Data Mining with WEKA

Original author: unknown. ? The WEKA team Additional material: Tim Menzies, 2010

WEKA

- Machine learning/data mining software written in Java
 - Used for research, education, and applications
 - Complements "Data Mining" by Witten & Frank
- Main features
 - Comprehensive set of data pre-processing tools, learning alg orithms and evaluation methods
 - Graphical user interfaces (incl. data visualization)
 - Environment for comparing learning algorithms

Access

WEKA is available at

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

- Also has a list of projects based on WEKA
- WEKA contributors:

Abdelaziz Mahoui, Alexander K. Seewald, Ashraf M. Kibriya, Bern hard Pfahringer, Brent Martin, Peter Flach, Eibe Frank, Gabi Schmidb erger, Jan H. Witten, J. Lindgren, Janice Boughton, Jason Wells, Len Trigg, Lucio de Souza Coelho, Malcolm Ware, Mark Hall, Remco Bou ckaert, Richard Kirkby, Shane Butler, Shane Legg, Stuart Inglis, Sylva in Roy, Tony Voyle, Xin Xu, Yong Wang, Zhihai Wang

Data Files

@relation heart-disease-simplified

@attribute age numeric

@attribute sex { female, male}

@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angi
 na}

@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

numeric attribute

nominal attribute





Explorer: pre-processing

Source

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

00	Weka Knowle	edge Explorer		
	Preprocess Classify Cluster As	ssociate Select at	tributes Visualize	
Open file	Open URL Oper	n DB	Undo	Save
Filter				
Choose None				Apply
Current relation		Selected attribute		
Relation: None Instances: None	Attributes: None	Name: None Missing: None	Distinct: None	Type: None Unique: None
Attributes				
				Visualize All
-Status Welcome to the Weka	a Knowledge Explorer	1	C	Log x 0

00	Weka Knowledge Explorer				
	Preprocess Classify Cluster A	ssociate Select attributes Visual	ize		
Open file	Open URL Open	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 e Unique: None		
Attributes					
			Visualize All		
Status					
Welcome to the Wek	a Knowledge Explorer		Log 🔊 X 0		

\varTheta 🖯 🖯 Weka Know	ledge Explorer	
Preprocess Classify Cluster A	Associate Select attributes	Visualize
Open file Open URL Ope	en DB Undo	Save
Filter		
Choose None		Apply
Current relation	Selected attribute	
Relation: iris Instances: 150 Attributes: 5	Name: sepallength Missing: 0 (0%) Disting	Type: Numeric ct: 35 Unique: 9 (6%)
Attributes	Statistic	Value
No. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class	Maximum Mean StdDev	7.9 5.843 0.828
	Colour: class (Nom)	Visualize All
Status		Log x 0

\varTheta 🖯 🖯 Weka Know	vledge Explorer	
Preprocess Classify Cluster	Associate Select attributes Visuali	ze
Open file Open URL Op	en DB Undo	Save
Filter		
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 Mean StdDev Colour: class (Nom) Colour: class (Nom)	Visualize All
OK		Log x 0

000	Weka Knowledge Explorer				
P	reprocess Classify Cluster As	ssociate Select attributes Visu	alize		
Open file	Open URL Oper	DB Undo	Save		
Choose None			Apply		
Current relation Relation: iris Instances: 150	Attributes: 5	Selected attribute Name: class Missing: 0 (0%) Distinct: 3	Type: Nominal Unique: 0 (0%)		
Attributes No. 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class	Name	Label Iris-setosa 50 Iris-versicolor 50 Iris-virginica 50	Count		
		Colour: class (Nom)	Visualize All		
		50 50	50		
OK			Log 💉 x 0		

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



😑 🖯 🖯 Weka Knowl	ledge Explorer
Preprocess Classify Cluster A	ssociate Select attributes Visualize
Open file Open URL Ope	en DB Undo Save
Filter	
Choose None	Apply
Current relation	Selected attribute
Relation: iris Instances: 150 Attributes: 5	Name: petallengthType: NumericMissing: 0 (0%)Distinct: 43Unique: 10 (7%)
Attributes	Statistic Value
No. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class	Minimum1Maximum6.9Mean3.759StdDev1.764
	Colour: class (Nom)
Status	Log x 0

000	Weka Knowle	edge Explorer	
	Preprocess Classify Cluster As	ssociate Select attributes	Visualize
Open file	Open URL Oper	n DB Undo	Save
Filter			
Choose None			Apply
Current relation		Selected attribute	
Relation: iris Instances: 150	Attributes: 5	Name: petallength Missing: 0 (0%) Disting	Type: Numeric t: 43 Unique: 10 (7%)
Attributes		Statistic	Value
No. 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class	Name n	Maximum Mean StdDev	6.9 3.759 1.764
		Colour: class (Nom)	Visualize All
		37 11 2 0 0 3 4 1	18 17 16 14 12 2 4 3.95 6.9
Status			
UK			Log

000	Weka Knowledge Explorer							
	Preprocess	Classify	Cluster	Associate	Select att	tributes	Visualize	
Open file	Оре	n URL) (Or	en DB	\supset	Undo		Save
Filter								Apply
🛛 🔻 🧊 filters								(1997)
👖 🛛 🔻 🧊 unsupervised	đ			Selected	d attribute			
 attribute instance 				Nan Missir	ne: petallen ng: 0 (0%)	igth Distin	ct: 43	Type: Numeric Unique: 10 (7%)
[-	Statistic	c		Value
				Minimu Maximu Mean StdDev	ım um		1 6.9 3.759 1.764	
				Colou	r: class (No	om)		Visualize All
				11	2 0 0	3 4	18 17 12 3.95	16 14 2 6.9
Status							0.00	0.9
ОК							C	Log 💉 🗴 V





000	Weka Knowle	edge Explorer	
	Preprocess Classify Cluster As	ssociate Select attributes	Visualize
Open file	Open URL Oper	n DB Undo	Save
Choose Discretiz	ze -B 10 -R first-last		Apply
Current relation		Selected attribute	
Relation: iris Instances: 150	Attributes: 5	Name: petallength Missing: 0 (0%) Disting	Type: Numeric ct: 43 Unique: 10 (7%)
Attributes		Statistic	Value
No. 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class	Name	Minimum Maximum Mean StdDev	1 6.9 3.759 1.764
Status		Colour: class (Nom)	Visualize All
OK			Log x 0

\varTheta 🖯 🖯 Weka Knowle	edge Explorer
Preprocess Classify Cluster As	ssociate Select attributes Visualize
Open file Open URL Oper	n DB Undo Save
Choose Discretize -B 10 -R first-last	Apply
Current relation	Selected attribute
Relation: iris Instances: 150 Attributes: 5	Name: petallengthType: NumericMissing: 0 (0%)Distinct: 43Unique: 10 (7%)
Attributes	Statistic Value
No. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class	Minimum 1 Maximum 6.9 Mean 3.759 StdDev 1.764
	Colour: class (Nom)
Status	Log x 0

🖻 🖯 🗍 Weka Knowledge Explorer				
Preprocess Classify	y Cluster Associate Select attributes Visualize			
Open file Open URL	Open DB Undo	Save		
Filter				
Choose Discretize -B 10 -R first-last	😝 🖯 💮 weka.gui.GenericObjectEditor	Apply		
Current relation	weka.filters.unsupervised.attribute.Discretize			
Relation: iris Instances: 150 Attributes:	About An instance filter that discretizes a range of numeric More	: Numeric : 10 (7%)		
Attributes	attributes in the dataset into nominal attributes.	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 4		
	1 3.95	6.9		
OK	Log	x 0		

😑 🖯 💮 Weka Knowledge Explorer							
,	Preprocess Classif	fy Cluster Associate Select attributes Visualize					
Ор	en file Open URL	Open DB Undo	Save				
Filter							
Choose	Discretize -B 10 -R first-last	\varTheta 🖯 🕤 weka.gui.GenericObjectEditor	Apply				
Current rel	ation	weka.filters.unsupervised.attribute.Discretize					
Relatio Instance	n: iris s: 150 Attributes	About An instance filter that discretizes a range of numeric More	: Numeric : 10 (7%)				
Attributes		attributes in the dataset into nominal attributes.	e				
No.	sepallength sepalwidth	attributeIndices first-last					
3	petallength	bins 10					
4 5	petalwidth class	findNumBins False					
		invertSelection False	Visualize All				
		makeBinary False					
		useEqualFrequency False					
		Spen Save OK Cancel					
			10 2 4				
Clature		1 3.95	6.9				
OK		Log) 💉 🔨				

😑 🖯 💮 Weka Knowledge Explorer								
Preprocess Classify	Cluster Associate Select attributes Visualize							
Open file Open URL	Open DB Undo	Save						
Filter	\varTheta 🖯 🖯 weka.gui.GenericObjectEditor	Apply						
	weka.filters.unsupervised.attribute.Discretize							
Relation: iris Instances: 150 Attributes:	About An instance filter that discretizes a range of numeric More	: Numeric : 10 (7%)						
Attributes No. Name	attributes in the dataset into nominal attributes.	e						
1 sepallength 2 sepalwidth	attributeIndices first-last							
4 petalwidth 5 class	findNumBins False							
	invertSelection False	Visualize All						
	makeBinary False							
	useEqualFrequency True							
	Open Save OK Cancel							
		10 2 4						
⊂ Status	1 3.95	6.9						
ок	Log) 🔊 x 0						

