Machine Learning with WEKA

Eibe Frank

Department of Computer Science, University of Waikato, New Zealand

- WEKA: A Machine Learning Toolkit
- The Explorer
 - Classification and Regression
 - Clustering
 - Association Rules
 - Attribute Selection
 - Data Visualization
- The Experimenter
- The Knowledge Flow GUI
- Conclusions

WEKA: the bird



Copyright: Martin Kramer (mkramer@wxs.nl)



WEKA: the software

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

WEKA: versions

There are several versions of WEKA:

- WEKA 3.0: "book version" compatible with description in data mining book
- WEKA 3.2: "GUI version" adds graphical user interfaces (book version is command-line only)
- WEKA 3.3: "development version" with lots of improvements
- This talk is based on the latest snapshot of WEKA 3.3 (soon to be WEKA 3.4)

WEKA only deals 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



. . .

WEKA only deals with "flat" files

numeric attribute @attribute age numeric nominal attribute @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

@relation heart-disease-simplified

. . .



2/22/2011



2/22/2011



2/22/2011

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, ...

00	Weka Knowledge Explorer							
	Preprocess Classify Cluster	Associate Select attri	butes Visualize					
Open file	Open URL O	pen 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								
				Visualize All				
Status Welcome to the Wek	a Knowledge Explorer		C	Log x 0				
freiconie to the wek	a knowledge Explorer			ALC: NOT				

163

Welcome to the Weka Knowledge Explorer



Welcome to the Weka Knowledge Explorer



x 0





OK















OK



x 0







Status -OK

Log 🦽

x 0





🕽 🖯 🕤 🦳 Weka Knowledge Explorer							
	Preprocess Classify	Cluster Associa	te Select attributes V	isualize			
Open file	Open URL	Open DB	Undo		Save		
Filter							
Choose Discret	ize -B 10 -R first-last	\varTheta 🖯 🖯 🛛 weka	a.gui.GenericObjectEditor		Apply		
Current relation		weka.filters.unsuperv	vised.attribute.Discretize				
Relation: iris Instances: 150	Attributes:	About An instance filter that d	iscretizes a range of numeric	More	: Numeric : 10 (7%)		
Attributes		attributes in the datase	t into nominal attributes.	More	e		
No. 1 sepallengt	Name	attributeIndices	first-last				
3 petallengt	h	bins	10				
4 petalwidth 5 class	1	findNumBins	False	;			
		invertSelection	False	•	Visualize All		
		makeBinary	False	•			
		useEqualFrequency	False	•			
		Open	iave OK	Cancel			
		11	2 0 0 3 4		10 2 4		
Status		- 1	3.9	15	6.9		
OK				Log) × 0		



00	O Weka Knowledge Explorer								
	Preprocess Classif	y Cluster Associa	te Select attributes Visualize						
Open file	Open URL	Open DB	Undo	Save					
Filter									
Choose Discret	ize -B 10 -R first-last	\varTheta 🖯 🖯 🛛 weka	a.gui.GenericObjectEditor	Apply					
Current relation		weka.filters.unsuperv	ised.attribute.Discretize						
Relation: iris Instances: 150	Attributes:	About An instance filter that d	iscretizes a range of numeric More	: Numeric : 10 (7%)					
Attributes		attributes in the datase	attributes in the dataset into nominal attributes.						
No. 1 sepallengt	Name h	attributeIndices	first-last						
3 petallengt	h	bins	10						
4 petalwidth 5 class	1	findNumBins	False	0					
		invertSelection	False	Visualize All					
		makeBinary	False						
		useEqualFrequency	True						
		Open	ave OK Cancel	\supset					
		11	2 0 0 3 4	10 2 ⁴					
Status		. 1	3.95	6.9					
OK			Lo	g 💉 x 0					

000	O Weka Knowledge Explorer								
	Preprocess Classify	Cluster Associa	te Select attributes Visualize						
Open file	Open URL	Open DB	Undo	Save					
Filter									
Choose Discreti	ze -B 10 -R first-last	\varTheta 🖯 🖯 🛛 weka	a.gui.GenericObjectEditor	Apply					
Current relation		weka.filters.unsuperv	vised.attribute.Discretize						
Relation: iris Instances: 150	Attributes:	About An instance filter that d	liscretizes a range of numeric More	: Numeric : 10 (7%)					
Attributes		attributes in the datase	et into nominal attributes.	e					
No. 1 sepallengt	Name h	attributeIndices	first-last]					
3 petallength	h	bins	10						
4 petalwidth 5 class		findNumBins	False 🛟						
		invertSelection	False 🗧	Visualize All					
		makeBinary	False						
		useEqualFrequency	True						
		Open	Gave OK Cancel)					
			2 0 0 3 4	10 2 4 6 9					
Status			0.00	0.0					
OK			Log	x0					







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, ...



Weka Knowledge Explorer

-

	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
Choose ZeroR							
Test options		Clas	ssifier output				
O Use training set							
Supplied test se	t Set						
Cross-validation	n Folds 10						
O Percentage split	% 66						
More op	otions						
(Nom) class		•					
Start	Stop						
Result list (right-click for	r options)						
Status							
ОК						C	Log 🔵 🛷 x O

Weka Knowledge Explorer

000			WERA KIN	owieuge Lxpi	orei		
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
Choose ZeroR							
Test options		Clas	ssifier output				
Use training set							
O Supplied test set	set						
Cross-validation	n Folds 10						
O Percentage split	% 66						
More op	otions						
(Nom) class		•					
Start	Stop						
Result list (right-click for	options)						

OK





000

Weka Knowledge Explorer

	()					()
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
🗐 weka							
v 🗊 classifiers			-				
📄 🕨 🧊 bayes			ifier output				
▶ 📁 functions							
🕨 🧊 lazy							
🕨 📁 🔰 meta							
🕨 🧊 misc							
🔻 🧊 trees							
🕨 📔 adtree							
Decision	Stump						
🕒 📄 ld3							
🛛 🔻 🧊 j48							
J48							
🔰 🕨 🔰 İmt							
📄 🕨 🎾 m5							
Random	Forest						
Random	Tree						
REPTree							
UserCla	ssifier						
rules							
Status							
ОК						C	Log 💉 🔊 x 0



.....

😇 🖯 🗧 Weka Knowledge Explorer									
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize]		
Classifier									
Choose J48 -C	Choose J48 -C 0.25 -M 2								
Test options		Clas	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									

ОК








	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	2
Classifier							
Choose J48 -C	0.25 –M 2	000) weka.gui	.GenericObje	ctEditor		
Test options		weka.clas	sifiers.tree	s.j48.J48			
O Use training set			binarySplits	False	•		
O Supplied test se	t Set	confi	denceFacto	r 0.25			
 Cross-validation Percentage split 	n Folds 10 % 66		minNumOb	j 2			
More op	otions		numFolds	; 3			
(Nom) class		reduced	ErrorPruning	False	•		
(Start)	Stop	savelr	nstanceData	a False	•		
Result list (right-click for	options)	sut	otreeRaising	True	•		
			unpruned	False	•		
			useLaplace	False	÷		
		Open	. Save.) ОК	Cancel		
Status							
ОК						C	Log 💉 X O



.

	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
Choose J48 -C 0	.25 -M 2	000	weka.gui	.GenericObje	ctEditor		
Test options		weka.clas	sifiers.tree	s.j48.J48			
O Use training set			binarySplits	False	•		
O Supplied test set	Set	confi	denceFacto	0.25			
Cross-validation	Folds 10		minNumOb	i 2			
O Percentage split	% 66	_		2			
More opt	ions		numFolds	3			
(Nom) class		reducedE	FrrorPruning	False	•		
		savelr	nstanceData	False	•		
Start Result list (right-click for c	Stop	sub	otreeRaising	True	•		
	59.000		unpruned	False	•		
			useLaplace	False	•		
		Open	Save.) (ОК	Cancel		
Status						6	



.....

