | | | | |
| Test options | | Classifi | er output | | | | | | | |
| 🔘 Use training set | | Time | tokon to b | uild model. | 0.24 accords | | | | | |
| Supplied test se | t Set | Time | taken to p | uild model: | 0.24 seconds | 1 | | | | |
| | | ==== E | valuation | on test spli | t === | | | | | |
| Cross-validation | n Folds 10 | | unnary | - | | | | | | |
| Percentage split | % 66 | Corre | ctly Class | ified Instan | ces | 49 | | 96.0784 % | | |
| More opt | ions | Kappa | statistic | issified inst | ances | ∠ 0.940 |)8 | 3.9210 % | | |
| | | Mean | absolute e | error | | 0.039 | 96 | | | |
| (Nom) class | | Root : | mean squar | ed error | | 0.157 | 79 70 & | | | |
| (NOIII) Class | | Root | relative s | quared error | | 33.409 | 91 % | | | |
| | Chan | Total | Number of | Instances | | 51 | | | | |
| Start | Stop |) === D | atailed Ac | Curacy By Cl | agg === | | | | | |
| Result list (right-click for | r options) | | ecarrea no | curacy by or | 455 | | | | | |
| 11:49:05 - trees.j48 | .]48 | TP Ra | te FP Ra | te Precisi | on Recall | F-Meas | sure Clas | 55 | | |
| | | 1 | 0 0 | 1 | 5 1 | 1 | Iri: | s-setosa | | |
| | | 0.8 | 82 0 | 1 | 0.882 | 0.9 | 938 Iris | s-versicolor s-virginica | | |
| | | | | | | | | | | |
| | | ==== C | onfusion M | latrix === | | | | | | |
| | | a | b c < | classified | as | | | | | |
| | | 15 | 0 0 a | = Iris-setos | a | | | | | |
| | | 0 1 | 90 b | = Iris-versi | color | | | | | |
| | | 0 | Z 12 C | = iris-virgi | nica | | | | | |

Ă ▼

× 0

Log

Status OK

Weka Knowledge Explorer

Classify Cluster Associate Select attributes Visualize

| | Preprocess | Classify | Cluster | Associate | Select att | tributes | Visualize | | |
|------------------------------|------------|---------------------------|---|---|----------------|-----------------------------|-------------------------|---|---------|
| Classifier | | | | | | | | | |
| Choose J48 -C 0 |).25 –M 2 | | | | | | | | |
| Test options | | Classifi | er output | | | | | | |
| O Use training set | | | | | | | | | 6 |
| O Supplied test set | Set | Time | taken to b | uild model: | 0.24 secor | las | | | |
| O Cross-validation | Folds 10 | === S | ummary === | on test spir | | | | | |
| Percentage split | % 66 | Corre | ctly Class | ified Instan | ces | 49 | | 96.0784 | 8 |
| More opti | ons |) Incor Kappa Mean | statistic absolute e | ssified Inst | ances | 2 0.94 0.03 | 08 | 3.9216 | 8 |
| (Nom) class | | Relat Root Total | mean squar ive absolu relative s Number of | ed error te error quared error Instances | | 0.15 8.89 33.40 51 | 979 % 979 % 991 % | | |
| Besult list (right-click for | options) | _ === D | etailed Ac | curacy By Cl | ass === | | | | |
| 11:49:05 - trees.i48. | Vie | w in main | window | | Recal | l F-Mea | sure Cla | ass | |
| | Vie Sav | w in separ ve result b | rate windo uffer | w | 1 1 0.88 | 1 0. 32 0. | 95 Ir: 938 Ir: | is-setosa is-versicol is-virginic | or a |
| | Loa | ad model | | | | | | | |
| | Sav | ve model | | | | | | | |
| | Re | -evaluate i | model on o | current test s | et | | | | |
| | Vis | ualize clas | sifer error | 'S | | | | | Ă |
| | Vis | ualize tree |) | | | | | | |
| Status | Vis | ualize mar | rgin curve | V0 | | | 0 | | |
| OK | Vis | ualize cos | t curve | ve | - i | | C | Log | × U |



Weka Knowledge Explorer

Classify Cluster Associate Select attributes Visualize

| | Preprocess | Classify | Cluster | Associate | Select attrib | outes | Visualize | | |
|--------------------------------|-----------------------|--|---|--------------------------------------|---------------------------|----------------------|---------------------------------------|---|-------|
| Classifier | | | | | | | | | |
| Choose J48 -C | 0.25 –M 2 | | | | | | | | |
| Test options | | Classifi | er output | | | | | | |
| O Use training set | | | | | | | | | 6 |
| O Supplied test set | Set | Time 1 | valuation | on test spli | 0.24 seconds | | | | |
| O Cross-validation | Folds 10 | === S | ummary === | ou cesc spir | 6 | | | | |
| Percentage split More opti | % 66 ons | Corre Incor Kappa Mean | ctly Class rectly Cla statistic absolute e | ified Instan ssified Inst rror | ces ances | 49 2 0.9400 | 8 | 96.0784 % 3.9216 % | |
| (Nom) class Start | Stop | Root Relat Root Total | Root mean squared error0.1579Relative absolute error8.8979 %Root relative squared error33.4091 %Total Number of Instances51 | | | | | | |
| Result list (right-click for | options) | D | etailed Ac | curacy By Cl | ass === | | | | |
| 11:49:05 - trees.j48, | J48 View i Save r | in main wii in separate result buffe | ndow e window er | | Recall 1 1 0.882 | F-Meast 1 0.99 | ure Clas Iris 5 Iris 38 Iris | s -setosa -versicolor -virginica | |
| | Load Save Re-ev | model model valuate mo | del on cur | rent test set | | | | | |
| | Visua Visua | lize classif lize tree | er errors | | lor ca | | | | A V |
| Status | Visua | lize margir | n curve | | | | | | |
| ОК | Visua Visua | lize thresh lize cost cu | old curve | | ► | | C | Log 🧳 | × x 0 |