Preprocess Classify Cluster Associate Select attributes Visualize Open file Open URL Open DB Undo Save Filter Choose Discretize -B 10 -R first-last Image: Comparison of the dataset into nominal attribute. Discretize Apply Current relation Relation: iris Attributes: About An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. More : Numeric : 10 (7%) Attributes No. Name attributes in the dataset into nominal attributes. More : 10 (7%) Attributes No. Name attributes in the dataset into nominal attributes. More : 10 (7%) Attributes InvertSelection False : Wisualize All MakeBinary False : Visualize All makeBinary False : Visualize All Use Equal Frequency True : 4 Open Save OK Cancel	000			Weka Knowledge	Explorer		
Open file Open URL Open DB Undo Save Filter Choose Discretize -B 10 -R first-last weka.gui.GenericObjectEditor Apply Current relation Relation: iris Attributes: About Instance filter that discretizes a range of numeric attributes. Discretize Invertice::: 10 (7%) Attributes No. Name Status Status More InvertSelection False Visualize All		Preprocess	Classify	Cluster Associa	te Select attributes Visi	ualize	
Filter Operation Apply Current relation weka.filters.unsupervised.attribute.Discretize About : Numeric Relation: iris Instances: 150 Attributes: About : Numeric Attributes About attributes in the dataset into nominal attributes. More : 10 (7%) Attributes No. Name attributelndices first-last e 1 sepallength 2 sepalwidth 3 petallength bins 10 findNumBins False visualize All MakeBinary False : invertSelection False : visualize All MakeBinary False : 0 3.95 e e Open Save OK Cancel 10 2.4 e e Status Log X Log X 0	0	pen file O	pen URL	Open DB	. Undo		Save
Choose Discretize -B 10 -R first-last Image: Webal we	Filter						
Current relation weka.filters.unsupervised.attribute.Discretize Relation: iris Attributes: Instances: 150 Attributes: Attributes An instance filter that discretizes a range of numeric attributes. More 1 sepallength 2 sepalwidth attributeIndices 3 petallength bins 4 petalwidth findNumBins 5 class invertSelection InvertSelection False invertSelection Visualize All makeBinary False 0pen Save Open Save 0pen save 1 <td>Choos</td> <td>e Discretize -B 10 -R</td> <td>first-last</td> <td>🖯 🖯 🖯 weka</td> <td>a.gui.GenericObjectEditor</td> <td></td> <td>Apply</td>	Choos	e Discretize -B 10 -R	first-last	🖯 🖯 🖯 weka	a.gui.GenericObjectEditor		Apply
Relation: iris Instances: 150 Attributes: Attributes No. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class Status OK Log Attributes	Current re	elation		weka.filters.unsuperv	vised.attribute.Discretize		
Attributes attributes in the dataset into nominal attributes. Mote No. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class invertSelection False invertSelection False visualize All makeBinary False 0pen Status OK	Relati Instanc	on: iris :es: 150	Attributes: !	About An instance filter that of	liscretizes a range of numeric	More	: Numeric : 10 (7%)
No. Name 1 sepallength attributeIndices 2 sepalwidth bins 3 petallength bins 4 petalwidth findNumBins 5 class findNumBins invertSelection False invertSelection False visualize All makeBinary False visualize All 0pen Save 0pen Save 0pen says 3.95 e.9	Attributes			attributes in the datase	t into nominal attributes.	WIDTE	e
bins 10 bins 10 findNumBins False invertSelection False useEqualFrequency True Open Save OK Log × 0	No.	N 1 sepallength	lame	attributeIndices	first-last		
A petalwidth 5 class findNumBins False invertSelection False makeBinary False UseEqualFrequency True Open Save OK Cancel 11 2 4 5 class Status OK	4	3 petallength		bins	10		
invertSelection False invertSelection False invertSelection False invertSelection False visualize All visualize All visualize All Open Open Save OK Cancel 10 2 4 10 2 4 10 2 4 10 2 4 10 2 4 10 10 2 4 10 10 10 10 10 10 10 10 10 10	-	4 petalwidth 5 class		findNumBins	False	•	
MakeBinary False useEqualFrequency True Open Save OK Cancel 11 12 4 5 5 5 5 5 5 5 5 5 5 5 6.9 5 5 5 5 5 5 5 5 5 5 5 5 5				invertSelection	False	;	Visualize All
UseEqualFrequency True Open Save OK Cancel 1 5 5 5 5 5 5 5 5 5 5 5 5 5				makeBinary	False	÷	
Open Save OK Cancel				useEqualFrequency	True	•	
Status OK				Open	Save OK	Cancel	
Status 1 3.95 6.9 OK Log					2 0 0 3 4		10 2 4
OK Log .x0	Obstar				3.95		6.9
	OK					Log	x 0

000	Weka Knowl	edge Explorer	
	Preprocess Classify Cluster A	ssociate Select attributes	Visualize
Open file	Open URL Ope	n DB Undo	Save
Filter			
Choose Discr	etize -F -B 10 -R first-last		Apply
Current relation		Selected attribute	
Relation: iris Instances: 150	Attributes: 5	Name: petallength Missing: 0 (0%) Disti	Type: Numeric nct: 43 Unique: 10 (7%)
Attributes		Statistic	Value
No. 1 sepallen 2 sepalwio 3 petallen 4 petalwio 5 class	Name ngth dth ngth dth	Minimum Maximum Mean StdDev	1 6.9 3.759 1.764
		Colour: class (Nom)	Visualize All
		37 11 2 0 0 3 4 1	18 17 16 14 12 10 2 4 3.95 6.9
OK			Log x 0

\varTheta 🖯 🖯 Weka Knowledge Explorer							
	Preprocess Classify Cluster	Associate Select attributes Visual	ize				
Open file	. Open URL Op	Den DB Undo	Save				
Filter							
Choose Disc	retize -F -B 10 -R first-last		Apply				
Current relation		Selected attribute					
Relation: iris Instances: 150	Attributes: 5	Name: petallength Missing: 0 (0%) Distinct: 43	7,/pe: Numeric / nique: 10 (7%)				
Attributes		Statistic	Value				
No. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class		Minimum 1 Maximum 6.9 Mean 3.759 StdDev 1.764	ə 1				
		Colour: class (Nom)	Visualize All				
Status							
OK			Log X V				

000	Weka Knowledge Explorer							
	Preprocess Classify Cluster As	ssociate Select attribut	es Visualize					
Open	file Open URL Oper	n DB Ur	ndo	Save				
Filter								
Choose	Discretize -F -B 10 -R first-last			Apply				
Current relati	on	Selected attribute						
Relation: Instances:	iris-weka.filters.unsupervised.attribute.Disc 150 Attributes: 5	Name: petallength Missing: 0 (0%) D	Distinct: 10	Type: Nominal Unique: 0 (0%)				
Attributes		Label		Count				
No. 1 se 2 se 3 pe 4 pe 5 cl	Name pallength epalwidth etallength etalwidth ass	'(-inf-1.45]' '(1.45-1.55]' '(1.55-1.8]' '(1.8-3.95]' '(3.95-4.35]' '(4.35-4.65]' '(4.65-5.05]'	23 14 11 13 14 15 18					
Shahua		Colour: class (Nom)	14 15	Visualize All				
OK			C	Log 💉 x 0				

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 v ector machines, multi-layer perceptrons, logistic regression, Bayes' nets, …
- "Meta"-classifiers include:
 - Bagging, boosting, stacking, error-correcting output codes, lo cally weighted learning, ...

Preprocess Classify Cluster Associate Select attributes Visualize Classifier Choose ZeroR Image: Classifier output Image: Classifier output Image: Classifier output Test options Use training set Supplied test set Set Image: Classifier output Image: Use training set Supplied test set Set Image: Classifier output Image: Classifier output Image: Cross-validation Folds 10 Image: Classifier output Image: Classifier output Image: Cross-validation Folds 10 Image: Classifier output Image: Classifier output Image: Cross-validation Folds 10 Image: Classifier output Image: Classifier output Image: Cross-validation Folds 10 Image: Classifier output Image: Classifier output Image: Cross-validation Folds 10 Image: Classifier output Image: Classifier output Image: Cross-validation Folds 10 Image: Classifier output Image: Classifier output Image: Cross-validation Folds 10 Image: Classifier output Image: Classifier output Image: Classifier output Ima	000) 🖯 🖯 Weka Knowledge Explorer						
Classifier Choose ZeroR Test options Use training set Supplied test set Set Cross-validation Folds 10 Percentage split % 66 More options (Nom) class Start Stop Result list (right-click for options)		Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Choose ZeroR Test options Classifier output Use training set Supplied test set Supplied test set Set Cross-validation Folds Percentage split % 66 More options More options (Nom) class ; Start Stop Result list (right-click for options) [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 for options) (Nom) class	Choose ZeroR							
Use training set Supplied test set Cross-validation Fercentage split % More options (Nom) class Start Start Stop Result list (right-click for options)	Test options		Clas	sifier output				
Supplied test set © Cross-validation Folds 10 Percentage split % 66 More options (Nom) class \$ Start Start Stop Result list (right-click for options)	🔘 Use training set							
Cross-validation Folds 10 Percentage split % 66 More options (Nom) class Start Stop Result list (right-click for options)	O Supplied test set	t Set						
Percentage split % 66 More options (Nom) class Start Stop Result list (right-click for options)	Cross-validation	Folds 10						
More options (Nom) class Start Stop Result list (right-click for options)	Percentage split	% 66						
(Nom) class	More op	otions						
(Nom) class								
Start Stop Result list (right-click for options)	(Nom) class		+					
Result list (right-click for options)	Start	Stop						
	Result list (right-click for	options)						