000	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										
O Supplied test set	set									
Cross-validation	Folds 10									
O Percentage split	% 66									
More op	tions									
(Nom) class		•								
(Start)	Stop									
Result list (right-click for	options)									

ОК





. 1.0

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

ОК





ociate	Select attributes	Visualize
ociate	Select attributes	Visualiz







ociate	Select attributes	Visualize
ourace	Select attributes	VISUAIIZE





N

	Colors attaileater	Visualizza
e	Select attributes	Visualize

	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											
O Cross-validation	n Folds 10		00	🖯 Classifie	er evaluation opt							
• Percentage split	% 66		🗹 Out	put model								
More op	otions		🗹 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	r options)		🗹 Stor	e prediction	s for visualization							
			🗌 Out	put text pre	dictions on test se	et						
			Cos	t-sensitive	evaluation Set							
			Random	n seed for X\	/al / % Split 1							
			\square		ОК	\supset						
1		194										









ociate	Select attributes	Visualize
ociate	Select attributes	Visualiz







Weka Knowledge Explorer

	r .		_
ssociate	Select attributes	Visualize	

	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
Choose J48 -C	0.25 -M 2						
Test options		Clas	sifier output	t			
🔘 Use training set							
O Supplied test se	t Set						
O Cross-validation	n Folds 10						
🖲 Percentage split	% 66						
More op	otions						
(Nom) class		+					
Start .	Stop						
Result list (right-click in	r options)						
	opuono,						
Status						6	
OK						C	Log X 0



Select attributes Visualize

Log



Status



Select attributes Visualize

Log



Status



Classify Cluster Associate Select attributes Vicualiza

Log

	Preprocess	Classify	Cluster	Associate	Select attrib	outes Visi	ualize	
Classifier								
Choose J48 -C 0.	25 -M 2							
Test options		Classifi	er output					
🔘 Use training set		Time	taken to bu	uild model: (1.24 seconds			
O Supplied test set	Set		valuation of	n tost split				
O Cross-validation	Folds 10	=== S	ummary ===	m cesc spir				
Percentage split More option	% 66 ns	Corre Incor Kappa	ctly Classi rectly Clas statistic	fied Instand sified Insta	ces ances	49 2 0.9408	96.0784 3.9216	8
(Nom) class		Root Relat Root	absolute er mean square ive absolut relative so	ror ed error ee error quared error		0.0396 0.1579 8.8979 % 33.4091 %		
Start	Stop	Total	Number of	Instances		51		
Result list (right-click for or	ptions)	D	etailed Acc	curacy By Cla	155 ===			
11:49:05 – trees.j48.J4	8	TP Rat 1 1 0.8	te FP Rat 0 0.06 82 0	e Precisio 1 3 0.901 1	on Recall 1 5 1 0.882	F-Measure 1 0.95 0.938	Class Iris-setosa Iris-versico Iris-virgini	lor .ca
		=== C a 15 0 1 0	onfusion Ma b c < 0 0 a = 9 0 b = 2 15 c =	classified a Iris-setosa Iris-versia Iris-virgin	as a color nica			

Status

ОК



V⁻

1 ve

~ × 0

Log

	Preprocess	Classify	Cluster	Associate	Select attrib	outes	Visualize		
Classifier Choose J48 -C	0.25 -M 2								
 Test options Use training set Supplied test set 	t Set	Classifi	eroutput	uild model:	0.24 seconds	;			Ĩ
Cross-validation Percentage split More opt (Nom) class	n Folds 10 % 66 ions	Correct Incorr Kappa Mean a Root r Relat Root	=== Evaluation on test split === === Summary === Correctly Classified Instances 49 96.0784 % Incorrectly Classified Instances 2 3.9216 % Kappa statistic 0.9408 Mean absolute error 0.0396 Root mean squared error 0.1579 Relative absolute error 8.8979 %						
Start Result list (right-click for 11:49:05 - trees.j48	Stop options) J48	Total Do TP Ray 1 0.89 Co a 1 15	Number of etailed Acc te FP Rat 0.00 32 0 onfusion Ma o c <	Instances curacy By Cl te Precisi 1 63 0.90 1 atrix === classified = Iris-setos	ass === on Recall 1 5 1 0.882 as	51 F-Measu 1 0.99 0.93	ure Clas Iris 5 Iris 38 Iris	s -setosa -versicolor -virginica	
		0 1	90 b: 215 c:	= Iris-versi = Iris-virgi	color nica				4

-Status -



	Preprocess	Classify	Cluster	Associate	Select at	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 t	valuation	on test spli	0.24 seco	nds			
O Cross-validation	Folds 10	==== St							
• Percentage split	% 66	Correc	ctly Class rectly Cla	ified Instan ssified Inst	ces ances	49 2		96.0784 % 3.9216 %	
More opti	ons	Mean a	statistic absolute e	rror		0.94	108 396 579		
(Nom) class		Relat:	ive absolu relative s	te error quared error		8.89	979 %)91 %		
Start Carl	Stop	D Total	etailed Ac	curacy By Cl	ass ===	51			
Result list (right-click for	Vie Vie	ew in main	window		Reca	ll F-Mea	asure Cla	ISS	
11:49:05 - trees.J48.	J48 Vie	ew in separ	ate windo	w	1	1	Iri	.s-setosa	
	Sa	ve result bi	uffer		1 0.8	0. 82 0.	.95 Iri .938 Iri	s-versicolo s-virginica	r
	Lo	ad model							
	Sa	ve model							
	Re	-evaluate r	model on (current test s	set				U
	Vis	sualize clas	sifer error	rs					Ă
	Vis	sualize tree							<u> </u>
Status	Vis	sualize mar	gin curve						
OK	VIS	sualize thre	esnola cur	ve			C	Log	• × 0
	VIS	sualize cos	curve						



Associate Select attributes

Log

x 0





	Preprocess	Classify	Cluster	Associate	Select attri	butes	Visualize		
Classifier									
Choose J48 -C	0.25 -M 2								
Test options		Classifi	er output						
🔘 Use training set		Time	takan ta b	udld model.	0.04 accords				
O Supplied test set	t Set) E	valuation	on test spli	t ===	5			
Cross-validation	n Folds 10	=== S	ummary ===						
• Percentage split	% 66	Corre	ctly Class rectly Cla	ified Instan ssified Inst	ces ances	49 2		96.0784 % 3.9216 %	
More opti	ions	Mean Root	statistic absolute e mean squar	error ed error		0.94	08 96 79		
(Nom) class		Relat Root	ive absolu relative s Number of	quared error		8.89 33.40 51	79 % 91 %		
Start Start	Stop) D	etailed Ac	curacy By Cl	ass ===				
11:49:05 - trees i48	View i	in main wir	ndow		Recall	F-Mea:	sure Cla	SS	
11.45.05 - ((cc3.)40.	View i Save i	in separate result buffe	e window er		1 1 0.882	1 0. 0.	Iri 95 Iri 938 Iri	s-setosa s-versicolo s-virginica	r
	Load Save r	model model	del en eur						
	Visua	lize classife	er errors	rent test set	lor ca				
	Visua	lize tree							
Status OK	Visua Visua	lize margir lize thresh	old curve		•		C	Log	x 0
	Visua	lize cost cu	irve		·				



Weka Knowledge Explorer



Status-

OK



🕋 x 0



Classify Cluster Associate Select attributes Vicualiza

Log

	Preprocess	Classify	Cluster	Associate	Select attrib	outes Visi	ualize				
Classifier											
Choose J48 -C 0.	25 -M 2										
Test options		Classifi	er output								
🔘 Use training set		Time	Time taken to build model: 0.24 seconds								
O Supplied test set	Set		=== Evaluation on test split ===								
O Cross-validation	Folds 10	=== S	ummary ===	m cesc spir							
Percentage split More option	% 66 ns	Corre Incor Kappa	ctly Classi rectly Clas statistic	fied Instand sified Insta	ces ances	49 2 0.9408	96.0784 3.9216	8			
(Nom) class		Root Relat Root	absolute er mean square ive absolut relative so	ror ed error ee error quared error		0.0396 0.1579 8.8979 % 33.4091 %					
Start	Stop	Total	Number of	Instances		51					
Result list (right-click for or	ptions)	D	etailed Acc	curacy By Cla	155 ===						
11:49:05 – trees.j48.J4	8	TP Rat 1 1 0.8	te FP Rat 0 0.06 82 0	e Precisio 1 3 0.901 1	on Recall 1 5 1 0.882	F-Measure 1 0.95 0.938	Class Iris-setosa Iris-versico Iris-virgini	lor .ca			
		=== C a 15 0 1 0	onfusion Ma b c < 0 0 a = 9 0 b = 2 15 c =	classified a Iris-setosa Iris-versia Iris-virgin	as a color nica						