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Log 🔨 🔨 x O

| | | | | | | | 1 | _ | | | | | |
|---------------------------------------|--|---|---|--|--------------------------------|--------------------------------|---|---|--|--|--|--|--|
| Classifier | | | | | | | | | | | | | |
| Choose J48 -C 0.25 -M 2 | Choose J48 -C 0.25 -M 2 | | | | | | | | | | | | |
| Test options | Classifi | er output | | | | | | _ | | | | | |
| O Use training set | Time | taken to b | wild model: | 0.24 seconds | | | | ŕ | | | | | |
| O Supplied test set Set | 5 | === Evaluation on test split === | | | | | | | | | | | |
| O Cross-validation Folds 10 | === St | === Evaluation on test split === === Summary === | | | | | | | | | | | |
| • Percentage split % 66 | Corre | Correctly Classified Instances 49 96.0784 % | | | | | | | | | | | |
| More options | Kappa Mean a | Incorrectly Classified Instances 2 3.9216 % Kappa statistic 0.9408 Mean absolute error 0.0396 | | | | | | | | | | | |
| (Nom) class | Relat: Root | ive absolu celative s | te error quared error | | 8.8979 33.4091 | 8 8 | | | | | | | |
| Start Stop | === De | etailed Ac | curacy By Cl | ass === | 51 | | | 1 | | | | | |
| Result list (right-click for options) | | | | | | | | | | | | | |
| 11:49:05 – trees.j48.J48 | TP Rat 1 0.88 === Co a 1 15 0 1 0 2 | ce FP Ra 0 0.0 20 0 0 0 0 a 0 b 2 15 c | lte Precisi 1 163 0.90 1 latrix === classified = Iris-setos = Iris-versi = Iris-virgi | on Recall 1 5 1 0.882 as a color nica | F-Measur 1 0.95 0.938 | e Clas Iris Iris Iris | s -setosa -versicolor -virginica | | | | | | |

Status

OK

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

×0

Log

| Classifier Choose J48 - C 0.25 - M 2 Test options Use training set Cross-validation Folds 10 Percentage split % 66 More options (Nom) class Start Stop Result list (right-click for options) 11:49:05 - trees.j48.J48 Classifier output Time taken to build model: 0.24 seconds === Evaluation on test split === === Summary === Correctly Classified Instances 49 96.0784 % Incorrectly Classified Instances 2 3.9216 % Kappa statistic 0.9408 Root mean asguared error 0.1579 Relative asguared error 8.8979 % Root relative squared error 33.4091 % Total Number of Instances 51 === Detailed Accuracy By Class === TF Rate FF Rate Precision Recall F-Measure Class 1 0 1 1 1 Iris-setosa 1 0.063 0.905 1 0.95 Iris-versicolor 0.882 0 1 0.882 0.938 Iris-virginica === Confusion Matrix === a b c < classified as 15 0 0 a = Iris-versicolor 0 1 0 1 b = Iris-versicolor | | | 1 | | | | | | | | | |
|--|---------------------------------------|--|---|---|--|--|--|--|--|--|--|--|
| Choose J48 - C 0.25 - M 2 Test options Use training se Supplied test set Supplied test set Cross-validation Folds 10 Percentage split % 66 More options Choose J48 - C 0.25 - M 2 Cross-validation Folds 10 Percentage split % 66 More options Correctly Classified Instances 2 Appa statistic 0.9408 More options Nomy class Start Stop Result list (right-click for options) 11:49:05 - trees.j48.J48 TP Rate FP Rate Precision Recall F-Measure Class 1 0 1 0 1 0 1 0.063 0 1 1 0.063 0 1 0 1 0 1 0 1 0 1 0 1 1 0 0 1 1 0 1 < | Classifier | | | | | | | | | | | |
| Test options Classifier output Use training set Supplied test set Supplied test set Set Cross-validation Folds 10 Percentage split % 66 More options Correctly Classified Instances 49 Yes 96.0784 % Incorrectly Classified Instances 2 3.9216 % More options More absolute error 0.0396 Root mean squared error 0.1579 Result list (right-click for options) The Result list (right-click for options) The Rate FP Rate Precision Recall F-Measure Class 1 0 1 1 Iris-setosa 1 0 1 1 Iris-versicolor 0.882 0 1 0.938 Iris-viriginica == Confusion Matrix === a b c < classified as 15 0 a = Iris-versicolor 0.938 Iris-viriginica | Choose J48 -C 0.25 -M 2 | | | | | | | | | | | |
| Ouse training set Supplied test set Set Orcoss-validation Folds 10 Fine taken to build model: 0.24 seconds Percentage split % 66 More options Correctly Classified Instances 49 96.0784 % Incorrectly Classified Instances 2 3.9216 % More options Correctly Classified Instances 1 3.4091 % Total Number of Instances 51 | Test options | Classifier output | | | | | | | | | | |
| Supplied test set Set Cross-validation Folds 10 Incorrectly Classified Instances 49 96.0784 % More options Correctly Classified Instances 2 3.9216 % Resultlist (right-click for options) Feature absolute error 0.0396 800 f 11:49:05 - trees.j48.J48 0 1 1 1 Iris-setosa 1 0 1 1 1 Iris-setosa 1 0 0 1 0.882 0.938 1 Iris-virginica a b c < classified as 15 0 1 15< | O Use training set | Time taken to build model: 0.1 | 24 seconds | - | | | | | | | | |
| Cross-validation Folds 10 Image: Summary === Orecentage split % 66 More options Start Start Stop Result list (right-click for options) TP Rate I1:49:05 - trees.j48.J48 TP Rate Precentage split 0 1 0 1 0 1 0.063 0 1 1:1:49:05 - trees.j48.J48 | O Supplied test set | === Evaluation on test split === | | | | | | | | | | |
| • Percentage split % 66 More options Correctly Classified Instances 49 96.0784 % Incorrectly Classified Instances 2 | O Cross-validation Folds 10 | === Summary === | | | | | | | | | | |
| More options More options Display for the instances Display for the instances (Nom) class Image: Start Stop Relative absolute error 0.0396 Start Stop Start Stop Result list (right-click for options) Result list (right-click for options) 11:49:05 - trees.j48.J48 TP Rate FP Rate Precision Recall F-Measure Class 1 1 0 1 1 1 0.063 0.905 1 0.95 I 0.063 0.905 1 0.938 I 0 1 0.882 0.938 Iris-versicolor 0.882 0 1 0.882 0.938 Iris-virginica === Confusion Matrix === a b c < classified as 15 0 a Iris-versicolor 0 1 1 b Iris-versicolor 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 <th>• Percentage split % 66</th> <th>Correctly Classified Instance</th> <th>s 49 Ces 2</th> <th>96.0784 %</th> | • Percentage split % 66 | Correctly Classified Instance | s 49 Ces 2 | 96.0784 % | | | | | | | | |
| (Nom) class Image: Constraint of the second sec | More options | Kappa statistic Mean absolute error | 0.9408 | 5.5210 % | | | | | | | | |
| Start Stop Result list (right-click for options) Total Number of Instances 51 11:49:05 - trees.j48.J48 TP Rate Precision Recall F-Measure Class 1 0 1 1 1 Iris-setosa 1 0.063 0.905 1 0.95 Iris-versicolor 0.882 0 1 0.882 0.938 Iris-virginica === Confusion Matrix === a b c < classified as 15 0 a = Iris-setosa 0 0 9 b = Iris-versicolor 0 2 15 b = Iris-versicolor 0 a = Iris-versicolor | (Nom) class | Root mean squared error 0.1579 Relative absolute error 8.8979 % Root relative squared error 33.4091 % | | | | | | | | | | |
| Result list (right-click for options) 11:49:05 - trees.j48.J48 TP Rate FP Rate Precision Recall F-Measure Class 1 0 1 1 1 Iris-setosa 1 0.063 0.905 1 0.95 Iris-versicolor 0.882 0 1 0.882 0.938 Iris-virginica === Confusion Matrix === a b c < classified as 15 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0.92 15 a = Iris-versicolor 0.92 15 a = Iris-versicolor | Start Stop | Total Number of Instances === Detailed Accuracy By Class | 51 s === | ſ | | | | | | | | |
| 11:49:05 - trees.j48.J48 TP Rate FP Rate Precision Recall F-Measure Class 1 0 1 1 1 Iris-setosa 1 0.063 0.905 1 0.95 Iris-versicolor 0.882 0 1 0.882 0.938 Iris-virginica === Confusion Matrix === a b c < classified as 15 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0 2 15 classified as | Result list (right-click for options) | | | | | | | | | | | |
| 0 Z 15 C = IIIS-VIIgINICA | <u>11:49:05 – trees.j48.J48</u> | TP Rate FP Rate Precision 1 0 1 1 0.063 0.905 0.882 0 1 === Confusion Matrix === a a b c classified as 15 0 a Iris-setosa 0 19 0 b Iris-versico 0 2 15 c Iris-virginio | Recall F-Measure 1 1 1 0.95 0.882 0.938 lor ca | Class Iris-setosa Iris-versicolor Iris-virginica | | | | | | | | |