Status		
OK	Log) 🛷 × 0

000	Weka Knowledge Explorer						
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
Choose ZeroR							
Test options		Class	sifier output				
Use training set							
Supplied test set	set						
Cross-validatior	n Folds 10						
Percentage split	% 66						
More op	otions						
(Nom) class		•					
Start	Stop						
Result list (right-click for	options)						
Status							

Status		
OK	Log) 🔊 x o

000	🖯 🖯 💭 Weka Knowledge Explorer						
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
📁 weka							
🖉 🔻 🧊 classifiers							
📄 🕨 🧊 bayes			ifier output	l			
functions							
🕨 🧊 lazy							
🕨 🧊 meta							
misc							
🔻 📁 trees							
adtree							
Decision	Stump						
¥ 🔰 J48							1111
J48							1111
							11111
Random	Forest						11111
Bandom	Tree						11111
REPTree	in de						
UserCla	ssifier						11111
▶ 📁 rules							11111
			e				11111
							11111
<u> </u>		11]
Status						0	00 × 0
OK						C	.og

00	🖻 🖯 😁 Weka Knowledge Explorer						
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
Choose J48 -C	0.25 –M 2						
Test options		Clas	sifier output				
O Use training set							
Supplied test set	set						
Cross-validation	Folds 10						
Percentage split	% 66						
More op	tions						
(Nom) class		+					
Start	Stop						
Result list (right-click for	options)						

Status	
ОК	Log 💉 X O

000	🕽 🖯 🖯 Weka Knowledge Explorer			
	Preprocess Classify Cluster Associate Select attributes Visualize			
Classifier				
Choose J48 -C	2 0.25 - M 2			
Test options	Classifier output			
O Use training set	t			
Supplied test set	et Set			
Cross-validatior	on Folds 10			
Percentage split	t % 66			
More op	options			
(Nom) class				
(INOM) Class				
Start	Stop			
Result list (right-click for	or options)			
Chatura	<u>1</u>]]1			

Status	
OK	Log 💉 X O

000		Weka Knowledge Explorer		
	Preprocess	Classify Cluster	Associate Select attributes Visualize	
Classifier				
Choose J48 -C	0.25 -M 2	\varTheta 🖯 🖯 weka.gui.(GenericObjectEditor	
Test options		weka.classifiers.trees.	.j48.J48	
O Use training set		binarySplits	False 🗘	
Cross-validation	n Folds 10	confidenceFactor	0.25	
O Percentage split	% 66	minNumObj	2	
More op	otions	numFolds	3	
(Nom) class		reducedErrorPruning	False	
		saveInstanceData	False 🗘	
Result list (right-click for options)		subtreeRaising	True	
		unpruned	False 🗘	
		useLaplace	False	
		Open Save	. OK Cancel	
Status				

Status	
ОК	Log 💉 🛚

000	Weka Knowledge Explorer								
Pro	eproce	ess	Classify Cluster	Associate	Select attributes	Visualiz	ze		
Classifier									
Choose J48 -C 0.25	5 -M 2		\varTheta 🖯 🖯 weka.gui.	GenericObje	ectEditor				
Test options			weka.classifiers.trees	.j48.J48]		
O Use training set			binarySplits	False	•				
O Supplied test set	S	et	confidenceFactor	0.25					
 Cross-validation Percentage split 	olds %	10 66	minNumObj	2					
More option	ns		numFolds	3					
(Nom) class			reducedErrorPruning	False	÷				
			saveInstanceData	False	•				
Start Stop		ор	subtreeRaising	True	;				
	,		unpruned	False	•				
			useLaplace	False	;				
			Open Save	ОК	Cancel				
Status									
OK						(
00) 🖯 🖯 Weka Knowledge Explorer								
------------------------------	-------------------------------	----------	---------------	-----------	-------------------	-----------	--	--	--
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize			
Classifier	Classifier								
Choose J48 -C	0.25 –M 2								
Test options		Clas	sifier output						
O Use training set									
Supplied test set	set								
Cross-validation	Folds 10								
Percentage split	% 66								
More op	tions								
(Nom) class		+							
Start	Stop								
Result list (right-click for	options)								

Status	
ОК	Log 💉 X O

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

Status	
ОК	Log 💉 🛛

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

Status	
ОК	Log 💉 X O

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 (Nom) class Start Stop Result list (right-click for options)	00			Weka Kn	owledge Expl	lorer		
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 (Nom) class Start Stop Result list (right-click for options)		Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Choose J48 - C 0.25 - M 2	Classifier							
Test options Classifier output Use training set Supplied test set Supplied test set Set Cross-validation Folds Percentage split % More options (Nom) class Start Stop Result list (right-click for options) (right-click for options)	Choose J48 -C 0).25 -M 2						
Use training set Supplied test set Cross-validation Folds 0 Percentage split % 66 More options (Nom) class Start Stop Result list (right-click for options)	Test options		Clas	sifier output				
Supplied test set Cross-validation Folds Percentage split % 66 More options (Nom) class Start Start Stop Result list (right-click for options)	Use training set							
Cross-validation Folds 10 Percentage split % 66 More options (Nom) class Start Stop Result list (right-click for options)	Supplied test set	Set						
Percentage split % 66 More options (Nom) class Start Start Stop Result list (right-click for options)	O Cross-validation	Folds 10						
More options (Nom) class Start Stop Result list (right-click for options)	Percentage split	% 66						
(Nom) class	More op	tions						
	(Nom) class Start Result list (right-click for	Stop options)						

Status	
ОК	Log 💉 X O

\varTheta 🖯 🖯 Weka Knowledge Explorer							
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
Choose J48 -C	0.25 -M 2						
Test options		Clas	sifier output				
🔘 Use training set							
Supplied test set	Set						
O Cross-validation	Folds 10		00	🖯 Classifie	er evaluation opt	Ì	
• Percentage split	% 66		🗹 Out	put model			
More op	tions		🗹 Out	put per-clas	s stats		
(Nom) class		•	Out	put entropy	evaluation measu	res	
Start	Stop		🗹 Out	put confusio	on matrix		
Result list (right-click for	options)		🗹 Stor	e prediction	s for visualizatior	1	
			🗌 Out	put text pred	dictions on test s	et	
			Cos	t-sensitive (evaluation Set.		
			Random	seed for XV	/al / % Split 1		
					ОК	\supset	

Status	
ОК	Log 💉 X O

000	Weka Knowledge Explorer
Preproces	S Classify Cluster Associate Select attributes Visualize
Classifier	
Choose J48 -C 0.25 -M 2	
Test options	Classifier output
O Use training set	
O Supplied test set Se	t
Cross-validation Folds	0 O Classifier evaluation opt
• Percentage split % E	Output model
More options	Output per-class stats
(Nom) class	Output entropy evaluation measures
Start Sto	Output confusion matrix
Result list (right-click for options)	Store predictions for visualization
	Output text predictions on test set
	Cost-sensitive evaluation Set
	Random seed for XVal / % Split 1
	ОК
Status	
ОК	Log 💉 x 0

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

Status	
ОК	Log 💉 X O

000			Weka Kn	owledge Expl	orer		
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
Choose J48 -C	0.25 -M 2						
Test options		Clas	sifier output				
🔘 Use training set							
O Supplied test set	Set						
O Cross-validation	Folds 10						
Percentage split	% 66						
More op	tions						
(Nom) class		÷					
Start	Stop						
Result list (right-click or	options)						
	•						
1							1

Status	
ОК	Log 💉 🛛



Log 💉 🗙 Log







Log 💉 🗙 X O

00	Weka Knowledge Explorer							
	Preprocess C	lassify C	luster As	sociate Se	elect attrib	outes Visu	ualize	
Classifier								
Choose J48 -C	0.25 -M 2							
Test options		Classifier o	utput					
O Use training set		Time tak	en to build	i model: 0.2	4 seconds			6
Supplied test set	Set	=== Eval	uation on t	est split =	==			
O Cross-validation	Folds 10	=== Summa	ary ===					
Percentage split	% 66	Correctly	y Classifie	ed Instances	0.5	49	96.0784 %	
More opti	ons	Kappa sta Mean abs	atistic olute error	r r	es	0.9408	5.5210 %	
(Nom) class	\$	Relative Root rela Total Nur	absolute e ative squar mber of Ins	error red error stances		8.8979 % 33.4091 % 51		
Start	Stop	=== Deta	iled Accura	acy By Class				
-Result list (right-click for 11:49:05 – trees.j48,	options)	TP Rate 1 0.882 === Confr a b 15 0 0 19 0 2 1	FP Rate 0 0.063 0 usion Matri c < cla 0 a = In 0 b = In 5 c = In	Precision 1 0.905 1 ix === assified as ris-setosa ris-versicol ris-virginic	Recall 1 0.882 or a	F-Measure 1 0.95 0.938	Class Iris-setosa Iris-versicolor Iris-virginica	
Status							7	