Status

ОК

 $\mathbf{0}$

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Log

		1 0.000011	C. aster	risserate	o creat atting			
Classifier								
Choose J48 -C	0.25 -M 2							
Test options		Classif	er output					
O Use training set		Time	taken to b	uild model:	0.24 seconds			
O Supplied test se	t Set		valuation	on test spli	+ ===			
O Cross-validation	n Folds 10	==== S	ummary ===	on cope opri				
• Percentage split	% 66	Corre	ctly Class	ified Instan ssified Inst	ces ances	49 2	96.0784 % 3.9216 %	
More opt	ions) Kappa Mean	statistic absolute e	rror		0.9408		
		Root	mean squar	ed error		0.1579		
(Nom) class		Relat Root	ive absolu relative s	te error quared error		8.8979 % 33.4091 %		
Start	Stop	Total	Number of	Instances		51		
Result list (right-click for	roptions)	=== D	etailed Ac	curacy By Cl	ass ===			
11:49:05 - trees.j48	.J48	TP Ra 1	te FP Ra O	te Precisi 1	on Recall 1	F-Measure 1	Class Iris-setosa	
		1	0.0	63 0.90 1	5 1	0.95	Iris-versicolor	
		=== c	onfusion M	atrix ===	0.002	0.950	iiis viiginida	
		a 15	b c <	classified	as			
		0 1	9 0 b	= Iris-versi	color			
		0	215 C	= Iris-virgi	nica			Ŧ
		=== C a 15 0 1 0	onfusion M b c < 0 0 a 9 0 b 2 15 c	atrix === classified = Iris-setos = Iris-versi = Iris-virgi	as a color nica			

Status-

00

Weka Knowledge Explorer

Select attributes Visualize

Preprocess Classify Cluster Associate Classifier weka classifiers T output bayes functions ken to build model: 0.24 seconds LeastMedSq LinearRegression luation on test split === mary === Logistic 🧊 neural T ly Classified Instances 49 96.0784 % ctly Classified Instances 2 3.9216 % NeuralNetwork 0.9408 statistic ► pace solute error 0.0396 supportVector an squared error 0.1579 re absolute error 8.8979 % SimpleLinearRegression lative squared error 33.4091 % SimpleLogistic lumber of Instances 51 VotedPerceptron ailed Accuracy By Class === Winnow 🔲 lazy FP Rate Precision Recall F-Measure Class meta 0 1 1 1 Iris-setosa 0.063 0.905 1 0.95 Iris-versicolor misc 0.882 0.938 Iris-virginica 0 1 trees 🧻 rules fusion Matrix === <-- classified as 15 0 a = Iris-setosa 0 b = Iris-versicolor 0 0 1 19 0 2 15 | c = Iris-virginica ÷





Select attributes

Log

Classify Cluster Associate Preprocess Visualize Classifier NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a Choose Test options -Classifier output -Use training set === Evaluation on test split === Supplied test set Set.... === Summary === Cross-validation Folds 10 Correctly Classified Instances 49 96.0784 % 3,9216 % Incorrectly Classified Instances 2 Percentage split % 66 Kappa statistic 0.9408 Mean absolute error 0.0396 More options... Root mean squared error 0.1579 Relative absolute error 8.8979 % Root relative squared error 33.4091 % + (Nom) class Total Number of Instances 51 === Detailed Accuracy By Class === Stop Start TP Rate FP Rate Precision Recall F-Measure Class Result list (right-click for options) Iris-setosa 1 0 1 1 1 11:49:05 - trees.j48.J48 0.063 0.905 0.95 Iris-versicolor 1 1 0.882 0.882 0 1 0.938 Iris-virginica === Confusion Matrix === <-- classified as bc а 15 0 0 | a = Iris-setosa 0 19 0 | b = Iris-versicolor 0 2 15 | c = Iris-virginica +

Status



Preprocess Classify Cluster Associate Select attributes Visualize

×0

Log

⊂Classifier							
Choose NeuralNetwork -1.0.3 -N	102 -N 50	0 - 1 0 - 5 0	-E 20 -H a				
	10.2 11 54		220 114				
Test options	-Classifier of	output					
O Use training set		on oo ooss					6
O Supplied test set Set	=== Eval	uation on t	test split ==	==			
	- Cult	in the second se					
Cross-validation Folds 10	Correctl	y Classifie	ed Instances		49	96.0784 %	
Percentage split % 66	Kappa st	tly Classi	fied Instance	es	2	3.9216 %	
	Mean abs	olute erro:	r		0.0396		
More options	Root mea	n squared (error		0.1579		
	Relative	absolute (error		8.8979 %		
(Nom) class	Root rel	ative squar	red error		33.4091 %		
	Total Nu	mber of In	stances		51		
Start Stop	=== Deta	iled Accura	acy By Class	===			
-Result list (right-click to options)	TP Rate	FP Rate	Precision	Recall	F-Measure	Class	
	1	0	1	1	1	Iris-setosa	
11:49:05 – trees.j48.J48	1	0.063	0.905	1	0.95	Iris-versicolor	
	0.882	U	1	0.882	0.938	iris-virginica	
· · · · · · · · · · · · · · · · · · ·	=== Conf	usion Matr	ix ===				
	15 0	c < cla	assilled as				
	0 19	$0 \mid b = I$	ris-versicol	or			
	0 2 1	5 c = I	ris-virginica	a			
							A I
							•

Status







Select attributes Visualize

Log

Classify Cluster Preprocess Associate Classifier NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R Choose) Test options -Classifier output -()Use training set === Evaluation on test split === === Summary === Supplied test set Set.... Correctly Classified Instances 98.0392 % 50 Cross-validation Folds 10 Incorrectly Classified Instances 1.9608 % 1 Kappa statistic 0.9704 Percentage split % 66 Mean absolute error 0.0239 Root mean squared error 0.1101 More options... Relative absolute error 5.3594 % Root relative squared error 23.2952 % Total Number of Instances 51 + (Nom) class === Detailed Accuracy By Class === Start Stop FP Rate Precision Recall F-Measure TP Rate Class 0 Iris-setosa 1 1 1 1 Result list (right-click for options) 0.95 0.031 1 0.974 Iris-versicolor 1 0.941 0.97 Iris-virginica 0.941 0 1 11:49:05 - trees.j48.J48 14:34:28 - functions.neural.NeuralNetwork === Confusion Matrix === b C <-- classified as а 15 0 0 | a = Iris-setosa 0 19 0 | b = Iris-versicolor 0 1 16 | c = Iris-virginica 4

Status



 $\mathbf{0}$

Weka Knowledge Explorer

Classify Cluster Associate Select attributes Visualize

Log

	Preprocess	Classify	Cluster	Associat	e Select at	tributes	Visualize					
Classifier Choose Neural	Choose NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R											
Test options O Use training set			Classifier out === Evalua === Summan	tput ation on t ry ===	est split ==							
Cross-validation Percentage split More	n Folds 10 % 66 options		Correctly Incorrect Kappa stat Mean absol Root mean Relative a Root relat	Classifie ly Classif tistic lute error squared e absolute e tive squar	d Instances ied Instance rror rror ed error	23	50 1 0.9704 0.0239 0.1101 5.3594 % 23.2952 %	98.0392 % 1.9608 %				
(Nom) class	Stor	•	=== Detail	led Accura	cy By Class	===	F-Moaguro	Class				
Result list (right-click fo 11:49:05 - trees.j48	r options) J48 s. neural Neura	Network	1 1 0.941	0 0.031 0	1 0.95 1	1 1 0.941	1 0.974 0.97	Iris-setosa Iris-versicolor Iris-virginica				
			=== Confus a b c 15 0 0 0 19 0 0 1 16	<pre>sion Matri < cla a = Ir b = Ir c = Ir</pre>	x === ssified as is-setosa is-versicolo is-virginica	or a		▲ ▼ ()				