Status

OK

| $\Theta \Theta \Theta$ | | | Weka Kn | owledge Exp | olorer | | | | | | |
|------------------------------------|--------------------------|--------------|----------------|-------------|----------------|-------------|-----------|----------------------|--|--|--|
| | Preprocess | Classify | Cluster | Associate | Select at | tributes | Visualize | | | | |
| Classifier | | | | | | | | | | | |
| 📁 weka | | | | | | | | | | | |
| v 🗊 classifiers | | | | | | | | | | | |
| 📑 🔻 🧊 bayes | | | assifier out | put | | | | | | | |
| AODE | | | == Evalua | ation on te | st split == | | | 6 | | | |
| BayesNe | tK2 | | -= Summary === | | | | | | | | |
| BayesNe | tB | | arrectly | Classified | Instances | | 50 | 98 0392 8 | | | |
| NaiveBay | /es | | acorrectl | y Classifi | ed Instance | es | 1 | 1.9608 % | | | |
| NaiveBay | /esMultinomial | | appa stat | istic | | | 0.9704 | | | | |
| NaiveBay | /esSimple | | ean absol | ute error | | | 0.0239 | | | | |
| NaiveBay | esUpdateable | | elative a | absolute er | ror | | 5.3594 % | | | | |
| ▶ 🗍 functions | | | pot relat | ive square | d error | | 23.2952 % | | | | |
| lazy | | | otal Numb | per of Inst | ances | | 51 | | | | |
| 🕨 🕨 🧊 meta | | | == Detail | ed Accurac | y By Class | === | | | | | |
| 🖌 🕨 🧊 misc | | | | | | | | | | | |
| ► 🚺 trees | | | ? Rate | FP Rate | Precision 1 | Recall 1 | F-Measure | Class Tris-setosa | | | |
| rules | | | 1 | 0.031 | 0.95 | 1 | 0.974 | Iris-versicolor | | | |
| | | | 0.941 | 0 | 1 | 0.941 | 0.97 | Iris-virginica | | | |
| : | | | == Confus | sion Matrix | === | | | | | | |
| | | | | | 1.01 | | | | | | |
| | | | a b c | < clas | sified as | | | | | | |
| | | | 0 19 0 | b = Iri | s-versicold | or | | | | | |
| | | | 0 1 16 | c = Iri | s-virginica | 1 | | | | | |
| | | | | | | | | A | | | |
| | | | | | | | | | | | |
| 1 | | | | | | | | | | | |
| QuickTime™ and a TIFF (LZW) decomp | pressor are needed to se | e this pictu | | | | | | | | | |
| | | | | | | | C | Log XO | | | |

Weka Knowledge Explorer

| elect | attributes | Visu |
|-------|------------|------|
| | | |

×0

Log

| Preprocess | Classify | Cluster | Associat | te Select at | ttributes | Visualize | |
|---|------------|---|--|--|-------------|-----------------------------|-----------------------|
| Classifier | | | | | | | |
| Choose NaiveBayes | | | | | | | |
| Test options | | Classifier ou | tput | | | | |
| O Use training set | | === Evalu | ation on t | est split == | | | |
| O Supplied test set | t | Summa | LY | | | | |
| Cross-validation Folds 10 |) | Correctly Incorrect | Classifie ly Classif | ed Instances fied Instance | es | 50 1 | 98.0392 % 1.9608 % |
| Percentage split % 66 | i i | Kappa sta Mean abso Root mean | tistic lute error squared e | rror | | 0.9704 0.0239 0.1101 | |
| More options | | Relative Root relation | absolute e tive squar | error red error | | 5.3594 % 23.2952 % 51 | |
| (Nom) class | • | === Detai | led Accura | acy By Class | | | |
| Start St | op | TP Rate | FP Rate | Precision 1 | Recall 1 | F-Measure 1 | Class Iris-setosa |
| Result list (right-click for options) | | 1 | 0.031 | 0.95 | 1 | 0.974 | Iris-versicolor |
| 11:49:05 - trees.j48.J48 14:34:28 - functions.neural.Neu | ralNetwork | 0.941 === Confu a b c 15 0 0 0 19 0 0 1 16 | 0 sion Matri < cla a = Ir b = Ir c = Ir | l assified as ris-setosa ris-versicolo ris-virginica | 0.941 | 0.97 | Iris-virginica |