OK



€ € €	Weka Knowledge Explorer		
Preprocess	Classify Cluster Associate Select a	attributes Vi	sualize
Classifier			
Choose J48 -C 0.25 -M 2			
Test options	Classifier output		
O Use training set	Time taken to build model: 0.24 seco	onds	
O Supplied test set Set	Fuelustion on test split		
O Cross-validation Folds 10	=== Summary ===		
Percentage split % 66	Correctly Classified Instances Incorrectly Classified Instances	49 2	96.0784 % 3.9216 %
More options	Mean absolute error	0.9408	
(Nom) class	Root mean squared error Relative absolute error Root relative squared error Total Number of Instances	0.1579 8.8979 33.4091 51	8
Start Stop Result list (right-click for options)	Detailed Accuracy By Class		
11:49:05 – trees.j48.J48	TP Rate FP Rate Precision Recall 1 0 1 1 1 0.063 0.905 1 0.882 0 1 0.8	all F-Measur 1 0.95 882 0.938	e Class Iris-setosa Iris-versicolor Iris-virginica
	=== Confusion Matrix ===		
	a b c < classified as 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0 2 15 c = Iris-virginica		4
Status OK			Log 💉 x (

00	Weka Knowledge Explorer						
Preprocess CI	assify Cluster Associate Sel	lect attributes Visualize					
Classifier							
Choose J48 -C 0.25 -M 2							
Test options	Classifier output						
O Use training set	Time taken to build model: 0.24	seconds					
O Supplied test set Set O Cross-validation Folds 10	=== Evaluation on test split === === Summary ===	=					
Percentage split % 66 More options	Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error	49 96.0784 % s 2 3.9216 % 0.9408 0.0396					
(Nom) class	Root mean squared error Relative absolute error Root relative squared error Total Number of Instances	0.1579 8.8979 % 33.4091 % 51	•				
Result list (right-click for options)	=== Detailed Accuracy By Class =	===					
11:49:05 - trees.j48.J48	TP Rate FP Rate Precision 1 0 1 1 0.063 0.905 0.882 0 1 === Confusion Matrix === a b c < classified as 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolos 0 2 15 c = Iris-virginica	Recall F-Measure Class 1 1 Iris-setosa 1 0.95 Iris-versicolor 0.882 0.938 Iris-virginica					
Status	5						

Log 💉 x 0

6	b	С	< classified as
15	0	0	a= Iris-setosa
0	19	0	b=Iris-versicolor
0	2	15	c=Iris-virginica

consider "TRUE" = iris-virginica and FALSE = everything else

	Ground truth				
	FALSE TRUE				
detector silent A =	=34	B = 2			
detector loud C=	= 0	D = 15			

accuracy	(A+D)/(A+B+C+D)	(34+15)/51	96%
recall (pd)	D/(B+D)	15/(2+15)	88%
false alarm (pf)	C/(A+C)	0/34	0%
precision	D/(C+D)	15/(15+0)	100%
f-measure	2*prec*pd/	2*1*0.88/	
	(prec+pd)	(1+0.88)	94%

Collect separately for each class. Repeat 10 times (re-ordering data) * 10-way Repeat for each learner * discretizer * x * y *

000	Weka Knowledge Expl	orer	
Prepro	ocess Classify Cluster Associate	Select attributes Vis	sualize
Classifier			
Choose J48 -C 0.25 -N	M 2		
Test options	Classifier output]
 Use training set Supplied test set Cross-validation Fold Percentage culit 	Set Set Set Set Time taken to build model: === Evaluation on test spli === Summary ===	0.24 seconds t ===	
More options	Correctly Classified Instan Incorrectly Classified Inst Kappa statistic Mean absolute error Root mean squared error	ices 49 ances 2 0.9408 0.0396 0.1579	96.0784 % 3.9216 %
(Nom) class	<pre>Relative absolute error Root relative squared error Total Number of Instances === Detailed Accuracy By Cl</pre>	8.8979 % 33.4091 % 51	
11:49:05 – trees.j48.J48	View in main window View in separate window Save result buffer	Recall F-Measure 1 1 1 0.95 0.882 0.938	Class Iris-setosa Iris-versicolor Iris-virginica
	Load model Save model Re-evaluate model on current test :	set	
	Visualize classifer errors Visualize tree		▲ ▼
Status OK	Visualize margin curve Visualize threshold curve Visualize cost curve	•	Log 💉 x 0





000	Weka Knowledge Explore	er	
Preprocess C	Classify Cluster Associate S	select attributes Vi	sualize
Classifier			
Choose J48 -C 0.25 -M 2			
Test options	Classifier output		
 Use training set 	Time taken to build model: 0.	24 seconds	
O Supplied test set Set O Cross-validation Folds 10	=== Evaluation on test split = === Summary ===		
Percentage split % 66 More options	Correctly Classified Instance Incorrectly Classified Instan Kappa statistic Mean absolute error	s 49 ces 2 0.9408 0.0396	96.0784 % 3.9216 %
(Nom) class	Relative absolute error Root relative squared error Total Number of Instainces === Detailed Accuracy by Class	8.8979 33.4091 51	s s
11:49:05 - trees.j48.J48 View in Save res	main window separate window ult buffer	Recall F-Measure 1 1 1 0.95 0.882 0.938	e Class Iris-setosa Iris-versicolor Iris-virginica
Load mo Save mo Re-eval	odel odel uate model on current test set		
Visualiz Status Visualiz OK Visualiz Visualiz Visualiz	e classifer errors e tree e margin curve e threshold curve e cost curve	lor ca	Log x 0

00	Weka Knowledge Explorer							
	Preprocess C	lassify C	luster As	sociate Se	elect attrib	outes Visu	ualize	
Classifier								
Choose J48 -C	0.25 -M 2							
Test options		Classifier o	utput					
O Use training set		Time tak	en to build	i model: 0.2	4 seconds			6
Supplied test set	Set	=== Eval	uation on t	est split =	==			
O Cross-validation	Folds 10	=== Summa	ary ===					
Percentage split	% 66	Correctly	y Classifie	ed Instances	0.5	49	96.0784 %	
More opti	ons	Kappa sta Mean abs	atistic olute error	r r	es	0.9408	5.5210 %	
(Nom) class	\$	Relative Root rela Total Nur	absolute e ative squar mber of Ins	error red error stances		8.8979 % 33.4091 % 51		
Start	Stop	=== Deta	iled Accura	acy By Class				
-Result list (right-click for 11:49:05 – trees.j48,	options)	TP Rate 1 0.882 === Confr a b 15 0 0 19 0 2 1	FP Rate 0 0.063 0 usion Matri c < cla 0 a = In 0 b = In 5 c = In	Precision 1 0.905 1 ix === assified as ris-setosa ris-versicol ris-virginic	Recall 1 0.882 or a	F-Measure 1 0.95 0.938	Class Iris-setosa Iris-versicolor Iris-virginica	
Status							7	

OK



000	Weka Knowledge Explorer									
	Preprocess Cl	assify Clu	uster As	sociate Se	elect attrib	outes Visu	alize			
Classifier										
Choose J48 -C	0.25 -M 2									
Test options		Classifier ou	tput							
O Use training set		Time taken to build model: 0.24 seconds								
 Supplied test set Cross-validation 	Folds 10	=== Evaluation on test split === === Summary ===								
Percentage split More opti	% 66 ions	Correctly Incorrect Kappa star Mean abso	Classifie ly Classif tistic lute error	lassified Instances Classified Instances stic te error			96.0784 % 3.9216 %			
(Nom) class	Stop	Root mean Relative a Root relat Total Num	squared e absolute e tive squar ber of Ins	error error ed error stances		8.8979 % 33.4091 % 51				
Result list (right-click for	options)	=== Detai	led Accura	cy By Class						
11:49:05 - trees.j48.	J48	TP Rate 1 0.882 === Confus a b c 15 0 0 0 19 0 0 2 15	FP Rate 0 0.063 0 sion Matri < cla a = Ir b = Ir c = Ir	Precision 1 0.905 1 .x === assified as ris-setosa ris-versicolo ris-virginica	Recall 1 0.882	F-Measure 1 0.95 0.938	Class Iris-setosa Iris-versicolor Iris-virginica			
Status							7			

Log 💉 X 0

000			Weka Kno	Weka Knowledge Explorer						
	Preprocess	Classify	Cluster	Associate	Select attrib	outes Visu	alize			
Classifier										
 weka classifiers p bayes functions 			·output							
LeastMedSo LinearRegr	iken to bu iluation o mary ===	ild model: (on test split).24 seconds							
 neural Neural pace supportVet SimpleLine SimpleLogi VotedPerce Winnow 	Network ctor arRegression stic eptron		ly Classi ctly Class statistic solute er an square re absolut lative sq Jumber of ailed Acc	fied Instand sified Insta or d error e error quared error Instances suracy By Cla	nces	49 2 0.9408 0.0396 0.1579 8.8979 % 33.4091 % 51	96.0784 % 3.9216 %			
 Withow Iazy meta misc trees rules 		15 0 0 19 0 2	<pre>FP Rat 0 0.06 0 ifusion Ma c < 0 a = 0 b = 15 c =</pre>	e Precisio 1 3 0.905 1 trix === classified a Iris-setosa Iris-versio Iris-virgin	on Recall 1 5 1 0.882 as color nica	F-Measure 1 0.95 0.938	Class Iris-setosa Iris-versicolor Iris-virginica			
<u>r</u>		7_6								