Status-



00

Weka Knowledge Explorer

Associate

Cluster

Classify

Preprocess

Select attributes

utes Visualize

Log 💉 🗴 0

Classifier		=							
📁 weka		1							
🔻 河 classifiers		1							
🐨 💓 bayes	assifier output								
AODE	-= Evaluation on test split ===	5							
BayesNetK2	== Summary ===								
BayesNetB	prrectly Classified Instances 50 98 0392 %								
NaiveBayes	correctly Classified Instances 1 1.9608 %								
NaiveBayesMultinomial	appa statistic 0.9704								
NaiveBayesSimple	ean absolute error 0.0239								
NaiveBayesUpdateable	elative absolute error 5.3594 %								
▶ 📁 functions	pot relative squared error 23.2952 %								
▶ 📁 lazy	otal Number of Instances 51								
▶ 📁 meta	== Detailed Accuracy By Class ===								
▶ 河 misc									
▶ 📋 trees	P Rate FP Rate Precision Recall F-Measure Class								
I ▶ 📋 rules	1 0.031 0.95 1 0.974 Iris-versicolor								
	0.941 0 1 0.941 0.97 Iris-virginica								
	== 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								
		Y							



Classify Cluster Associate Select attributes Visualiza

×0

Log

	Preprocess	Classify	Cluster	Associate	Select at	ttributes	Visualize	
Classifier Choose NaiveBa	iyes							
Test options			Classifier out	put				
🔘 Use training set			=== Evalua	ation on te	st split ==			
O Supplied test set	Set	.)	=== Summaı	су ===				
O Cross-validation	Folds 10		Correctly Incorrectl	Classified y Classifie	Instances ed Instance	es	50 1	98.0392 % 1.9608 %
Percentage split	% 66		Kappa stat Mean absol	istic ute error			0.9704	
More of	options		Root mean Relative a Root relat	squared er absolute er ive square	ror ror d error		0.1101 5.3594 % 23.2952 %	
(Nom) class		+	Total Numb	per of Insta	ances		51	
			=== Detail	ed Accurac	y By Class	===		
Start	Stop		TP Rate	FP Rate	Precision	Recall	F-Measure	Class
Result list (right-click for	options)		1	0.031	1 0.95	1	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.Neural	Network	=== Confus	sion Matrix	=== sified as			
			15 0 0 0 19 0 0 1 16	a = Iri: b = Iri: c = Iri:	s-setosa s-versicolo s-virginica	or a		
)4+



Preprocess Classify Cluster Associate Select attributes Visualize

	.ep.eccos	C,								
Classifier										
Choose NaiveBay	es									
Test options			Classifier ou	tput						
🖯 Use training set			=== Evaluation on test split ===							
O Supplied test set	Set	.)	=== Summa:							
Cross-validation	Folds 10		Correctly Classified Instances 50 98.0392 % Incorrectly Classified Instances 1 1.9608 %							
Percentage split	% 66		Kappa stat Mean abso	tistic lute error			0.9704			
More opt	tions	\square	Root mean Relative a Root relat	squared es absolute es tive square	rror rror ed error		0.1101 5.3594 % 23.2952 %			
(Nom) class		÷	Total Num	per of Ins	tances		51			
			=== Detai.	led Accura	cy By Class					
Start	Stop		TP Rate	FP Rate	Precision	Recall	F-Measure	Class		
Result list (right-click for on	otions)		1	0.031	1	1	1	Iris-setosa Iris-versicolor		
11:49:05 - trees.j48.J4	8		0.941	0	1	0.941	0.97	Iris-virginica		
14:34:28 - functions.n	eural. Veural	Network	=== Confu	sion Matri:	x ===					
	•		a b c 15 0 0 0 19 0 0 1 16	< cla a = Ir b = Ir c = Ir	ssified as is-setosa is-versicolo is-virginica	or 1				
) 4 1 4		



 $\mathbf{0}$

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Log

	1			1		I	·		
Classifier									
Choose NaiveBayes									
Test options		Classifier ou	tput						
O Use training set	-	=== Evalua	ation on te	st split ==			0		
O Supplied test set Set		=== Summa:	ry ===	Trakanaa		40	04 1176 6		
Correctly Classified Instances 48 94.1176 % Incorrectly Classified Instances 3 5.8824 %									
Percentage split % 66		Kappa sta Mean abso	tistic lute error			0.9113			
More options		Root mean Relative a Root relat	squared er absolute er tive square	ror ror d error		0.1722 10.0365 % 36.4196 %			
(Nom) class Total Number of Instances 51									
		=== Detail	led Accurac	y By Class					
Start Stop		TP Rate	FP Rate	Precision	Recall	F-Measure	Class		
Result list (right-click for options)		1 0.947	0 0.063	1 0.9	1 0.947	1 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 - functions.neural.NeuralNetv 14:48:05 - bayes.NaiveBayes	vork	=== Confu	sion Matrix						
		a b c 15 0 0 0 18 1 0 2 15	< clas a = Iri b = Iri c = Iri	sified as s-setosa s-versicolo s-virginica	or I				
)∢►		
Status									

00

OK

Weka Knowledge Explorer

Select attributes

Log

Classify Cluster Preprocess Associate Visualize Classifier NaiveBayes Choose) Test options Classifier output - \bigcirc Use training set === Evaluation on test split === === Summary === Supplied test set Set.... Correctly Classified Instances 94.1176 % 48 Cross-validation Folds 10 5.8824 % Incorrectly Classified Instances 3 Kappa statistic 0.9113 Percentage split % 66 Mean absolute error 0.0447 Root mean squared error 0.1722 More options... Relative absolute error 10.0365 % Root relative squared error 36.4196 % Total Number of Instances 51 + (Nom) class === Detailed Accuracy By Class === Start Stop FP Rate Precision Recall F-Measure Class TP Rate 0 Iris-setosa 1 1 1 1 Result list (right-click for options) 0.9 0.947 0.063 0.947 0.923 Iris-versicolor 0.882 0.029 0.938 0.882 0.909 Iris-virginica 11:49:05 - trees.j48.J48 14:34:28 – functions.neural.NeuralNetwork === Confusion Matrix === 14:48:05 - bayes.NaiveBayes <-- classified as а b C 15 0 0 | a = Iris-setosa 0 18 1 | b = Iris-versicolor 0 2 15 | c = Iris-virginica 4 1 Status

 $\bigcirc \bigcirc \bigcirc \bigcirc$

Weka Knowledge Explorer

	Preprocess	Classify	Cluster	Associate	Select at	tributes	Visualize	
Classifier Choose NaiveBa	iyes							
Test options			Classifier out	put				
O Use training set	Set		=== Evalua === Summar	ation on tes :y ===	st split ==			<u> </u>
 Cross-validation Percentage split 	Folds 10 % 66		Correctly Incorrectl Kappa stat Mean absol	Classified y Classifie istic ute error	Instances d Instance	25	48 3 0.9113 0.0447	94.1176 % 5.8824 %
(Nom) class	Root mean squared error Relative absolute error Root relative squared error Total Number of Instances				0.1722 10.0365 % 36.4196 % 51			
Start Result list (right-click for 11:49:05 - trees.j48, 14:24:28 - functions	View in ma View in se Save resul	ain window parate win t buffer	=== Detail / dow	ed Accuracy	Precision 1 0.9 0.938	=== Recall 1 0.947 0.882	F-Measure 1 0.923 0.909	Class Iris-setosa Iris-versicolor Iris-virginica
14:34:28 - functions 14:48:05 - bayes.Na	Load mode Save mode Re-evaluat Visualize c Visualize t Visualize r Visualize t	el te model o classifer en ree nargin cur hreshold o	n current 1 rors ve :urve	est set :is :is :is	=== ified as -setosa -versicolo -virginica ris-setosa	or 1		A 7
Status OK	Visualize o	ost curve		•	ris–versico ris–virgini	olor ca	C	Log x 0