QuickTime[™] and a TIFF (LZW) decompressor are needed to see this pictur

 $\bigcirc \bigcirc \bigcirc$

Weka Knowledge Explorer

| Preprocess (| Classify | Cluster | Associate | Select attributes | Visualize |
|--------------|----------|---------|-----------|-------------------|-----------|
|--------------|----------|---------|-----------|-------------------|-----------|

| Classifier | | | | | | |
|---|------------------------|--------------------------|-------------------------------|--------|-----------------|--------------------------------|
| Choose NaiveBayes | | | | | | |
| Test options | Classifier ou | ıtput | | | | |
| Use training set | === Evalu | ation on t | est split == | | | |
| O Supplied test set Set | === Summa | ry === | | | | |
| Cross-validation Folds 10 | Correctly Incorrect | Classifie | ed Instances fied Instance | 23 | 50 1 | 98.0392 % 1.9608 % |
| Percentage split % 66 | Kappa sta | tistic | | | 0.9704 | |
| | Mean abso Root mean | squared e | error | | 0.0239 | |
| More options | Relative | absolute e | error | | 5.3594 % | |
| | Total Num | tive squar ber of Ins | tances | | 23.2952 % 51 | |
| (Nom) class | D. I. I. | | 2 | | | |
| | === Detai | led Accura | acy By Class | | | |
| Start Stop | TP Rate | FP Rate | Precision | Recall | F-Measure | Class |
| Result list (right-click for options) | 1 | 0.031 | 0.95 | 1 | 0.974 | Iris-setosa Iris-versicolor |
| 11:49:05 – trees.j48.J48 | 0.941 | 0 | 1 | 0.941 | 0.97 | Iris-virginica |
| 14:34:28 - functions.neural.YeuralNetwork | === Confu | sion Matr: | x === | | | (|
| • • • • • • • • • • • • • • • • • • • | abo | < cla | assified as | | | |
| | 15 0 0 | a = I1 | cis-setosa | | | |
| | 0 19 0 | $ b = I_1$ | cis-versicolo | or | | |
| | 0 1 10 | | in triginitio | - | | |
| | | | | _ | | |
| 1 | | | | | | |

QuickTime[™] and a TIFF (LZW) decompressor are needed to see this pictur



+

x 0

 $\mathbf{\Theta} \mathbf{\Theta} \mathbf{\Theta}$

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Log 💉 🗴 V

| | Treprocess | clussily | cluster | Associate | Sciectur | tinbutes | Visualize | | | | | |
|---|------------------------------|----------|--|-------------|--------------|----------|-----------|--------------------------------|--|--|--|--|
| Classifier | | | | | | | | | | | | |
| Choose NaiveB | ayes | | | | | | | | | | | |
| Test options | | | Classifier ou | tput | | | | | | | | |
| 🔘 Use training set | | | === Evalua | ation on te | est split == | | | \sim | | | | |
| Supplied test se | t Set | | === Summar | cy === | | | | | | | | |
| Cross-validation | n Folds 10 | | Correctly Classified Instances 48 94.1176 % Incorrectly Classified Instances 3 5.8824 % | | | | | | | | | |
| Percentage split | % 66 | | Kappa statistic 0.9113 Mean absolute error 0.0447 | | | | | | | | | |
| More | options | | Root mean squared error0.1722Relative absolute error10.0365 %Root relative squared error36.4196 % | | | | | | | | | |
| (Nom) class | | + | Total Numb | per of Inst | ances | | 51 | | | | | |
| | | | === Detai. | led Accurac | y By Class | === | | | | | | |
| Start |) Stop | | TP Rate | FP Rate | Precision | Recall | F-Measure | Class | | | | |
| Result list (right-click fo | r options) | | 0.947 | 0.063 | 0.9 | 0.947 | 0.923 | Iris-setosa Iris-versicolor | | | | |
| 11:49:05 - trees.j48 | .J48 | | 0.882 | 0.029 | 0.938 | 0.882 | 0.909 | Iris-virginica | | | | |
| 14:34:28 - function: 14:48:05 - bayes.Na | s.neural.Neural aiveBayes | Network | === Confus | sion Matrix | c === | | | | | | | |
| | | | a b c < classified as 15 0 0 a = Iris-setosa 0 18 1 b = Iris-versicolor 0 2 15 c = Iris-virginica | | | | | | | | | |
| | | | | | | | |) 4 ► | | | | |
| | | | | | | | | | | | | |

Status OK

Weka Knowledge Explorer

Classify Cluster Associate Select attributes Visualize

Log 💉 🗴 O

| Preprocess | Classify | Cluster | Associate | e Select at | tributes | Visualize | |
|---|-----------|--|--|--|-------------------------------|----------------------------------|---|
| Classifier Choose NaiveBayes | | | | | | | |
| Test options | | Classifier out | put | | | | |
| Use training set | | === Evalua | tion on te | est split == | | | <u>_</u> |
| O Supplied test set Set | | === Summar | :y === | Tratanaa | | 40 | 04 1126 8 |
| Cross-validation Folds 10 | | Incorrectly | y Classified | ied Instances | es | 3 | 5.8824 % |
| Percentage split % 66 | | Kappa stat Mean absol | istic ute error | | | 0.9113 | |
| More options | \supset | Root mean Relative a Root relat | squared en bsolute en tive square | rror cror ed error | | 0.1722 10.0365 % 36.4196 % | |
| (Nom) class | ÷ | Detail | ed legura | Bu Class | | 51 | |
| Start Stop Result list (right-click for options) 11:49:05 - trees.j48.J48 14:34:28 - functions.neural.NeuralN 14:48:05 - bayes.NaiveBayes | etwork | <pre>TP Rate 1 0.947 0.882 === Confus a b c 15 0 0 0 18 1 0 2 15</pre> | FP Rate 0 0.063 0.029 sion Matrix < clas a = Iri b = Iri c = Iri | Precision 1 0.9 0.938 c === ssified as is-setosa is-versicolo is-virginica | Recall 1 0.947 0.882 | F-Measure 1 0.923 0.909 | Class Iris-setosa Iris-versicolor Iris-virginica |

ОК

 $\mathbf{\Theta} \mathbf{\Theta} \mathbf{\Theta}$

Weka Knowledge Explorer

Classify Cluster Associate Select attributes Visualize

| | Preprocess | Classify | Cluster | Associate | Select at | tributes | Visualize | |
|---|--|--|---------------------------------------|------------------------------------|--|-------------------------------|----------------------------------|---|
| Classifier | | | | | | | | |
| Choose NaiveBa | ayes | | | | | | | |
| Test options | | | Classifier out | put | | | | |
| 🔘 Use training set | | | === Evalua | ation on tes | t split == | = | | <u> </u> |
| O Supplied test set | Set | .) | === Summar | cy === | | | | |
| Cross-validation Percentage split | Folds 10 % 66 | | Correctly Incorrectl Kappa stat | Classified y Classifie istic | Instances d Instance | S | 48 3 0.9113 0.0447 | 94.1176 % 5.8824 % |
| More o | options | | Root mean Relative a Root relat | squared err absolute err | or or error | | 0.1722 10.0365 % 36.4196 % | |
| (Nom) class | | • | Total Numb | er of Insta led Accuracy | nces By Class | | 51 | |
| Start Result list (right-click for 11:49:05 - trees.j48. 14:34:28 - functions 14:48:05 - bayes.Na | View in ma View in sep Save result Load mode Save mode | in window parate wind t buffer el el | dow | P x | recision 1 0.9 0.938 === ified as | Recall 1 0.947 0.882 | F-Measure 1 0.923 0.909 | Class Iris-setosa Iris-versicolor Iris-virginica |
| | Visualize c Visualize t Visualize n Visualize t | lassifer eri ree nargin cur hreshold c | rors ve | :is :is | -setosa -versicolo -virginica ris-setosa | r | | ▲ ▼ ▼ |
| Status OK | Visualize c | ost curve | | ► | ris–versico ris–virgini | olor ca | C | Log 💉 🔊 |