Status	
ОК	Log 💉 X O

000		Weka Knowledge Explorer								
	Preprocess CI	assify C	luster As	sociate Se	lect attrib	outes Visu	alize			
Classifier										
Choose Neural	Network -L 0.3 -M	1 0.2 -N 50	0 -V 0 -S 0	-E 20 -H a						
Test options		Classifier o	utput							
🔘 Use training set								0		
O Supplied test set	Set	=== Eval === Summ	=== Evaluation on test split === === Summary ===							
O Cross-validation	Folds 10	Correctl	y Classifie	d Instances		49	96.0784 %			
Percentage split	% 66	Incorrec Kappa st	tly Classif atistic	ied Instance	es	2	3.9216 %			
More opti	ons	Mean abs	olute error	:		0.0396				
		Root mean Relative	n squared e absolute e	rror		0.1579 8.8979 %				
(Nom) class	\$	Root rel Total Nu	ative squar mber of Ins	ed error		33.4091 % 51				
Start	Stop	=== Deta	iled Accura	acy By Class						
Result list (right-click for	options)	TP Rate	FP Rate	Precision	Recall	F-Measure	Class			
11:49:05 - trees.j48.	J48	1	0.063	0.905	1	0.95	Iris-versicolor			
		0.882	0	1	0.882	0.938	Iris-virginica			
		=== Conf	usion Matri	x ===						
		a b	c < cla	ssified as						
		15 0	0 a = I1	is-setosa						
		0 19	0 b = 1 5 c = 1	is-versicolo)r 1					
								*		
								Y		
Status							7			

Log 💉 X 0

OK

000	Weka Knowledge Explorer								
Preprocess C	assify Cluster Associate Sele	ect attributes Visualia	ze						
Classifier									
Choose NeuralNetwork -L 0.3 -N	10.2 -N 500 -V 0 -S 0 -E 20 -H a								
Test options	Classifier output								
O Use training set									
O Supplied test set Set	=== Evaluation on test split === === Summary ===	=== Evaluation on test split === === Summary ===							
Cross-validation Folds 10 Percentage split % 66 More options	Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error Relative absolute error Root relative squared error	49 2 0.9408 0.0396 0.1579 8.8979 % 33.4091 %	96.0784 % 3.9216 %						
(Nom) class	Total Number of Instances === Detailed Accuracy By Class =	51							
Result list (right-click to options)	TP Rate FP Rate Precision 1 1 0 1 1 0.063 0.905 0.882 0 1 === Confusion Matrix === a b c < classified as 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0 2 15 c = Iris-virginica	Recall F-Measure C 1 1 I 1 0.95 I 0.882 0.938 I	lass ris-setosa ris-versicolor ris-virginica						
Status									

OK

Log 💉 🗴 0





000	Weka Knowledge Explorer								
	Preprocess Classify	Cluster Associate Select attributes	Visualize						
Classifier									
Choose Neural	Network -L 0.3 -M 0.2 -	N 500 - V 0 - S 0 - E 20 - H a - G - R							
Test options		Classifier output							
O Use training set		=== Evaluation on test split ===		6					
Supplied test set	t Set	Summary							
O Cross-validation	Folds 10	Correctly Classified Instances Incorrectly Classified Instances	50 98.0392 % 1 1.9608 %						
Percentage split	% 66	Kappa statistic Mean absolute error	0.9704 0.0239						
More o	options	Root mean squared error	0.1101						
		Root relative squared error	23.2952 %						
(Nom) class	•	Total Number of instances	51						
	(Chan	=== Detailed Accuracy By Class ===							
Start	Stop	TP Rate FP Rate Precision Recall 1 0 1 1	F-Measure Class 1 Iris-setosa						
Result list (right-click for	roptions)	1 0.031 0.95 1	0.974 Iris-versicolo	r					
14:34:28 - functions	s.neural.NeuralNetwork	0.941 0 1 0.941	0.57 III3-VIIginida						
		=== Confusion Matrix ===							
		a b c < classified as 15 0 0 a = Iris-setosa							
		0 19 0 b = Iris-versicolor 0 1 16 c = Iris-virginica							
				¥					
		())++	•					
Status									
ОК			Log 💞	x 0					

00	Weka Knowledge Explorer								
	Preprocess	Classify	Cluster	Associat	e Select at	ttributes	Visualize		
Classifier									
Choose Neural	Network -L	0.3 -M 0.2 -	N 500 -V 0	-S 0 -E 20	-H a -G -R				
Test options			Classifier ou	tput					
🔘 Use training set			=== Evalu	ation on t	est split =				
O Supplied test set	t Se	t)	=== Summa	ry ===					
Cross-validation	Folds 10		Correctly Incorrect	Classifie lv Classif	ed Instances lied Instance	25	50 1	98.0392 % 1.9608 %	
Percentage split	% 66		Kappa sta	tistic			0.9704		
Mara	ntions		Root mean	squared e	error		0.1101		
More c	options		Relative absolute error				5.3594 %		
(Nom) class		+	Total Num	ber of Ins	stances		51		
			=== Detai	led Accura	ncy By Class	===			
Start) St	op	TP Rate	FP Rate	Precision	Recall	F-Measure	Class	
Result list (right-click for	options)		1	0	1	1	1	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.Neu	ralNetwork	=== Confu	sion Matri	x ===				
			a b c 15 0 0 0 19 0 0 1 16	< cla a = In b = In c = In	assified as tis-setosa tis-versicolo tis-virginica	or a			
				* * * *	****) 4 + (
Status									

OK Log X 0

	Weka Kn	owledge Exp	olorer			
Preprocess Classify	Cluster	Associate	Select at	tributes	Visualize	
Classifier						
weka						
🔽 🔽 📁 bayes	assifier out	tput				
AODE	== Evalua	ation on te	st split ==			0
BayesNetK2	== Summar	су ===				
BayesNetB NaiveBayes NaiveBayesMultinomial NaiveBayesSimple NaiveBayesUpdateable functions I azy meta	prrectly appa stat an absol pot mean alative a pot relat ptal Numb == Detail	Classified ly Classified tistic lute error squared er absolute er tive square ber of Inst	Instances ed Instance ror d error ances y By Class	===	50 1 0.9704 0.0239 0.1101 5.3594 % 23.2952 % 51	98.0392 % 1.9608 %
Trees rules	? Rate 1 1 0.941	FP Rate 1 0 0.031 0	Precision 1 0.95 1	Recall 1 1 0.941	F-Measure 1 0.974 0.97	Class Iris-setosa Iris-versicolor Iris-virginica
	== Confus a b c 15 0 0 0 19 0 0 1 16	<pre>sion Matrix < class a = Iris b = Iris c = Iris </pre>	=== sified as s-setosa s-versicolo s-virginica	or a		
1						

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



000	Weka Knowledge Explorer								
	Preprocess	Classify	Cluster	Associat	e Select at	ttributes	Visualize		
Classifier									
Choose NaiveBa	ayes								
Test options			Classifier out	tput					
🔘 Use training set			=== Evalua	ation on t	est split ==			0	
Supplied test set	t Set	.)	=== Summan	cy ====					
Cross-validation	Folds 10		Correctly	Classifie	d Instances	99	50	98.0392 %	
Percentage split	% 66		Kappa stat	istic	100 1100010		0.9704	1.0000 0	
C			Mean absol Root mean	lute error squared e	rror		0.0239		
More o	options		Relative a	absolute e	rror		5.3594 %		
			Total Numb	per of Ins	ed error tances		23.2952 % 51		
(Nom) class		•	=== Detail	ed Accura	cv By Class				
Start	Stor								
Dentriticitation			TP Rate	FP Rate 0	Precision 1	Recall 1	F-Measure 1	Class Iris-setosa	
Result list (right-click for	roptions)		1	0.031	0.95	1	0.974	Iris-versicolor	
11:49:05 - trees.j48	.J48 . noural Noural	Notwork	0.941	U	1	0.941	0.97	Iris-virginica	
14:54:28 - Tunctions	s.neurai.neurai	Network	=== Confus	sion Matri	x ===				
			a b c	< cla	ssified as				
			15 0 0	a = Ir	is-setosa				
			0 1 16	c = Ir	is-virginica	а			
								*	
			_	-				7	
1		11							

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



Preprocess Classify Cluster Associate Select attributes Visualize Classifier Choose NaiveBayes Image: Classifier output Image:
Classifier Choose NaiveBayes Test options Classifier output Use training set Classifier output Supplied test set Set Cross-validation Folds 10 Percentage split % 66 More options Correctly Classified Instances 50 98.0392 % Root mean squared error 0.0239 0.0239 Root mean squared error 0.1101 Relative absolute error 0.3594 % Root relative squared error 23.2952 %
Choose NaiveBayes Test options Ouse training set Use training set Supplied test set Supplied test set Set Cross-validation Folds 10 Percentage split % 66 More options Correctly Classified Instances More options 0.0239 Root mean squared error 0.1101 Relative absolute error 5.3594 % Root relative squared error 23.2952 %
Classifier output O Use training set O Supplied test set Set O Cross-validation Folds 10 Percentage split % 66 More options 66 More options Correctly Classified Instances Supplied test set Set Correctly Classified Instances 1 1.9608 % 1.9608 % More options 66
O Use training set === Evaluation on test split === O Supplied test set Set O Cross-validation Folds 10 Incorrectly Classified Instances 50 98.0392 % Incorrectly Classified Instances 1 1.9608 % More options More options 0.0239 Root mean squared error 0.1101 Relative absolute error 5.3594 % Root relative squared error 23.2952 %
Supplied test set Set Cross-validation Folds 10 Percentage split % 66 More options Correctly Classified Instances 1 More options 0.9704 Relative absolute error 0.1101 Relative absolute error 5.3594 % Root relative squared error 23.2952 %
O Cross-validation Folds 10 O Cross-validation Folds 10 O Percentage split % More options 66 More options 0.9704 Relative absolute error 0.101 Relative absolute error 5.3594 % Root relative squared error 23.2952 %
• Percentage split * 66 Kappa statistic 0.9704 Mean absolute error 0.0239 Root mean squared error 0.1101 Relative absolute error S.3594 % Root relative squared error 23.2952 %
(Nom) class
=== Detailed Accuracy By Class ===
StartStopResult list (right-click for options)TP RateFP RatePrecisionRecallF-MeasureClass101111Iris-setosa11:49:05 - trees.j48.J48010.9510.9410.974Iris-versicolor
14:34:28 - functions.neural.leuralNetwork === Confusion Matrix ===
a b c < classified as 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0 1 16 c = Iris-virginica