Weka Knowledge Explorer



Log

 $\mathbf{0}$

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Log

	1			1			·
Classifier							
Choose NaiveBayes							
Test options		Classifier ou	tput				
O Use training set	-	=== Evaluation on test split ===					
O Supplied test set Set		=== Summa:	ry ===	Trakanaa		40	04 1176 6
Cross-validation Folds 10		Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error Relative absolute error Root relative squared error				48 3	5.8824 %
Percentage split % 66						0.9113	
More options						0.1722 10.0365 % 36.4196 %	
(Nom) class	•	Total Num	per of Inst	ances		51	
		=== Detailed Accuracy By Class ===					
Start Stop		TP Rate	FP Rate	Precision	Recall	F-Measure	Class
Result list (right-click for options)		1 0.947	0 0.063	1 0.9	1 0.947	1 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 – functions.neural.NeuralNetv 14:48:05 – bayes.NaiveBayes	vork	=== Confusion Matrix ===					
		a b c 15 0 0 0 18 1 0 2 15	< clas a = Iri b = Iri c = Iri	sified as s-setosa s-versicolo s-virginica	or I		
) 4 ► (
Status							
Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

- × 0

Log

		1	1
Classifier			
Choose NaiveBayes			
Test options	Classifier output		
O Use training set	=== Evaluation on test split ===		0
O Supplied test set	=== Summary ===		
O Cross validation Folds 10	Correctly Classified Instances	48	94.1176 %
Percentage split % 66	Kappa statistic	0.9113	5.8824 %
ercentage spint % 00	Mean absolute error	0.0447	
More options	Relative absolute error	10.0365 %	
	Root relative squared error	36.4196 % 51	
(Nom) class	Total Malber of Instances	<u></u>	
	=== Detailed Accuracy By Class ===		
Start Stop	TP Rate FP Rate Precision Recal	I F-Measure	Class
Result list (right-click for options)	1 0 1 1 0.947 0.063 0.9 0.94	1 7 0.923	Iris-setosa Iris-versicolor
11:49:05 - trees.j48.J48	0.882 0.029 0.938 0.88	2 0.909	Iris-virginica
14:34:28 - functions.neural.NeuralNetwork	=== Confusion Matrix ===		
14:48:05 – bayes.NaiveBayes	a h c < classified as		
	15 0 0 a = Iris-setosa		
	0 18 1 b = Iris-versicolor 0 2 15 c = Iris-virginica		
	· · · · · · · · · · · · · · · · · · ·		×
			V
Status			

OK

OK

Weka Knowledge Explorer

Select attributes Visualize

Log

Preprocess Classify Cluster Associate Classifier weka classifiers T assifier output bayes functions == Evaluation on test split === == Summary === lazy meta orrectly Classified Instances 94.1176 % 48 ncorrectly Classified Instances 5.8824 % misc 3 appa statistic 0.9113 trees ean absolute error 0.0447 Image: A laboration of the second oot mean squared error 0.1722 elative absolute error 10.0365 % DecisionStump oot relative squared error 36.4196 % ld3 otal Number of Instances 51 i48 🚺 ► == Detailed Accuracy By Class === lmt m5 P Rate FP Rate Precision Recall F-Measure Class RandomForest 0 Iris-setosa 1 1 1 1 0.9 0.947 0.923 0.947 0.063 Iris-versicolor RandomTree 0.882 0.029 0.938 0.882 0.909 Iris-virginica REPTree UserClassifier == Confusion Matrix === 间 rules <-- classified as a b C 15 0 0 | a = Iris-setosa 0 18 1 | b = Iris-versicolor 0 2 15 | c = Iris-virginica 4 1 Status

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Log

x 0

				1			
Classifier							
Choose							
Test options		Classifier out	tput				
 Use training set 		=== Evalua	ation on te	st split ==			\sim
O Supplied test set Set		=== Summan	су ===				
Cross-validation Folds 10		Correctly Incorrectl	Classified Ly Classifi	Instances ed Instance	s	48 3	94.1176 % 5.8824 %
Percentage split % 66		Kappa stat	istic ute error			0.9113	
More options	\supset	Root mean Relative a Root relat	squared er absolute er tive square	ror ror d error		0.1722 10.0365 % 36.4196 %	
(Nom) class	+	Total Numb	per of Inst	ances		51	
		=== Detail	led Accurac	y By Class	===		
Start Stop		TP Rate	FP Rate	Precision	Recall	F-Measure	Class
Result list (right-click for options)		1 0.947	0.063	1 0.9	1 0.947	1 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 - functions.neural. Jeural 14:48:05 - bayes NaiveBayes	letwork	=== Confus	sion Matrix				
		a b c 15 0 0 0 18 1 0 2 15	< clas a = Iri b = Iri c = Iri	sified as s-setosa s-versicolo s-virginica	er 1		
			+ + + + + +) 4 ► (
Status							

ОК



x 1

Log

Status

Building model on training data...



	Preprocess	Classify	Cluster	Associate	Select attributes	Visualiz	2		
Classifier									
Choose UserClassifier									
Test options	0								
O Use training set			Tree Vis	ualizer Dat	a Visualizer				
 Supplied test set Cross-validation 	X: petalle	ngth (Num)	• Y: p	etalwidth (Num)		•		
Percentage split	Colour: c	lass (Nom)		🗘 Poly	/line		•		
More	Submit	Clear	Sa	ve Jit	ter 😑		—		
(Nom) class	Plot: iris			Xxxx XX	××× 13				
Start	1.3-		× × , , , , , , , , , , , , , , , , , ,	Å.					
11:49:05 - trees.j48. 14:34:28 - functions	0.1	X X X X X	··· × × × · ××			ê. Bili. Ê			
14:48:05 - bayes.Na	1		3,95		6.9				
15.20.57 - (1995.036	-Class colour								
		Iris-setosa Iris-versicolor Iris-virginica							
1									
Status									
Building model on training data							LOG		





Select attributes Visualize

Classify Cluster Associate Preprocess Classifier UserClassifier Choose 00 Test options Data Visualizer Tree Visualizer \bigcirc Use training set Supplied test set ()÷ X: petallength (Num) ÷ Y: petalwidth (Num) Cross-validation ÷ Colour: class (Nom) + Polyline Percentage split lacksquareMore c Submit Clear Save Jitter 🤇 Plot: iris (Nom) class 2.5 Start 1.3-Result list (right-click for 11:49:05 - trees.j48. 14:34:28 - functions 0.1 14:48:05 - bayes.Nai 3.95 15:26:57 - trees.Use Class colour Iris-setosa Iris-versicolor Iris-virginica

Status Building model on training data...



x 1



)		
ciate	Select attributes	Visualize

	Preprocess	Classify	Cluster	Associate	Select attrib	utes	Visualize	
Classifier								
Choose UserCla	ssifier							
Test options	$\bigcirc \bigcirc \bigcirc$							
O Use training set			Tree Vis	sualizer D	ata Visualizer			
O Supplied test set	Tree View							
O Cross-validation	n							
Percentage split								
More	21							
(Nom) class	-			Split on petallength / petalwidt	ND			
Start				\wedge				
Result list (right-click for	·c			True E	alse			
11:49:05 - trees.j48	J			/	\ \			
14:34:28 - functions 14:48:05 - baves.Na	i.		[Iris-versio	volor 1 01	is-setosa 50.01	1		
15:26:57 - trees.Use	r		[Iris-virgin	ica, 48.0] [Iris	-versicolor, 49.0]			
				14	5 Virginiou, 2.01			
						_		

Status-

Building model on training data...



