

1.

操作流程:

先打開檔案

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

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose **None** [Apply] [Stop]

Current relation: Relation: weather.symbolic, Instances: 14, Attributes: 5, Sum of weights: 14

Attributes: [All] [None] [Invert] [Pattern]

No.	Name
1	<input checked="" type="checkbox"/> outlook
2	<input checked="" type="checkbox"/> temperature
3	<input checked="" type="checkbox"/> humidity
4	<input checked="" type="checkbox"/> windy
5	<input checked="" type="checkbox"/> play

Selected attribute: Name: outlook, Missing: 0 (0%), Distinct: 3, Type: Nominal, Unique: 0 (0%)

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

Class: play (Nom) [Visualize All]

接著進 classify 選 J48 並在 cross-validation 輸入 10 再按 start

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) play [Start] [Stop]

Result list (right-click for options): 13:33:47 - trees.J48

Classifier output:

Number of Leaves : 5
Size of the tree : 8
Time taken to build model: 0 seconds

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

Metric	Value	Std. Error	Percentage
Correctly Classified Instances	7		50 %
Incorrectly Classified Instances	7		50 %
Kappa statistic	-0.0426		
Mean absolute error	0.4167		
Root mean squared error	0.5984		
Relative absolute error	87.5 %		
Root relative squared error	121.2987 %		
Total Number of Instances	14		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.556	0.600	0.625	0.556	0.588	-0.043	0.633	0.758	yes
	0.400	0.444	0.333	0.400	0.364	-0.043	0.633	0.457	no
Weighted Avg.	0.500	0.544	0.521	0.500	0.508	-0.043	0.633	0.650	

=== Confusion Matrix ===

a b <-- classified as

	a	b
5 4 a = yes		
3 2 b = no		

(1)

```
=== Confusion Matrix ===  
  
 a b  <-- classified as  
 5 4 | a = yes  
 3 2 | b = no
```

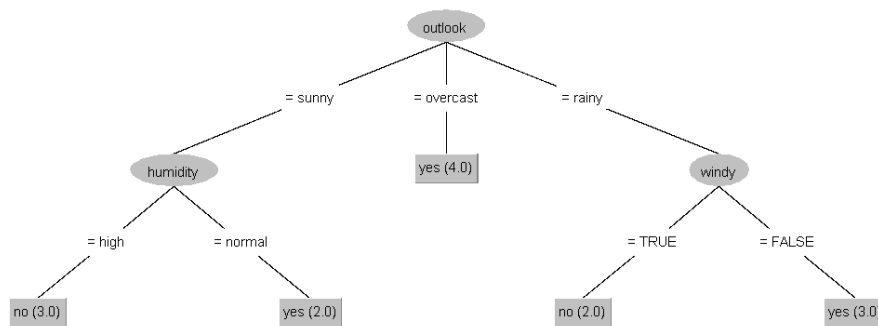
(2)

$$\text{Sensitivity} = \frac{5}{5+4} = 0.55$$

(3)

$$\text{Specificity} = \frac{2}{3+2} = 0.4$$

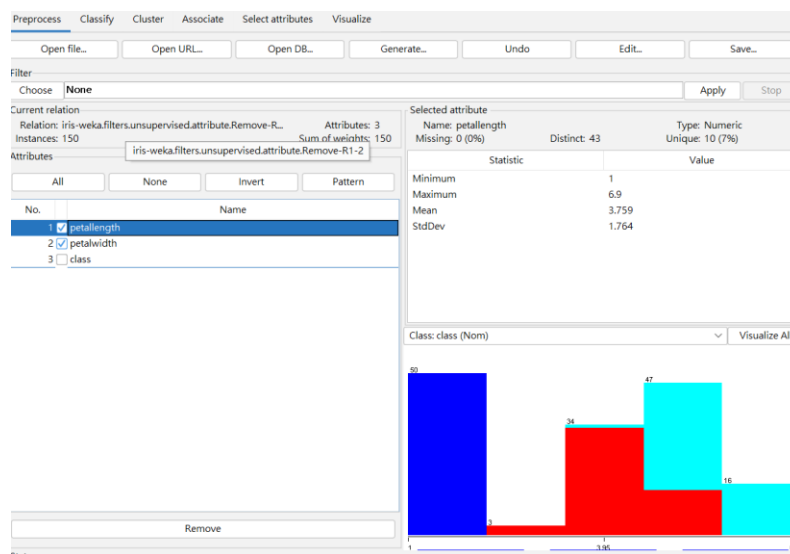
(4)



2.