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



000	Weka Knowledge Explorer								
	Preprocess	Classify	Cluster	Associate	Select at	tributes	Visualize		
Classifier									
Choose NaiveBa	ayes								
Test options			Classifier ou	tput					
🔘 Use training set			=== Evalua	ation on to	est split ==			6	
Supplied test set	t Set		=== Summa:	cy ===					
	Eolds 10		Correctly	Classifie	d Instances		48	94.1176 %	
Percentage split			Kappa stat	tistic	led instance	25	0.9113	5.8824 %	
• Percentage spin	% 00		Mean absol	lute error			0.0447		
More o	options		Relative a	absolute e	rror		10.0365 %		
			Root relat	tive square	ed error		36.4196 %		
(Nom) class		\$	rocar Num	Jer of 1113	cances		51		
			=== Detail	led Accura	cy By Class				
Start) Sto	p)	TP Rate	FP Rate	Precision	Recall	F-Measure	Class	
Result list (right-click for	r options) ——		1	0 063	1	1	1	Iris-setosa	
11:49:05 - trees.j48	.J48		0.882	0.029	0.938	0.882	0.909	Iris-virginica	
14:34:28 - functions	s.neural.Neura	alNetwork	=== Confu	sion Matri	× ===				
14:48:05 - bayes.Na	iveBayes		001120						
			a b c 15 0 0	< cla:	ssified as is-setosa				
			0 18 1	b = Ir	is-versicolo	or			
			0 2 15	c = Ir:	is-virginica	3		Ų	
								-	
			C) 4 1 4	
Statue									

OK Log 💉 X 0

00	Weka Knowledge Explorer								
	Preprocess	Classify	Cluster	Associat	e Select at	tributes	Visualize		
Classifier									
Choose NaiveBa	ayes								
Test options			Classifier ou	tput					
🔘 Use training set			=== Evalu	ation on t	est split ==			-	
Supplied test set	se Se	et)	=== Summa	ry ===					
O Cross-validation	Folds 10)	Correctly Incorrect	Classifie ly Classif	d Instances ied Instance	es	48 94.11 3 5.88	94.1176 % 5.8824 %	
Percentage split	% 66	õ	Kappa statistic 0.9113 Mean absolute error 0.0447						
More o	options		Root mean squared error Relative absolute error Root relative squared error				0.1722 10.0365 % 36.4196 %		
(Nom) class		•	Total Num	per of Ins	tances		51		
Start	Ct	00	=== Detai	led Accura	CY BY CLASS				
Result list (right-click for	options)	op	TP Rate 1 0.947	FP Rate 0 0.063	Precision 1 0.9	Recall 1 0.947	F-Measure 1 0.923	Class 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 - functions 14:48:05 - bayes.Na	iveBayes	rainetwork	=== Confu	sion Matri	x ===				
			a b c 15 0 0 0 18 1 0 2 15	< cla a = Ir b = Ir c = Ir	ssified as is-setosa is-versicolo is-virginica	or			
1				****	****) 4 Þ (
Status									

OK Log 💉 X 0

🖯 🖯 🖯 Weka Knowledge Explorer								
	Preprocess	Classify	Cluster	Associate	Select at	tributes	Visualize	
Classifier								
Choose NaiveBa	yes							
Test options			Classifier outp	ut				
🔘 Use training set			=== Evaluat	ion on tes	t split ==	=		-
O Supplied test set	Set.		=== Summary	a ===	Taskanasa		40	04 1176 8
O Cross-validation	Folds 10		Incorrectly C	correctly Classified Instances				5.8824 %
Percentage split	% 66		Kappa stati	stic			0.9113	
More o	ptions	•	Root mean s Relative ab Root relati Total Numbe	quared erro solute erro ve squared er of Instan	or or error nces By Class		0.1722 10.0365 % 36.4196 % 51	
Start	View in ma	ain window	v	P	recision	Recall	F-Measure	Class
Result list (right-click for	View in se	parate win	ndow		1 1	1	1	Iris-setosa
11:49:05 - trees.j48.J	Save resul	t buffer			0.938	0.882	0.909	Iris-virginica
14:34:28 - functions. 14:48:05 - bayes.Nai	Load model Save model Re-evaluate model on current test set							
					setosa	~		
	Visualize o	lassifer er	rors	:is	-virginica	-		
	Visualize t	ree						A
Visualize margin cu			rve			_		Y A
1	Visualize t	hreshold o	curve		ris-setosa			111
OK	Visualize o	cost curve			ris-versico ris-virginio	ior :a	C	Log 💉 🔊

🖯 🖯 Weka Knowledge Explorer										
	Preprocess	Classify	Cluster	Associate	Select at	tributes	Visualize			
Classifier										
Choose NaiveBa	ayes									
Test options			Classifier ou	tput						
🔘 Use training set			=== Evalua	ation on to	est split ==			6		
Supplied test set	t Set		=== Summa:	cy ===						
	Eolds 10		Correctly	Classifie	d Instances		48	94.1176 %		
Percentage split	* 66		Kappa stat	tistic	led instance	25	0.9113	5.8824 %		
• Percentage spin	% 00		Mean absol	lute error			0.0447			
More o	options		Relative a	absolute e	rror		10.0365 %			
			Root relat	tive square	ed error		36.4196 %			
(Nom) class		\$	rocar Num	Jer of 1113	cances		51			
			=== Detail	led Accura	cy By Class					
Start) (Sto	p)	TP Rate	FP Rate	Precision	Recall	F-Measure	Class		
Result list (right-click for	r options) ——		1	0 063	1	1	1	Iris-setosa		
11:49:05 - trees.j48	.J48		0.882	0.029	0.938	0.882	0.909	Iris-virginica		
14:34:28 - functions	s.neural.Neura	alNetwork	=== Confu	sion Matri	× ===					
14:48:05 - bayes.Na	iveBayes		001120							
			a b c 15 0 0	< cla:	ssified as					
			0 18 1	b = Ir	is-versicolo	or				
			0 2 15	c = Ir:	is-virginica	3		Ų		
								-		
			C) 4 + (
Statue										

OK Log 💉 X 0

000) 🖯 Weka Knowledge Explorer									
	Preprocess	Classify	Cluster	Associat	e Select at	tributes	Visualize			
Classifier										
Choose NaiveBa	ayes									
Test options			Classifier out	tput						
 Use training set Supplied test set 	t Set		=== Evalua === Summan	ation on t	est split ==			<u> </u>		
Cross-validation Percentage split More of (Nom) class	n Folds 10 % 66 options	•	Correctly Incorrectl Kappa stat Mean absol Root mean Relative a Root relat Total Numb	Classifie Ly Classif Listic Lute error squared e absolute e Live squar per of Ins	d Instances ied Instance rror ed error tances cy By Class	===	48 3 0.9113 0.0447 0.1722 10.0365 % 36.4196 % 51	94.1176 % 5.8824 %		
Start Result list (right-click for 11:49:05 - trees.j48 14:34:28 - functions 14:48:05 - bayes.Na	Stop roptions) J48 s.neural.Neural iveBayes	Network	TP Rate 1 0.947 0.882 === Confus a b c 15 0 0 0 18 1 0 2 15	FP Rate 0 0.063 0.029 sion Matri < cla a = Ir b = Ir c = Ir	Precision 1 0.9 0.938 x === 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		
Status										

Log 💉 x 0

😑 🖯 🕤 Weka Knowledge Explorer									
	Preprocess	Classify	Cluster	Associat	e Select at	tributes	Visualize		
Classifier									
Questioners Questioners and a TMF (L2W) deco	ompressor are needed to see the								
Test options			Classifier ou	tput					
🔘 Use training set			=== Evalu	ation on t	est split ==			6	
Supplied test set	t Set.		=== Summa	ry ===					
Cross-validation	Folds 10		Correctly	Classifie	d Instances		48	94.1176 %	
 Percentage split 	% 66		Kappa sta	tistic	ied instance	28	0.9113	5.0024 %	
More o		Root mean Relative Root rela	squared e squared e absolute e tive squar	rror rror ed error	0.0447 0.1722 10.0365 % 36.4196 %				
(Nom) class		÷	Total Num	per of Ins	tances		51		
			=== Detai	led Accura	CY BY Class	===			
- Result list (right click for	Sto	•	TP Rate	FP Rate 0	Precision 1	Recall 1	F-Measure 1	Class Iris-setosa	
11:40:05 trees ide			0.947	0.063	0.9	0.947	0.923	Iris-versicolor	
14:34:28 - functions	s.neural. Veura	INetwork	=== Confu	sion Matri	x ===	0.001	0.000		
14.40.05 Dayes.ru	ivebuyes 🗸		a b c 15 0 0 0 18 1 0 2 15	< cla a = Ir b = Ir c = Ir	ssified as is-setosa is-versicolo is-virginica	or a			
				++++	++++)∢►	
Status									