Dropro

Log

Treprocess classify	Cluster	Associat	e Select at	induces	Visualize	
Classifier						
Choose UserClassifier						
Test options	Classifier out	put				
O Use training set	=== Evalua	ation on t	est split ==			\sim
O Supplied test set Set	Correctly	Classific	d Instances		10	96 0794 \$
Cross-validation Folds 10	Incorrectl	y Classif	ied Instance	es	2	3.9216 %
Percentage split % 66	Kappa stat	istic			0.9408	
O relectinge spint in ou	Mean absol	ute error	rror		0.0319	
	Relative a	absolute e	error		7.1634 %	
	Root relat	ive squar	ed error		34.312 %	
QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.	Total Numb	per of Ins	stances		51	
	=== Detail	ed Accura	cy By Class			
	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	1	0 063	1	1	1	Iris-setosa
11:49:05 - trees i48 148	0.882	0.085	1	0.882	0.938	Iris-virginica
14:34:28 - functions neural NeuralNetwork						
14:48:05 – bayes NaiveBayes	=== Confus	sion Matri	x ===			
15:44:32 - trees.UserClassifier	abc	< cla	ssified as			
	15 0 0	a = I1	is-setosa			
	0 19 0 b = Iris-versicolor					
	0 2 15	c = I1	is-virginica	1		U
						-
	C	****) ()

OK

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

x 0

Log

Classifier			-				
Choose UserClassifier							
Test options	Classifier of	utput					5
O Use training set	=== Evalu	ation on t	est split ==				1
O Supplied test set Set	Compatib	. Glassifis	d Trotorooo		40	06 0704 %	
O Cross-validation Folds 10	Incorrectly	ly Classifie	ied Instances	es	2	3.9216 %	
Percentage split % 66	Kappa sta Mean abso	atistic plute error			0.9408		
QuickTime [™] and a TIFE (I ZW) decompressor are needed to see this purve	Root mear Relative Root rela Total Nur	absolute e absolute e ative squar mber of Ins	rror rror ed error tances		0.1622 7.1634 % 34.312 % 51		
	=== Detai	iled Accura	cy By Class			1	1
	TP Rate 1 1	FP Rate 0 0.063	Precision 1 0.905	Recall 1 1	F-Measure 1 0.95	Class Iris-setosa Iris-versicolor	
11:49:05 – trees.j48.J48	0.882	0	1	0.882	0.938	Iris-virginica	
14:34:28 - functions.neural.NeuralNetwork 14:48:05 - bayes.NaiveBayes	=== Confu	usion Matri	x ===				
15:44:32 – trees.UserClassifier	a b 0 15 0 0 0 19 0 0 2 15	c < cla) a = Ir) b = Ir 5 c = Ir	ssified as is-setosa is-versicolo is-virginica	or 1			
	<u> </u>				****)+ +]

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Log

		-	
Classifier			
Choose UserClassifier			
Test options	Classifier output		
🔘 Use training set	=== Evaluation on test split ===		<u></u>
O Supplied test set Set	=== Summary ===		
Cross-validation Folds 10	Correctly Classified Instances Incorrectly Classified Instances	49 2	96.0784 %
Percentage split % 66	Kappa statistic	0.9408	
	Root mean squared error	0.1622	
(Num) sepallength	Relative absolute error Root relative squared error	7.1634 % 34.312 %	
(Num) sepalwidth	Total Number of Instances	51	
(Num) petalwidth	=== Detailed Accuracy By Class ===		
(Nom) class	TR Pate FR Pate Provision Possil	E-Monguro	Class
Result list (right-click for options)	1 0 1 1	1 1	Iris-setosa
11:49:05 - trees i48 I48	1 0.063 0.905 1 0.882 0 1 0.882	0.95	Iris-versicolor Iris-virginica
14:34:28 – functions.neural.NeuralNetwork			
14:48:05 – bayes.NaiveBayes	=== Confusion Matrix ===		
15:44:32 - trees.UserClassifier	a b c < classified as		
	0 19 0 b = Iris-versicolor		
	0 2 15 c = Iris-virginica		U I
			•
Status			

ОК

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes

~		
1	Manuelles	
-	visualize	

x 0

Log

Classifier							7
Choose UserClassifier							
Test options	Classifier or	utput					5
 Use training set Supplied test set Set Cross-validation Folds 10 Percentage split % 66 	=== Evalu === Summa Correctly Incorrect Kappa sta Mean abso Root mear Relative Root rela Total Num	Ary === y Classifie tly Classifie atistic olute error h squared e absolute e ative squar mber of Ins	est split == d Instances ied Instance rror rror ed error tances	-=	49 2 0.9408 0.0319 0.1622 7.1634 % 34.312 % 51	96.0784 % 3.9216 %	1
Quick I ime I ^M and a TIFF (LZW) decompressor are needed to see this picture.	=== Detai TP Rate 1 1	iled Accura FP Rate 0 0.063	cy By Class Precision 1 0.905	Recall 1 1	F-Measure 1 0.95	Class Iris-setosa Iris-versicolor	
11:49:05 - trees.j48.J48	0.882	0	1	0.882	0.938	Iris-virginica	
14:34:28 - functions.neural.NeuralNetwork 14:48:05 - bayes.NaiveBayes 15:44:32 - trees.UserClassifier	=== Confu a b c 15 0 (0 19 (0 2 15	usion Matri c < cla) a = Ir) b = Ir 5 c = Ir	x === ssified as is-setosa is-versicolo is-virginica	or 1			•
Status							

-

	Preprocess	Classify	Cluster	Associat	e Select at	tributes	Visualize	
Classifier								
🗐 weka								
v 🖾 classifiers								
bayes			assifier ou	tput				
functions			= Evalu	ation on t	est split ==			6
▶ 🗿 lazy			== Summa:	ry ===	esc spire			
🕨 🧊 meta				0)				00.0304.0
🕨 🧊 misc			correctly	Classifie lv Classif	d Instances ied Instance	es	2	96.0784 % 3.9216 %
v 🗿 trees			appa stat	tistic			0.9408	
▶ 🗿 adtree			ean absol	lute error			0.0319	
Decision	Stump		elative a	squared e absolute e	rror		7.1634 %	
L Id3			pot relat	tive squar	ed error		34.312 %	
► 📜 i48			otal Num	per of Ins	tances		51	
▶ 📁 Imt			== Detail	led Accura	cv Bv Class			
v 🗐 m5					-1 -1			
MSP			? Rate	FP Rate	Precision	Recall	F-Measure	Class Tria cotoco
Random	Forest		1	0.063	0.905	1	0.95	Iris-setosa Iris-versicolor
Random	Tree		0.882	0	1	0.882	0.938	Iris-virginica
: REPTree			- Confin	alon Matul				
: UserCla	ssifier		== Conrus	sion Matri	x ===			
	5511101		a b c	< cla	ssified as			
P J Tuics			15 0 0	a = Ir	is-setosa			
			0 19 0	D = Ir c = Ir	is-versicold	a a		
				,		-		×
								v
				****	*****	* * * *	*****) ↓ ►
Status								



x 0

Weka Knowledge Explorer

Select attributes

Log



Status OK

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Log

Classifier	
Choose M5P -M 4.0	
Test options	Classifier output
	Class=iris=Virginica > 0.5 : LM3 (50/1/.598%)
Use training set	
O Supplied test set Set	LM num: 1 Linear Regression Model
O Cross-validation Folds 10	petallength =
Percentage split % 66	
	0.4957 * petalwidth +
More options	1.345
	LM num: 2
(Num) petallength	Linear Regression Model
	petallength =
Start Stop	Foodraad
	0.4208 * sepallength +
Result list (right-click for options)	1.2692 * petalwidth +
11:49:05 - trees.j48.J48	0.0755
14:34:28 - functions.neural.NeuralNetwork	LM num: 3
14:48:05 – bayes.NaiveBayes	Linear Regression Model
15:44:32 - trees.UserClassifier	petallength =
15:49:03 - trees.m5.M5P	
	0.7501 * sepallength +
	0.0103
	Number of Rules : 3
	▼

Status OK



OK

Weka Knowledge Explorer

Classify Cluster Associate Select attributes Visualize

Log 💉 🗴 0

	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize			
Classifier									
Choose M5P -M	1 4.0								
Test options			Classifier out	put					
Use training set				0.0 +	and have been a second s				
O Supplied test set Set			1.26	92 * petalwi	idth +				
O Cross-validation	n Folds 10		0.07	95					
Percentage split	% 66		LM num: 3 Linear Reg	ression Mode	21				
More	options		petallengt	:h =					
(Num) petallength	1	•	0.75	01 * sepalle .05	ength +				
Start) Stop		Number of	Rules : 3					
Result list (right-click for	r options)		Time taken to build model: 1.31 seconds						
11:49:05 - trees.j48	.J48	Network	=== Evaluation on test split ===						
14:34:28 - functions 14:48:05 - bayes.Na	s.neural.Neural liveBayes	INetwork	Correlatio	on coefficier	nt	0.9889			
15:44:32 - trees.Use	rClassifier		Mean absol	ute error		0.1861			
15:49:03 - trees.m5	.M5P		Root mean Relative a	squared erro bsolute erro	or or	0.255 11.9578 %			
			Root relat Total Numb	ive squared er of Instar	error	14.9153 % 51			
							* *		
Status									