Explorer: Clustering

• WEKA contains many clustering implementations:

Works with both discrete and numerical data

Example of K-means





| 0 0 | | Weka Explorer | | |
|---------------------------------------|----------------------------|-------------------|-------------------|-----------|
| | Preprocess Classify Clus | ster Associate | Select attributes | Visualize |
| Clusterer | | | | |
| Choose EM -I | 100 -N -1 -M 1.0E-6 -S 100 | | | |
| Cluster mode | | Clusterer output- | | |
| 💿 Use training set | | | | |
| Supplied test set | Set | | | |
| O Percentage split | % 66 | | | |
| O Classes to cluste | ers evaluation | | | |
| (Nom) class | * * | | | |
| Store clusters for | r visualization | | | |
| Ignor | re attributes | | | |
| Start | Stop | | | |
| Result list (right-click | k for options) | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| Status | | | | |

| Status | | |
|--------|-----|-----|
| OK | Log | × 0 |
| | | |

Weka Explorer 0 Preprocess Classify Cluster Associate Select attributes Visualize Clusterer Choose EM -I 100 -N -1 -M 1.0E-6 -S 100 Cluster mode Clusterer output Use training s Set... Supplied test set Percentage split % 66 Classes to clusters evaluation (Nom) class * Store clusters for visualization Ignore attributes Start Stop Result list (right-click for options)

| | Status | | |
|-------|--------|-----|-----|
| . YO. | OK | Log | × 0 |

| 0 0 | | | We | ka Explorer | | | |
|------------------------|-------------------|--------------|--------------|--------------|------------------------|-----------|--|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | |
| Clusterer | | | | | | | |
| Choose Simp | oleKMeans -N | 1 2 - A "wek | a.core.Eucli | deanDistance | e -R first-last" -I 50 | 0 -S 10 | |
| Cluster mode | | | Clus | terer output | | | |
| 🖲 Use training se | | | | | | | |
| O Supplied test s | et | Set | | | | | |
| O Percentage spli | it | % 66 | | | | | |
| O Classes to clus | ters evaluatio | n | | | | | |
| (Nom) class | | A V | | | | | |
| Store clusters f | or visualizatio | on | | | | | |
| Ign | ore attributes | | \supset | | | | |
| Start | $\supset \square$ | Stop | \supset | | | | |
| Result list (right-cli | ick for options | ;) | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

| Status | | |
|--------|-----|-------|
| OK | Log | ••••• |

| 000 | | | | We | eka Explorer | |
|-------------------|------------------------------|------------|-------------------|--------------|---------------|---|
| | P | reproces | s Classify | Cluster | Associate | Select attributes Visualize |
| Clusterer | | | | | | |
| Choose | \varTheta 🔿 🔿 🛛 weka. | gui.Gene | ericObjectEdito | or | | -S 10 |
| -Cluster mode- | weka.clusterers.SimpleKMea | ns | | | | |
| Use trainir | About | | | | | -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S |
| Supplied t | Cluster data using the k mea | ns algorit | hm. | | More | |
| O supplied t | | | | Cap | abilities | |
| O Percentage | | | | | | |
| Classes to | displayStdDevs | False | | | \$ | |
| | distanceFunction | Cho | ose Euclida | anDistan | n _ P first_l | a ba set === |
| Store clus | distancer unetion | | Lucinu | canoistan | | |
| | dontReplaceMissingValues | False | | | ; | |
| C Start | maxIterations | 500 | | | | |
| Start | | | | | | th mean/mode |
| Result list (righ | numClusters | 3 | | | | Cluster# |
| 15:11:12 - 500 | preserveInstancesOrder | False | | | ÷ | 0 1 (100) (50) |
| | hees | 10 | | | | 6.262 5.006 |
| | 3660 | 10 | | | | 2.872 3.418 4.906 1.464 |
| | Open Save | e | ОК | $ \subset $ | Cancel | Iris-versicolor Iris-setosa |
| | | | | - | | |
| | | | Time taken t | o build mo | del (full tr | caining data) : 0.02 seconds |
| | | | === Model and | d evaluati | on on traini | ing set === |
| | | | Clustered In | stances | | |
| | | | 0 100 (1 50 (| 67%) 33%) | | |
| | | | (| , | | |
| | | | (| | | |
| Status | | | | | | |
| OK | | | | | | Log x |

Weka Explorer

| Prepr | ocess Classify Cluster Associate Select attributes Visualize | | | | | | | | |
|--|---|--|--|--|--|--|--|--|--|
| Clusterer | | | | | | | | | |
| Choose SimpleKMeans -N 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10 | | | | | | | | | |
| Cluster mode | Clusterer output | | | | | | | | |
| Use training set Supplied test set Percentage split Classes to clusters evaluation (Nom) class Store clusters for visualization | Relation: iris Instances: 150 Attributes: 5 sepallength petallength class Test mode:evaluate on training data === Model and evaluation on training set === | | | | | | | | |
| Ignore attributes Start Stop | <pre>kMeans Number of iterations: 3 Within cluster sum of squared errors: 7.817456892309574 Missing values globally replaced with mean/mode</pre> | | | | | | | | |
| Result list (right-click for options) | Cluster centroids: Cluster# | | | | | | | | |
| 15:11:12 - SimpleKMeans | Attribute Full Data 0 1 2 (150) (50) (50) (50) | | | | | | | | |
| 15:12:39 – SimpleKMeans | sepallength 5.8433 5.936 5.006 6.588 sepalwidth 3.054 2.77 3.418 2.974 petallength 3.7587 4.26 1.464 5.552 petalwidth 1.1987 1.326 0.244 2.026 class Iris-setosa Iris-versicolor Iris-setosa Iris-virginica | | | | | | | | |
| | Time taken to build model (full training data) : 0 seconds | | | | | | | | |
| | === Model and evaluation on training set === | | | | | | | | |
| | Clustered Instances | | | | | | | | |
| | 0 50 (338) 1 50 (338) 2 50 (338) | | | | | | | | |
| | | | | | | | | | |
| itatus OK | Log " , | | | | | | | | |