OK Log CK
Explorer: clustering data

- WEKA contains "clusterers" for finding groups of similar instances in a dataset
- Implemented schemes are:
 - *k*-Means, EM, Cobweb, *X*-means, FarthestFirst
- Clusters can be visualized and compared to "t rue" clusters (if given)
- Evaluation based on loglikelihood if clustering scheme produces a probability distribution

🖯 🖯 💮 Weka Knowledge Explorer								
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize		
Clusterer								
Choose EM -I 1	Choose EM -I 100 -N -1 -S 100 -M 1.0E-6							
Cluster mode			Clust	erer output				
🖲 Use training set								
Supplied test set	t Se	t						
O Percentage split		% 66						
Classes to cluste	ers evaluation							
(Nom) class		A V						
Store clusters fo	or visualization							
Igno	ore attributes		\supset					
Start		Stop	\supset					
Result list (right-click for	roptions)							
р <u>г</u>			110					

Status	
OK	Log 💉 X O

🖯 🖯 🖯 Weka Knowledge Explorer						
	Preprocess Classify	Cluster	Associate	Select attributes	Visualize	
Clusterer						
Choose EM -I 1	00 -N -1 -S 100 -M 1.0E-	6				
Cluster mode		Cluster	er output			
💿 Use training set						
Supplied test set	set					
Percentage split	% 66					
Classes to cluste	ers evaluation					
(Nom) class	4 7					
Store clusters fo	r visualization					
Igno	ore attributes	\supset				
Start	Stop	\supset				
Result list (right-click for	options)					
		15				42

Status	
ОК	Log 💉 🛛

000	Weka Knowledge Explorer	r Confere
Preprocess Classify	Cluster Associate Select attributes Visualize	
Clusterer		
Vusterer weka Clusterers EM SimpleKMeans Cobweb FarthestFirst XMeans	77387815	
Status		

ОК



000			Weka Kn	owledge Expl	orer		
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Clusterer							
Choose Cobwel	b -A 1.0 -C 0.0	02820947	917738781	5			
Cluster mode			Cluste	erer output			
🖲 Use training set							
O Supplied test set	t Se	t					
O Percentage split		% 66					
Classes to clust	ers evaluation						
(Nom) class		A Y					
Store clusters fo	or visualization						
Igno	ore attributes		\supset				
Start		Stop	\supset				
Result list (right-click for	r options)						
Status							
OK						Log) ×0

1 100

000			Weka Kn	owledge Expl	orer		
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Clusterer							
Choose Cobwel	b -A 1.0 -C 0.0	02820947	917738781	5			
Cluster mode			Cluste	erer output			
🖲 Use training set							
Supplied test set	t Se	t					
O Percentage split		% 66					
Classes to clust	ers evaluation						
(Nom) class		Å ¥					
Store cousters fo	or visualization						
land	ore attributes						
Igin	ore attributes						
Start)	Stop					
Result list (right-click for	r options)						
Status						<i>C</i> .	
OK						Log) X0









1	Status	
	OK	Log 💉 x 0



Status	
OK	Log 💉 X O











Explorer: finding associations

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

000			Weka Kn	owledge Exp	olorer				
	Preprocess	Classify	Cluster	Associate	Select at	tributes	Visual	ize	
Associator									
Choose Apriori	-N 10 -T 0 -C 0	.9 -D 0.05	5 -U 1.0 -	M 0.1 -S -1.	0				
	(the second	Asso	ciator outp	ut					
Start	Stop								
Result list (right-click for	r options)								
Status		110							1
ОК								Log) 💉 🥎

Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori - N 10 - T 0 - C 0.9 - P.0.05 - U 1.0 - M 0.1 - S - 1.0 Associator output Start Stop Associator output Associator output Result list (right-click for options) Associator output Image: Classify output Image: Classify output Status OK Image: Classify output Image: Classify output Image: Classify output	000		Weka K	nowledge Explo	orer			
Associator Choose Apriori - N 10 - T 0 - C 0.9 = 10.05 - U 1.0 - M 0.1 - S - 1.0 Start Stop Result list (right-click for options) Status OK Log x 0		Preprocess	Classify Cluster	Associate	Select attributes	Visualize	2	
Choose Apriori - N 10 - T 0 - C 0.9 - 1 0.05 - U 1.0 - M 0.1 - S - 1.0 Start Stop Result list (right-click for options) Result list (right-click for options) Status OK Log	Associator							
Start Stop Associator output	Choose Ap	riori – N 10 –T 0 –C 0	.9 -1 0.05 -U 1.0	-M 0.1 -S -1.0				
Start Stop Result list (right-click for options) Status OK Log			Associator out	out				
Status OK	Start	Stop						
	Result list (right-cl	ick for options)						
Status OK								
Status OK								
Status OK								
Status OK								
Status OK								
Status OK								
Status OK								
Status OK								
Status OK								
Status OK								
Status OK								
Status Log Log x 0								
OK Log x0	-Status		431					
	OK					(Log	× 0

\varTheta 🖯 🖯 Weka Knowle	edge Explorer
Preprocess Classify Cluster As	ssociate Select attributes Visualize
Open file Open URL Open	n DB Undo Save
Choose None	Apply
Current relation	Selected attribute
Relation: vote Instances: 435 Attributes: 17	Name: handicapped-infantsType: NominalMissing: 12 (3%)Distinct: 2Unique: 0 (0%)
Attributes	Label Count
No.Name1handicapped-infants2water-project-cost-sharing3adoption-of-the-budget-resolution4physician-fee-freeze5el-salvador-aid6religious-groups-in-schools7anti-satellite-test-ban8aid-to-nicaraguan-contras9mx-missile10immigration11synfuels-corporation-cutback12education-spending13superfund-right-to-sue14crime15duty-free-exports16export-administration-act-south-africa17Class	y 187 Colour: Class (Nom) Visualize All



\varTheta 🖯 🖯 Weka Knowl	ledge Explorer
Preprocess Classify Cluster A	ssociate Select attributes Visualize
Open file Open URL Ope	en DB Undo Save
Choose None	Apply
Current relation Relation: vote Instances: 435 Attributes: 17	Selected attribute Name: handicapped-infants Missing: 12 (3%) Distinct: 2 Unique: 0 (0%)
Attributes 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 11 synfuels-corporation-cutback 12 education-spending 13 superfund-right-to-sue 14 crime 15 duty-free-exports 16 export-administration-act-south-africa 17 Class 17 class	Label Count n 236 y 187 Colour: Class (Nom) Visualize All

OK Log X 0

000			Weka Kn	owledge Exp	olorer				
	Preprocess	Classify	Cluster	Associate	Select at	tributes	Visual	ize	
Associator									
Choose Apriori	-N 10 -T 0 -C 0	.9 -D 0.05	5 -U 1.0 -	M 0.1 -S -1.	0				
	(the second	Asso	ciator outp	ut					
Start	Stop								
Result list (right-click for	r options)								
Status		110							1
ОК								Log) 💉 🥎

000		Weka Knowledge Exp	orer	
	Preprocess Classify	Cluster Associate	Select attributes V	isualize
Associator				
Choose Apriori	-N 10 -T 0 -C 0.9 -D 0.0	5 -U 1.0 -M 0.1 -S -1.0)	
Start	Stop	ociator output		
Result list (right-click for	options)			
Status				
ОК				Log 💉 X O

000	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	
Start Stop	Associator output	
Result list (right-click for optic	Minimum metric <confidence>: 0.9 Number of cycles performed: 11</confidence>	0
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 Best rules found:	
	1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democr 2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-co 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 con 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 9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197 cor 10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210	:at int : (1 if: 2(if: •
Status		

ОК

Log 💉 X 0

Explorer: attribute selection

- Panel that can be used to investigate which (s ubsets of) attributes are the most predictive o nes
- Attribute selection methods contain two parts:
 - A search method: best-first, forward selection, random, exha ustive, genetic algorithm, ranking
 - An evaluation method: correlation-based, wrapper, informati on gain, chi-squared, ...
- Very flexible: WEKA allows (almost) arbitrary combinations of these two

00		Weka Knowledge Explorer
	Preprocess Class	ify Cluster Associate Select attributes Visualize
Attribute Evaluator		
Choose CfsSubs	setEval	
Search Method		
Choose BestFirs	t −D 1 −N 5	
Attribute Selection Mode	9	Attribute selection output
💿 Use full training	set	
Cross-validation	Folds 10	
	Seed 1	
(Nom) Class	•	
Start	Stop	
Result list (right-click for	options)	
1		
Status		