Classify Cluster Associate Select attributes Visualize

100

	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
Choose M5P -M	1 4.0						
Test options			Classifier out	put			
🔘 Use training set		1					6
O Supplied test set	t Set		0.42	208 * sepalle 592 * petalwi	ength + .dth +		
O Cross-validation	n Folds 10		0.07	95			
Percentage split	% 66		LM num: 3 Linear Reg	ression Mode	21		
More	options		petallengt	:h =			
(Num) petallength	ו	•	0.75	01 * sepalle .05	ength +		
Start) Stop	,	Number of	Rules : 3			
Result list (right-click for	roptions) ——		Time taker	to build mo	del: 1.31 seconds		
11:49:05 - trees.j48	.J48		=== Evalua	ation on test	: split ===		
14:34:28 - functions	s.neural.Neural	Network	=== Summar	:y ===			
14:48:05 - bayes.Na	iveBayes		Correlatio	on coefficier	nt	0.9889	
15:44:32 - trees.Use	rClassifier		Mean absol Root mean	ute error	ar and a second s	0.1861	
15:49:05 - trees.ms.			Relative a	bsolute erro)r	11.9578 %	
			Root relat	ive squared	error	14.9153 %	
			Total Numb	er of Instar	ices	51	
							Ť
Status							LEI
OK						C	Log x 0



ОК

Log

🔈 x 0



ОК

Log

🔊 x 0



OK

Log

P. × 0

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 "true" clusters (if given)
- Evaluation based on loglikelihood if clustering scheme produces a probability distribution



Males K

000			weка кл	owledge Explo	orer		
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Clusterer							
Choose EM -I 1	00 -N -1 -S 1	.00 -M 1.0E-	-6				
Cluster mode			Cluste	erer output			
💿 Use training set							
O Supplied test set	: (5	Set					
🖯 Percentage split		% 66					
O Classes to clust	ers evaluation						
(Nom) class		A Y					
Store clusters fo	r visualizatior	n					
			_				
Igno	ore attributes						
Start	\supset	Stop	\supset				
Result list (right-click for	options)						





Weles Knowledge Explo

Preprocess Classify Cluster Associate Select attributes Visualize Clusterer Choose EM -I 100 -N -1 -S 100 -M 1.0E-6 Cluster mode Cluster mode Cluster routput Output Use training set Supplied test set Set Percentage split % 66 Classes to clusters evaluation (Nom) class \$\$ Store clusters for visualization Ignore attributes Start Stop Result list (right-click for options)	000		. We	eka Knowledge Expl	orer		
Clusterer Choose EM -I 100 -N -1 -S 100 -M 1.0E-6 Cluster mode Use training set Supplied test set Set Percentage split % 66 Classes to clusters evaluation (Nom) class \$ Store clusters for visualization Ignore attributes Start Stop Result list (right-click for options)		Preprocess	Classify Clu	uster Associate	Select attributes	Visualize	
Choose EM -I 100 -N -1 -S 100 -M 1.0E-6 Cluster mode Ouse training set Supplied test set Set Percentage split % 66 Classes to clusters evaluation (Nom) class \$ Store clusters for visualization Ignore attributes Start Stop Result list (right-click for options)	Clusterer						
Cluster mode Use training 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)	Choose EM -I 1	00 -N -1 -S 100) -M 1.0E-6				
 Use training set Supplied test set Set Percentage split 66 Classes to clusters evaluation (Nom) class Store clusters for visualization Ignore attributes Start Stop Result list (right-click for options) 	Cluster mode			Clusterer output			
Supplied test set Set Percentage split % 66 Classes to clusters evaluation (Nom) class (Nom) class ‡ Store clusters for visualization Ignore attributes Start Stop Result list (right-click for options) Image: Stop option (Stop)	Use training set						
 Percentage split % 66 Classes to clusters evaluation (Nom) class ÷ Store clusters for visualization Ignore attributes Start Stop Result list (right-click for options) 	O Supplied test set	t Set					
Classes to clusters evaluation (Nom) class Store clusters for visualization Ignore attributes Start Stop Result list (right-click for options)	O Percentage split		% 66				
(Nom) class + Store clusters for visualization Ignore attributes Start Stop Result list (right-click for options)	O Classes to clust	ers evaluation					
Store clusters for visualization Ignore attributes Start Stop Result list (right-click for options)	(Nom) class		A T				
Ignore attributes Start Stop Result list (right-click for options)	Store clusters fo	or visualization					
Start Stop Result list (right-click for options)	Igno	ore attributes					
Result list (right-click for options)	Start) (s	top				
	Result list (right-click for	r options)					



Preprocess Classify Cluster Associate Select attributes Visualize

Confere

_			Preprocess	Classify	Cluster	Associate	Select attributes	visualize	
Г	Clusterer								
	河 weka				7738781	.5			
L	🔻 🧊 clu	isterers			_				
Г		EM			Clust	erer output			
		SimpleKMea	ns						
		Cobweb							
		FarthestFirs	t		-				
		Xmeans							
									11111
									11111
									1111
L									
	(
									1111
	(
	-								
I	-								
									1111
									11111
	_	_		_					11111
					1				
[Status							6	
	OK							0	







	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize		
Clusterer								
Choose Cobwel	b -A 1.0 -C 0.	0028209479	917738781	5				
Cluster mode			Clust	erer output				
🔘 Use training set								n
O Supplied test set	t S	et						
🔘 Percentage split		% 66						
Classes to clust	ers evaluation							
(Nom) class		+						
Store clusters fo	or visualization	1						
Igno	ore attributes		\supset					
Start	\supset	Stop	\supset					
Result list (right-click for	r options)							
								v
			C		******			• •
Status								
OK						C	Log	P. X 0



Classify Cluster Associate Select attributes Visualize

	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize		
Clusterer								,]
Choose Cobwel	b -A 1.0 -C 0.	0028209479	917738781	5				
Cluster mode			Cluste	erer output				
Use training set							r	١
Supplied test set	t S	et						
Percentage split		% 66						
• Classes to clust	ers evaluation	_						
(Nom) class		÷						
Store clusters fo	or visualization	I						
Igno	ore attributes		\supset					
Start	\supset	Stop	\supset					
Result list (right-click for	otions) —							
	•							
								-
			C) + + (,
Status								
ОК						C	Log 🔪 💉 🗙	0

Select attributes Visualize

Cluster Preprocess Classify Associate Clusterer Cobweb -A 1.0 -C 0.0028209479177387815 Choose) Cluster mode Clusterer output- \bigcirc Use training set === Run information === Supplied test set Set.... weka.clusterers.Cobweb -A 1.0 -C 0.002820947917 Scheme: Relation: iris Percentage split % 66 150 Instances: Attributes: 5 Classes to clusters evaluation sepallength + (Nom) class sepalwidth petallength Store clusters for visualization petalwidth Ignored: class Classes to clusters evaluation on training data Test mode: Ignore attributes === Clustering model (full training set) === Stop Start Number of merges: 0 Result list (right-click for options) Number of splits: 0 Number of clusters: 3 16:05:58 - Cobweb node 0 [150] leaf 1 [96] node 0 [150] leaf 2 [54] === Evaluation on training set ===

Status OK



Select attributes

Visualize

Cluster Preprocess Classify Associate Clusterer Cobweb -A 1.0 -C 0.0028209479177387815 Choose Cluster mode Clusterer output- \bigcirc Use training set === Run information === Supplied test set Set.... weka.clusterers.Cobweb -A 1.0 -C 0.002820947917 Scheme: Relation: iris Percentage split % 66 150 Instances: Attributes: 5 Classes to clusters evaluation sepallength + (Nom) class sepalwidth petallength Store clusters for visualization petalwidth Ignored: class Ignore attributes Test mode: Classes to clusters evaluation on training data === Clustering model (full training set) === Stop Start Number of merges: 0 Result list (right-click for options) Number of splits: 0 Number of clusters: 3 16:05:58 - Cobweb node 0 [150] leaf 1 [96] node 0 [150] leaf 2 [54] === Evaluation on training set ===