Computin





Explorer: Finding Associations

- WEKA contains an implementation of the Apriori algorithm for learning association rules
 - Works only with discrete data
- Can identify statistical dependencies between groups of attributes:
 - o milk, butter ⇒ bread, eggs (with confidence 0.9 and support 2000)
- Apriori can compute all rules that have a given minimum support and exceed a given confidence





| 000 | 🕽 🖯 🕤 🦳 Weka Knowledge Explorer | | | | | | | | |
|------------------------------|---------------------------------|------------|------------|---------------|-----------|----------|---------|-----|-------|
| | Preprocess (| lassify | Cluster | Associate | Select at | tributes | Visuali | ze | |
| Associator | | | | | | | | | |
| Choose Apriori | -N 10 -T 0 -C 0 | .9 -D 0.05 | -U 1.0 -I | M 0.1 -S -1.0 | 0 | | | | |
| | | Assoc | ator outpu | Jt | | | | | |
| Start | Stop | | | | | | | | |
| Result list (right-click for | r options) | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| Status | | | | | | | | 100 | 0 x 6 |
| OK | | | | | | | | LUG | |

| 000 | | Weka Knowledge Exp | lorer | |
|------------------------------|--------------------------|--------------------------|-------------------|-----------|
| | Preprocess Classify | Cluster Associate | Select attributes | Visualize |
| Associator | | | | |
| Choose Apriori | -N 10 -T 0 -C 0.9 -L 0.0 | 05 -U 1.0 -M 0.1 -S -1.0 |) | |
| | As | sociator output | | |
| Start | Stop | | | |
| Result list (right-click for | options) | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| Status | 4_4 | | | Y |
| ОК | | | | Log 💉 X 0 |
| 000 | Weka Knowledge Explorer | | | | | |
|--|--|--------------------------------------|-----------------|---------------|--|--|
| | Preprocess Classify Cluster | Associate Select attri | ibutes Visualiz | e | | |
| Open file | Open URL O | pen DB | Undo | Save | | |
| Filter Choose None | | | | Apply | | |
| Current relation Relation: vote | Attributos: 17 | Selected attribute Name: handicap | ped-infants | Type: Nominal | | |
| Attributes No. | Name | Label | 236 | Count | | |
| 1 handicapped-infants 2 water-project-cost-sharing 3 adoption-of-the-budget-resolution 4 physician-fee-freeze 5 el-salvador-aid 6 religious-groups-in-schools | | y 187 | | | | |
| 7 anti-sate 8 aid-to-n 9 mx-miss 10 immigrat 11 synfuels 12 education | icaraguan-contras icaraguan-contras ile ion -corporation-cutback n-spending | Colour: Class (Nom) Visualize | | | | |
| 13 Superior 14 crime 15 duty-free 16 export-a 17 Class | e-exports dministration-act-south-africa | | | | | |
| -Status OK | | | (| Log 💉 X O | | |

| 🖯 🖯 🧶 Weka Knowledge Explorer | | | | | | | |
|--|---|--|--|--|--|--|--|
| | Preprocess Classify Cluster A | ssociate Select attributes Visualize | | | | | |
| Op | Open file Open URL Open DB Undo Save | | | | | | |
| Choose | Choose None Apply | | | | | | |
| Current re Relatio | lation on: vote es: 435 Attributes: 17 | Selected attribute Name: handicapped-infants Type: Nominal Missing: 12 (3%) Distinct: 2 Unique: 0 (0%) | | | | | |
| No. Name 1 handicapped-infants 2 water-project-cost-sharing 3 adoption-of-the-budget-resolution 4 physician-fee-freeze 5 el-salvador-aid 6 religious-groups-in-schools 7 anti-satellite-test-ban 8 aid-to-nicaraguan-contras 9 mx-missile 10 immigration | | n 236 y 187 Colour: Class (Nom) Visualize All | | | | | |
| 11 12 13 14 15 16 17 | synfuels-corporation-cutback education-spending superfund-right-to-sue crime duty-free-exports export-administration-act-south-africa Class | | | | | | |
| OK | | Log 💉 🛛 | | | | | |

| 000 | 🖯 🖯 Weka Knowledge Explorer | | | | | | | | |
|------------------------------|-----------------------------|------------|------------|---------------|------------|---------|----------|-----|--|
| | Preprocess C | lassify | Cluster | Associate | Select att | ributes | Visualiz | e | |
| Associator | Associator | | | | | | | | |
| Choose Apriori | -N 10 -T 0 -C 0. | .9 -D 0.05 | -U 1.0 -I | M 0.1 -S -1.0 |) | | | | |
| | | Assoc | ator outpu | ıt | | | | | |
| Start | Stop | | | | | | | | |
| Result list (right-click for | r options) | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| Status | | | | | | | (| Log | |
| UK | | | | | | | (| | |

| 000 | | Weka Knowledge Ex | olorer | |
|------------------------------|--------------------------|------------------------|-------------------|-----------|
| | Preprocess Classify | Cluster Associate | Select attributes | Visualize |
| Associator | | | | |
| Choose Apriori | -N 10 -T 0 -C 0.9 -D 0.9 | 05 -U 1.0 -M 0.1 -S -1 | 0 | |
| | | sociator output | | |
| Start | Stop | | | |
| Result list (right-click for | options) | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| ⊂Status | 131 | | | |
| OK | | | | Log 💉 X 0 |

| 00 | Weka Knowledge Explorer | | | | | | |
|--|--|--|--|--|--|--|--|
| Pre | process Classify Cluster Associate Select attributes Visualize | | | | | | |
| Associator | | | | | | | |
| Choose Apriori -N 1 | 0 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 | | | | | | |
| Cture Stop | Associator output | | | | | | |
| Result list (right-click for optic | Minimum metric <confidence>: 0.9 Number of cycles performed: 11</confidence> | | | | | | |
| 16:29:37 - Apriori Generated sets of large itemsets: | | | | | | | |
| | Size of set of large itemsets L(1): 20 | | | | | | |
| | Size of set of large itemsets L(2): 17 | | | | | | |
| | Size of set of large itemsets L(3): 6 | | | | | | |
| Size of set of large itemsets L(4): 1 | | | | | | | |
| | <pre>1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-cont 3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210 4. physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201 conf: (5. physician-fee-freeze=n 247 ==> Class=democrat 245 conf: (0.99) 6. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197 conf: 7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204 conf: (0.98) 8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 20 9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197 conf: 10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210</pre> | | | | | | |
| | | | | | | | |
| Status OK | Log 💉 X O | | | | | | |