操作流程

開檔案



接著把前 2 個 attribute 做 discretize (bin = 5)

Choose **Discretize -B 5 -M -1.0 -R first-last -precision 6** attributes in the dataset into nominal attributes. Capabilities

Current relation
Relation: iris-weka.filters.unsupervised.attribute.Remove-R...
Instances: 150

Attributes

attributeIndices first-last

binRangePrecision 6

bins 5

All None Invert

進 classify 選 NaiveBayes, cv 輸入 10, 接著按 start

Classifier

Choose **NaiveBayes**

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)

13:40:45 - bayes.NaiveBayes

Classifier output

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	142	94.6667 %
Incorrectly Classified Instances	8	5.3333 %
Kappa statistic	0.92	
Mean absolute error	0.0549	
Root mean squared error	0.1769	
Relative absolute error	12.3468 %	
Root relative squared error	37.5223 %	
Total Number of Instances	150	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	Iris-
0.920	0.040	0.920	0.920	0.920	0.920	0.880	0.977	0.957	Iris-
0.920	0.040	0.920	0.920	0.920	0.920	0.880	0.978	0.956	Iris-
Weighted Avg.	0.947	0.027	0.947	0.947	0.947	0.920	0.985	0.971	

=== Confusion Matrix ===

```

a b c <-- classified as
50 0 0 | a = Iris-setosa
0 46 4 | b = Iris-versicolor
0 4 46 | c = Iris-virginica

```

(1)

```

=== Confusion Matrix ===

a b c <-- classified as
50 0 0 | a = Iris-setosa
0 46 4 | b = Iris-versicolor
0 4 46 | c = Iris-virginica

```

(2)

Sensitivity :

$$\text{Setosa} = \frac{50}{50+0+0} = 1$$

$$\text{versicolor} = \frac{46}{0+46+4} = 0.92$$

$$\text{virginica} = \frac{46}{0+4+46} = 0.92$$

(3)

Specificity:

$$\text{Setosa} = \frac{46+46}{50+50} = 0.92$$

$$\text{versicolour} = \frac{50+46}{50+50} = 0.96$$

$$\text{virginica} = \frac{50+46}{50+50} = 0.96$$

3.

操作過程

開檔案

The screenshot shows the Weka software interface with the 'Preprocess' tab selected. The 'Filter' dropdown is set to 'Discretize - B 5 - M -1.0 -R first-last -precision 6'. The 'Current relation' is 'iris' with 150 instances and 5 attributes. The 'Attributes' list on the left shows 'sepalength', 'sepalwidth', 'petalength', 'petalwidth', and 'class', all of which are checked. The 'Selected attribute' panel on the right shows statistics for 'sepalength': Minimum (4.3), Maximum (7.9), Mean (5.843), and StdDev (0.828). Below this, a histogram shows the distribution of 'sepalength' values, with the x-axis ranging from 4.0 to 7.0 and the y-axis showing counts (16, 30, 34, 28, 25, 10, 7). The histogram bars are colored red and blue.

進 classify 選 IBk

The screenshot shows the 'Classifier' tab in Weka. A tree view on the left lists various classifiers under the 'weka' folder. The 'IBk' classifier is selected and highlighted in blue. Other visible classifiers include 'KStar', 'LWL', 'meta', 'misc', 'rules', and 'trees'.

進 IBk 調參數(k = 3, distance = Euclidean distance)

weka.classifiers.lazy.IBk

About

K-nearest neighbours classifier.

More

Capabilities

KNN 3

batchSize 100

crossValidate False

debug False

distanceWeighting Weight by 1/distance

doNotCheckCapabilities False

meanSquared False

nearestNeighbourSearchAlgorithm Choose LinearNNSearch -A "weka.core.EuclideanDistance"

numDecimalPlaces 2

windowSize 0

Cv 輸入 10 後, 按 start

Classifier

Choose IBk -K 3 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last""

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

13:46:19 - lazy.IBk

Classifier output

using 3 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	143	95.3333 %
Incorrectly Classified Instances	7	4.6667 %
Kappa statistic	0.93	
Mean absolute error	0.0357	
Root mean squared error	0.1673	
Relative absolute error	8.0416 %	
Root relative squared error	35.4899 %	
Total Number of Instances	150	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Iris-se
	0.940	0.040	0.922	0.940	0.931	0.896	0.981	0.972	Iris-ve
	0.920	0.030	0.939	0.920	0.929	0.895	0.981	0.958	Iris-vi
Weighted Avg.	0.953	0.023	0.953	0.953	0.953	0.930	0.987	0.977	

=== Confusion Matrix ===

a	b	c	<-- classified as
50	0	0	a = Iris-setosa
0	47	3	b = Iris-versicolor
0	4	46	c = Iris-virginica

(1)

```
=== Confusion Matrix ===  
  
  a  b  c  <-- classified as  
50  0  0 | a = Iris-setosa  
 0 45  5 | b = Iris-versicolor  
 0  4 46 | c = Iris-virginica
```

(2)

Sensitivity:

$$\text{Setosa} = \frac{50}{50+0+0} = 1$$

$$\text{versicolour} = \frac{45}{0+45+5} = 0.90$$

$$\text{virginica} = \frac{46}{0+4+46} = 0.92$$

(3)

Specificity:

$$\text{Setosa} = \frac{45+46}{50+50} = 0.91$$

$$\text{versicolour} = \frac{50+46}{50+50} = 0.96$$

$$\text{virginica} = \frac{50+45}{50+50} = 0.95$$