Status OK





Associate

Cluster

Classify

Preprocess

Select attributes

Visualize

Clusterer		
Choose Cobweb -A 1.0 -C 0.00282094791773	387815	
Cluster mode	Clusterer output	
🔘 Use training set	Number of clusters: 3	<u> </u>
O Supplied test set Set	node 0 [150]	
O Percentage split % 66	node 0 [150]	
Classes to clusters evaluation	leaf 2 [54]	
(Nom) class	Clustered Instances	
Store clusters for visualization	1 100 (67%) 2 50 (33%)	
Ignore attributes	Class attribute: class Classes to Clusters:	
Start Stop	1 2 < assigned to cluster	
Result list (right-click for options)	0 50 Iris-setosa 50 0 Iris-versicolor	
10:05:58 - CODWED	50 0 Iris-virginica	
	Cluster 1 < Iris-versicolor Cluster 2 < Iris-setosa	
	Incorrectly clustered instances : 50.0 33.3333 %	Ų
		Y
Clabua		-
OK	Log	× x 0
UK		



Select attributes

Visualize



Select attributes

Visualize





Select attributes





Weka Knowledge Explorer

Cluster

Classify

Preprocess

Associate Select attributes Visualize

Log



OK

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:
 - ◆ 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
| $\Theta \Theta \Theta$ | | | Weka Kr | nowledge Expl | orer | | |
|------------------------------|---------------|--------------|--------------|---------------|-------------------|-----------|--------|
| | Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | |
| Associator | | | | | | | |
| Choose Apriori | -N 10 -T 0 -C | C 0.9 –D 0.0 | 5 -U 1.0 - | M 0.1 -S -1.0 | | | |
| Start | Stop | Ass | ociator outp | ut | | | |
| Result list (right-click for | r options) | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| Status | | | | | | | |
| OK | | | | | | C | Log X0 |

00		Weka Kr	nowledge Expl	orer		
	Preprocess Class	ify Cluster	Associate	Select attributes	Visualize	
Associator						
Choose Apriori	-N 10 -T 0 -C 0.9 -	0.05 -U 1.0 -	M 0.1 -S -1.0			
Start	Stop	Associator outp	ut			
Result list (right-click for	r options)					
Status						
ОК						og) x0



Log



Log

Status-

OK

$\Theta \Theta \Theta$			Weka Kr	nowledge Expl	orer		
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Associator							
Choose Apriori	-N 10 -T 0 -C	C 0.9 –D 0.0	5 -U 1.0 -	M 0.1 -S -1.0			
Start	Stop	Ass	ociator outp	ut			
Result list (right-click for	r options)						
Status							
OK						C	Log X0

00			Weka Kr	nowledge Expl	orer		
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
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	Ass	ociator outp	ut			
Result list (right-click fo	options)						
Status						6	Log x0
UN						<u> </u>	ALC: NOT



	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize)	
Associator								
Choose Apriori	-N 10 -T 0 -C	C 0.9 -D 0.0	5 -U 1.0 -	M 0.1 -S -1.0				
Start Stop Result list (right-click for 16:29:37 – Apriori	Associato Minimum Number Generat Size of Size of Size of Size of Best ru 1. ado 2. ado 3. phy 4. phy 5. phy 6. el- 7. el- 8 ado	routput metric <co of cycles p ed sets of set of lan set of lan set of lan set of lan set of lan les found: ption-of-th ption-of-th sician-fee- sician-fee- salvador-ai salvador-ai</co 	onfidence> performed: large ite rge itemse rge itemse	: 0.9 11 msets: ts L(1): 20 ts L(2): 17 ts L(2): 17 ts L(3): 6 ts L(4): 1 resolution=y aid-to-nicar education-sp 247 ==> Clas =democrat 20 => aid-to-ni	physician-fee-fre physician-fee-fre aguan-contras=y 21 ending=n 202 ==> 0 s=democrat 245 0 ==> aid-to-nicar caraguan-contras=y aid-to-nicaraguan	eeze=n 219 = eeze=n aid-t 1 ==> Class class=democr conf: (0.99) raguan-contr 204 con	=> Class=democra o-nicaraguan-con =democrat 210 at 201 conf: f:(0.98) Class=democrat 2	at ni (:
	9. el- 10. aid	salvador-ai -to-nicarao	id=n aid-t guan-contr	o-nicaraguan as=y Class=d	-contras=y 204 ==> emocrat 218 ==> ph	Class=demo nysician-fee	-freeze=n 210	f A V
Status								





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: WEKA allows (almost) arbitrary combinations of these two



	Weka Kn	owledge Expl	orer		
Classify	Cluster	Associate	Select attributes	Visualize	

Log

x 0

ocaron mounou	-S	e	а	rc	h	М	e	t	n	o	c
---------------	----	---	---	----	---	---	---	---	---	---	---

Attribute Evaluator

Choose	BestFirst -D 1 -N 5
Attribute Selec	tion Mode

Choose CfsSubsetEval

Preprocess











Select attributes Visualize





000

Weka Knowledge Explorer

Classify Select attributes Preprocess Cluster Associate Visualize Attribute Evaluator Choose CfsSubsetEval Search Methoo Choose BestFirst -D 1 -N 5 Attribute Selection Mode Attribute selection outputduty-free-exports Use full training set export-administration-act-south-africa Cross-validation Folds 10 Class Evaluation mode: evaluate on all training data Seed 1 + (Nom) Class === Attribute Selection on all input data === Start Stop Search Method: Best first. Result list (right-click for options) Start set: no attributes Search direction: forward 16:39:40 - BestFirst + CfsSubsetEval 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

OK



00

Weka Knowledge Explorer

Select attributes Classify Visualize Preprocess Cluster Associate Attribute Evaluator weka attributeSelection CfsSubsetEval ClassifierSubsetEval WrapperSubsetEval te selection output-ConsistencySubsetEval ReliefFAttributeEval duty-free-exports export-administration-act-south-africa InfoGainAttributeEval Class GainRatioAttributeEval evaluate on all training data ation mode: SymmetricalUncertAttributeEval OneRAttributeEval ChiSquaredAttributeEval Attribute Selection on all input data === PrincipalComponents ch Method: 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 bute Subset Evaluator (supervised, Class (nominal): 17 Class): CFS Subset Evaluator Selected attributes: 4 : 1 physician-fee-freeze

F



-	- I	~	~
- 1			
- 1		1	









-Status OK



Explorer: data visualization

- Visualization very useful in practice: e.g. helps to determine difficulty of the learning problem
- WEKA can visualize single attributes (1-d) and 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





	Preprocess	Classify	Cluster	Associate	Select attribut	es Visualize		
Plot Matrix	RI	Na		Mg	AI	Si	к	1
	·							١
Туре								1
Fe			•	4				
						1111111 111111		
		•	∥ •			•) 4 F	
PlotSize: [100]				_			A	
PointSize: [1]	0			_ (Update)		
Jitter:	0				elect Attributes)		
Colour: Type	e (Nom)			;	SubSample % :) 100		
Class Colour								
build wind	float build wind	non-float v	ehic wind	float vehic	wind non-float	containers tabl	eware headlamps	
Status								







x 0

Log



00

Weka Knowledge Explorer

x 0

Log



OK

00

Weka Knowledge Explorer

Select attributes

Visualize

Log

x 0



OK



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



build wind float	build wir	nd 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



Class colour -

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



Class colour -

build wind float vehic wind non-float build wind non-float

tableware

vehic wind float headlamp



2/22/2011

University of Waikato



2/22/2011

University of Waikato

Conclusion: try it yourself!

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, Bernhard Pfahringer, Brent Martin, Peter Flach, Eibe Frank, Gabi Schmidberger ,Ian H. Witten, J. Lindgren, Janice Boughton, Jason Wells, Len Trigg, Lucio de Souza Coelho, Malcolm Ware, Mark Hall, Remco Bouckaert, Richard Kirkby, Shane Butler, Shane Legg, Stuart Inglis, Sylvain Roy, Tony Voyle, Xin Xu, Yong Wang, Zhihai Wang