Explorer: Attribute Selection

- Panel that can be used to investigate which (subsets of) attributes are the most predictive ones
- Attribute selection methods contain two parts:
 - A search method:
 - best-first, forward selection, random, exhaustive, genetic algorithm, ranking
 - An evaluation method:
 - × correlation-based, wrapper, information gain, chi-squared, ...
- Very flexible: allows arbitrary combinations of these two





| 000 | | | Weka Kn | owledge Expl | orer | | |
|---|--|----------|----------------|--------------|-------------------|-----------|--|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | |
| Attribute Evaluator | | | | | | | |
| Choose CfsSub | setEval | | | | | | |
| Search Method | | | | | | | |
| Choose BestFirs | st -D 1 -N 5 | | | | | | |
| Attribute Selection Mode | e | Attr | ibute selectio | on output | | | |
| Use full training Cross-validation | O Cross-validation Folds 10 Seed 1 | | | | | | |
| (Nom) Class | | • | | | | | |
| Start | Stop | | | | | | |
| Result list (right-click for options) | | | | | | | |
| OK | | | | | | | |

1000

| 00 |) 🖯 Weka Knowledge Explorer | | | | | | | |
|------------------------------|-----------------------------|----------|----------------|-----------|-------------------|-----------|--|---|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | | |
| Attribute Evaluator | | | | | | | | |
| Choose CfsSubs | setEval | | | | | | | |
| Search Method | | | | | | | | |
| Choose BestFirs | st -D 1 -N 5 | | | | | | | |
| Attribute Selection Mode | 9 | Attri | ibute selectio | on output | | | | |
| • Use full training | set | | | | | | | |
| O Cross-validation | Folds 1 | .0 | | | | | | |
| | 2660 1 | · | | | | | | |
| (Nom) Class | | • | | | | | | |
| Start | Stop | | | | | | | |
| Result list (right-click for | options) | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| Status | | 11 | | | | | | 1 |
| Log 💉 X O | | | | | | | | |



Status

OK





Status

OK

Log x 0

 $\mathbf{\Theta} \mathbf{\Theta} \mathbf{\Theta}$

OK

Weka Knowledge Explorer

Associate

Cluster

Classify

Preprocess

Select attributes

| Attribute Evaluator | |
|--------------------------------|--|
| 🕼 weka | |
| attributeSelection | |
| S CfsSubsetEval | |
| ClassifierSubsetEval | |
| WrapperSubsetEval | |
| A ConsistencySubsetEval | te selection output |
| ReliefFAttributeEval | duty-free-exports |
| InfoGainAttributeEval | export-administration-act-south-africa |
| GainRatioAttributeEval | Class ation mode: evaluate on all training data |
| SymmetricalUncertAttributeEval | |
| OneRAttributeEval | |
| ChiSquaredAttributeEval | Attribute Selection on all input data === |
| PrincipalComponents | |
| SVMAttributeEval | th Method: |
| F | Start set: no attributes |
| 1 | Search direction: forward |
| | Stale search after 5 node expansions |
| | Merit of best subset found: 0.729 |
| | |
| | bute Subset Evaluator (supervised, Class (nominal): 17 Class): |
| | CFS Subset Evaluator |
| Sele | cted attributes: 4 : 1 |
| | physician-fee-freeze |
| | * |
| | * |
| Status | |



| 000 | Weka Knowledge Explorer | | | |
|---|--|--|--|--|
| Preprocess Classify | Cluster Associate Select attributes Visualize | | | |
| Attribute Evaluator | | | | |
| Choose InfoGainAttributeEval | | | | |
| Search Method | | | | |
| weka | 308 -N -1 | | | |
| BestFirst | te selection output | | | |
| ForwardSelection RaceSearch GeneticSearch RandomSearch ExhaustiveSearch Ranker RankSearch | <pre>duty-free-exports export-administration-act-south-africa Class pation mode: evaluate on all training data Attribute Selection on all input data === th Method: Best first. Start set: no attributes Search direction: forward Stale search after 5 node expansions Total number of subsets evaluated: 83 Merit of best subset found: 0.729 ibute Subset Evaluator (supervised, Class (nominal): 17 Class): CFS Subset Evaluator pted attributes: 4 : 1 physician-fee-freeze *</pre> | | | |
| Status | | | | |



OK





| 0 | |
|---|-----|
| | ĸ |
| - | n l |

Explorer: Data Visualization

Visualization very useful in practice:
 e.g. helps to determine difficulty of the learning problem

WEKA can visualize single attributes and pairs of attributes
 To do: rotating 3-d visualizations (Xgobi-style)

- Color-coded class values
- "Jitter" option to deal with nominal attributes (and to detect "hidden" data points)
- "Zoom-in" function





| 000 | Weka Knowledge Explorer | | | | | |
|--|--------------------------------|---|--|--|--|--|
| | Preprocess Classify Cluster As | ssociate Select attributes | Visualize | | | |
| Open file | Open URL Oper | n DB Undo | Save | | | |
| Current relation Relation: Glass Instances: 214 | Attributes: 10 | Selected attribute Name: RI Missing: 0 (0%) Distinct: | Type: Numeric 178 Unique: 145 (68%) | | | |
| Attributes No. 1 RI 2 Na 3 Mg 4 AI 5 Si 6 K | Name | Statistic Minimum Maximum Mean StdDev | Value 1.511 1.534 1.518 0.003 | | | |
| 7 Ca 8 Ba 9 Fe 10 Type | | Colour: Type (Nom) | 9 5 1 .52 Visualize All | | | |
| OK | | | Log 💉 x 0 | | | |

000

Weka Knowledge Explorer

Visualize Preprocess Cluster Associate Select attributes Classify Plot Matrix RI Na Mg AI Si κ Type Fe ļ, • A 4 1 A 7 PlotSize: [100] PointSize: [1] Update Jitter: Select Attributes SubSample % : + Colour: Type (Nom) 100 Class Colour build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps Status OK



000

Weka Knowledge Explorer













Class colour

build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps



Class colour

build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps

| 000 |) | Weka Knowledge Explorer: Vis | sualizing Glass |
|----------------------|-----------------------------|--|----------------------------------|
| X: AI (| Num) | Y: Ca (Num) | • |
| Colou | r: Type (Nom) | Rectangle | • |
| Re | eset C | Clear Save | Jitter 🔵 |
| ⊢Plot: Glas | SS | | |
| 9.77 8.8- | × × × × × × × × | \times | <pre></pre> |
| 7.83 | 5 | 1.27 | × 1.49 |
| -Class col | lour | | |
| build wi vehic wi | ind float ind non-float | build wind non-float containers tablewa | vehic wind float re headlamps |