Log 💉 X O

$\Theta \Theta \Theta$		Weka Knowledge Explorer
	Preprocess Class	ify Cluster Associate Select attributes Visualize
Attribute Evaluator		
Choose CfsSub	setEval	
Search Method		
Choose BestFirs	st -D 1 -N 5	
Attribute Selection Mode	9	Attribute selection output
🖲 Use full training	set	
Cross-validation	Folds 10	
	Seed 1	
(Nom) Class	•	
Start	Stop	
Result list (right-click for	options)	
	•	
Status		

Status	
OK	Log 💉 X O

000		Weka Knowledge Explorer
	Preprocess Clas	sify Cluster Associate Select attributes Visualize
Attribute Evaluator		
Choose CfsSub	setEval	
Search Method		
Choose BestFirs	st -D 1 -N 5	
Attribute Selection Mod	e	Attribute selection output
 Use full training Cross-validation 	set Folds 10 Seed 1	duty-free-exports export-administration-act-south-africa Class Evaluation mode: evaluate on all training data
(Nom) Class	Stop	=== Attribute Selection on all input data === Search Method:
Result list (right-click for 16:39:40 – BestFirst	+ CfsSubsetEval	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 Attribute Subset Evaluator (supervised, Class (nominal): 17 Class): CFS Subset Evaluator Selected attributes: 4 : 1 physician-fee-freeze

Status	
ОК	Log x 0



j	Status	1
	OK	Log x 0

000	Weka Knowledge Explorer
Preprocess Classify	Cluster Associate Select attributes Visualize
Attribute Evaluator	
weka weka CfsSubsetEval ClassifierSubsetEval WrapperSubsetEval	
A ConsistencySubsetEval	te selection output
ReliefFAttributeEval InfoGainAttributeEval GainRatioAttributeEval SymmetricalUncertAttributeEval OneRAttributeEval ChiSquaredAttributeEval PrincipalComponents	duty-free-exports export-administration-act-south-africa Class mation mode: evaluate on all training data Attribute Selection on all input data ===
F I SVMAttributeEval	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 ected attributes: 4 : 1 physician-fee-freeze

1	Status	1
	OK	Log 💉 X O

000		Weka K	nowledge Expl	orer		
	Preprocess	Classify Cluster	Associate	Select attributes	Visualize	
Attribute Evaluator						
Choose InfoGa	inAttributeEva	al				
Search Method						
🧊 weka	on	308 -N	-1			
BestFirst		te selecti	on output			
ForwardSelection RaceSearch GeneticSearch RandomSearch ExhaustiveSearch Ranker RankSearch		Attribut Attribut Th Metho Best Stat Stat Stat Tota Mer: ibute St CFS	duty-free- export-adm Class mode: evalu- te Selection of od: t first. rt set: no at rch direction le search aft al number of a it of best sub ubset Evaluate Subset Evaluate tributes: 4 : phys	exports inistration-act-so uate on all traini on all input data tributes : forward er 5 node expansio subsets evaluated: bset found: 0.7 or (supervised, Cl ator 1 sician-fee-freeze	uth-africa ng data === ns 83 29 ass (nominal): 17 C	lass):
Status						

Log 💉 X O





000	Weka Knowledge Explorer
Preproce	ess Classify Cluster Associate Select attributes Visualize
Attribute Evaluator	
Choose InfoGainAttribute	eEval
Search Method	
Choose Ranker -T -1.797	76931348623157E308 -N -1
Attribute Selection Mode	Attribute selection output
💿 Use full training set	Information Gain Ranking Filter
Cross-validation Fold	s 10 d 1 Ranked attributes: 0.7078541 4 physician-fee-freeze 0.4185726 3 adoption-of-the-budget-resolution 0.4028397 5 el-salvador-aid
(Nom) Class	0.34036 12 education-spending 0.3123121 14 crime 0.3095576 8 aid-to-nicaraguan-contras 0.2856444 9 mx-missile
Result list (right-click for options)	0.2121705 13 superfund-right-to-sue 0.2013666 15 duty-free-exports
16:39:40 - BestFirst + CfsSubs 16:43:05 - Ranker + InfoGainA	<pre>setEval ().1902427 7 anti-satellite-test-ban ().1404643 6 religious-groups-in-schools ().1211834 1 handicapped-infants ().1007458 11 synfuels-corporation-cutback ().0529956 16 export-administration-act-south-africa ().0049097 10 immigration ().0000117 2 water-project-cost-sharing Selected attributes: 4,3,5,12,14,8,9,13,15,7,6,1,11,16,10,2 : 16 ().1007458 11 synfuels-corporation-cutback ().1007458 11 synfuels-corporation-cutback ().1007458 11 synfuels-corporation-cutback ().1007458 11 synfuels-corporation-cutback ().1007458 12 synfuels-corporation-cutback ().1007458 13 synfuels-corporation-cutback ().1007458 14 synfuels-corporation-cutback ().1007458 15 synfuels-corporation-cutback ().1007458 15 synfuels-corporation-cutback ().1007458 15 synfuels-corporation-cutback ().1007458 15 synfuels-corporation-cutback ().1007458 16 synfuels-corporation-cutback ().1007458 16 synfuels-corporation-cutback ().1007458 16 synfuels-corporation-cutback ().1007458 16 synfuels-corporation-cutback ().1007458 17 synfuels-corporation-cutback ().1007458 18 synfuels-corporation-cutback ().1007458 18 synfuels-corporation-cutback ().1007458 19 synfuels-corporation-cutback ().1007458 19 synfuels-corporation-cutback ().1007458 18 synfuels-corporation-cutback ().1007458 18</pre>
Status	
OK	Log 🔊 x 0

Which attribute selector?

Best: WRAPPER

- Slow: O(2^N) search through all attribute combinations
- The "wrapped" learner called to assess each combination
- Some heuristics to prune the search; but does not scale

If not WRAPPER

- Use InfoGain / OneR for very big datasets
- Use CFS otherwise
- Don't use PCA
 - This is an unsupervised selector
 - So it is uninformed on how dimensions help classification

Explorer: data visualization

- Visualization very useful in practice: e.g. help s to determine difficulty of the learning proble m
- WEKA can visualize single attributes (1-d) an d pairs of attributes (2-d)
 - 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

00	Weka Knowle	edge Explorer	
	Preprocess Classify Cluster As	ssociate Select attributes	Visualize
Open file	Open URL Oper	n DB Undo	Save
Choose None			Apply
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 7 Ca 8 Ba 9 Fe 10 Type	Name	Statistic Minimum Mean StdDev Colour: Type (Nom)	Value 1.511 1.534 1.518 0.003 Visualize All
Status OK			Log x 0

🖯 🖯 Weka Knowledge Explorer							
	Preprocess	Classify C	Cluster Associate	Select attribute	s Visualize		
Plot Matrix	RI	Na	Mg	AI	Si	к	
	· · · · · · · · ·				··· · ····· ·		
Туре							U
			· · · · ·				
Fe							
							Ă
	Ċ			· .			
PlotSize: [100]	PlotSize: [100]						
PointSize: [1] Update							
Jitter: Select Attributes							
Colour: Type (Nom) SubSample % : 100					100		
Class Colour							
build wind	build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps						
Status							
OK					C	Log 🖉 🛷 '	x 0








Limitations

- Loads all data into ram prior to learning
 - Problem for large data sets
- Not good for complex experiments
- IMHO, discourages experimentation with new learners
 - The "WEKA effect"
 - Try every learner till something works
- Still, very useful for
 - Initial investigations
 - Learning data mining
 - Or as a sub-routine of other systems

Alternate tools: Orange



Written in Python

Simpler specification (but see WEKA's KnowledgeFlow Environment).

Also, less community support/ debugging. So sometimes frustrated by random bugs

Alternate tools: RapidMiner



Experiments specified in an XML tree syntax

In theory, possible to share experimental descriptions

Alternate tools: OurMine

Java=\$Base/lib/java

Weka="java -Xmx2048M -cp \$Java/weka.jar " Clusterers="java -Xmx1024M -jar \$Java/Clusterers.jar " Reducers="java -Xmx1024M -jar \$Java/Reduce.jar "

nb() {

```
local learner=weka.classifiers.bayes.NaiveBayes
$Weka $learner -p 0 -t $1 -T $2
```

nb10() {

```
local learner=weka.classifiers.bayes.NaiveBayes
$Weka $learner -i -t $1
```

}

}

```
j48() {
```

```
local learner=weka.classifiers.trees.J48
$Weka $learner -p 0 -C 0.25 -M 2 -t $1 -T $2
}
```

Forget the visuals.

Make WEKA a sub-routine inside Bash script

Now you can mix WEKA's JAV A with learners written in your favorite language.

But how do you find the magic command strings?

Why go to all that trouble?

analysis1(){ local origdata=\$1 local outstats=\$2 local nattrs="2 4 6 8 10 12 14 16 18 20" local learners="nb10 j4810 zeror10 oner10 adtree10" local reducers="infogain chisquared oneR" local tmpred=\$Tmp/red echo "n,reducer,learner,accuracy" > \$outstats

```
for n in $nattrs; do
for reducer in $reducers; do
$reducer $origdata $n $tmpred
for learner in $learners; do
accur=`$learner $tmpred.arff | acc
out="$n,$reducer,$learner,$accur"
blabln $out
echo $out >> $outstats
done
done
```

Complex experiments, specifie d succinctly.

Experiments can now be revie wed, audited, by others.

Also, in 12 months time when Reviewer2 wants a tiny extension to the old paper, you don't have to remember all that clicking you did: just rerun the script.



ОК