Performing Experiments

- Experimenter makes it easy to compare the performance of different learning schemes
- For classification and regression problems
- Results can be written into file or database
- Evaluation options: cross-validation, learning curve, hold-out
- Can also iterate over different parameter settings
- Significance-testing built in!





| 🖯 🖯 🧶 Weka Experiment Environment | | | | |
|-----------------------------------|--------|--------------------------------------|-----------------|--|
| Set | up Ru | in Analyse | | |
| Experiment Configuration Mode: | | Simple | Advanced | |
| Open | Sa | ve | New | |
| Results Destination | | | | |
| JDBC database 🗘 Filename: | | | Browse | |
| Experiment Type | | Iteration Control | | |
| Cross-validation | A T | Number of repetitions | | |
| Number of folds: | | Data sets first | | |
| Classification Regression | | Algorithms first | | |
| Datasets | | Algorithms | | |
| Add new Delete selected | d | Add new | Delete selected | |
| Notes | | | | |

| 🖯 🖯 🔿 Weka Experiment Environment | | | | |
|--|--------|--------------------------------------|-----------------|--|
| Setup | Ru | n Analyse | | |
| Experiment Configuration Mode: | | Simple | Advanced | |
| Open | Sa | ve | New | |
| Results Destination | | | | |
| JDBC database 🗘 Filename: | | | Browse | |
| Experiment Type | | Iteration Control | | |
| Cross-validation | × v | Number of repetitions: | | |
| Number of folds: | | Data sets first | | |
| Classification Regression | | Algorithms first | | |
| Datasets | | Algorithms | | |
| Add new Delete selected | | Add new | Delete selected | |
| | | | | |
| | | | | |

|) 🖯 🖯 Weka Experiment Environment | | | | |
|---|---|--|--|--|
| Setup R | un Analyse | | | |
| Experiment Configuration Mode: | Simple Advanced | | | |
| Open Sa | ave New | | | |
| Results Destination | | | | |
| JDBC database 🗘 URL: jdbc:idb=experiments.prp | Browse) | | | |
| Experiment Type | Iteration Control | | | |
| Cross-validation 🗘 | Number of repetitions: 10 | | | |
| Number of folds: 10 | Oata sets first | | | |
| Classification Regression | Algorithms first | | | |
| Datasets | Algorithms | | | |
| Add new Delete selected | Add new Delete selected | | | |
| Use relative paths | 148 -C 0.25 -M 2 | | | |
| /Users/eibe/Documents/datasets/UCI/iris.arff | NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a | | | |
| /Users/eibe/Documents/datasets/UCI/vote.arff | NaiveBayes | | | |
| /Users/eibe/Documents/datasets/UCI/glass.arff | | | | |
| | | | | |
| Notes | 9 PL 3 | | | |
| | | | | |
| | | | | |
| | | | | |
| 1 1 | | | | |

| 🖯 🖯 Weka Experim | Weka Experiment Environment | | | |
|--|---|--|--|--|
| Setup R | un Analyse | | | |
| Experiment Configuration Mode: | Shaple Advanced | | | |
| Open Sa | ave New | | | |
| Results Destination | | | | |
| JDBC database 🗘 URL: jdbc:idb=experiments.prp | Browse) | | | |
| Experiment Type | Iteration Control | | | |
| Cross-validation 🗘 | Number of repetitions: 10 | | | |
| Number of folds: 10 | Data sets first | | | |
| Classification Regression | Algorithms first | | | |
| Datasets | Algorithms | | | |
| Add new Delete selected | Add new Delete selected | | | |
| Use relative paths | 148 - C 0 25 - M 2 | | | |
| /Users/eibe/Documents/datasets/UCI/iris arff | NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a | | | |
| /Users/eibe/Documents/datasets/UCI/vote.arff | NaiveBayes | | | |
| /Users/eibe/Documents/datasets/UCI/glass.arff | | | | |
| | | | | |
| | | | | |
| Notes | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

| $\Theta \Theta \Theta$ | 😑 🖯 Weka Experiment Environment | | | |
|------------------------|---------------------------------|---|--|--|
| | Setup Run Analyse | | | |
| Start | Stop |) | | |
| Log | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| Status | | | | |
| Not running | | | | |

| 000 | Weka Experiment Envi | ronment |
|-------------|----------------------|---------|
| (| Start | Stop |
| Log | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| Status | | |
| Not running | | |

| 00 | Weka Experiment Environment | | | |
|--|-----------------------------|--|--|--|
| | Setup Run Analyse | | | |
| Start | Stop | | | |
| Log | | | | |
| 10:33:04: Started 13:41:15: Finished 13:41:15: There were 0 errors | | | | |
| Status | | | | |
| Not running | | | | |

| 00 | Weka Experiment Environment | | |
|--|-----------------------------|--|--|
| | Setup Run Analyse | | |
| Start | Stop | | |
| Log | | | |
| 10:33:04: Started 13:41:15: Finished 13:41:15: There were 0 errors | | | |
| Status | | | |
| Not running | | | |

| 000 | We | Weka Experiment Environment | | | |
|----------------------|-------------|-----------------------------|---------|---------------------|--|
| | | Setup Run | Analyse | | |
| Source | | | | | |
| No source | | | File D | atabase) Experiment | |
| Configure test | | Test output | | | |
| Row key fields | Select keys | | | | |
| Run field | Å T | | | | |
| Column key fields | Select keys | | | | |
| Comparison field | Å V | | | | |
| Significance 0.05 | | | | | |
| Test base | Select base | | | | |
| Show std. deviations | | | | | |
| Perform test | Save output | | | | |
| Result list | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| 000 | W | eka Expe | eriment I | Invironmer | nt |
|----------------------|-------------|----------|-----------|------------|--------------------------|
| | | Setup | Run | Analyse | |
| Source | | | | | |
| No source | | | | | File Database Experiment |
| Configure test | | Test out | tput | | |
| Row key fields | Select keys | | | | |
| Run field | (A Y | | | | |
| Column key fields | Select keys | | | | |
| Comparison field | ▲ ▼ | | | | |
| Significance | 0.05 | | | | |
| Test base | Select base | | | | |
| Show std. deviations | | | | | |
| Perform test | Save output | | | | |
| Result list | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |



Conclusion: Try It Yourself!

WEKA is available at

http://www.cs.waikato.ac.nz/ml/weka

Launch WEKA:

- Enter in the weka directory within a terminal
- java –Xmx2048m –jar weka.jar



