

## Assignment 7 Logic-based approaches

HAITAO ZHOU (HAZ59)

YINGZHI YANG(YIY50)

XIN JIN(XIJ21)

### 1.Iris

#### (1) Introduction of data

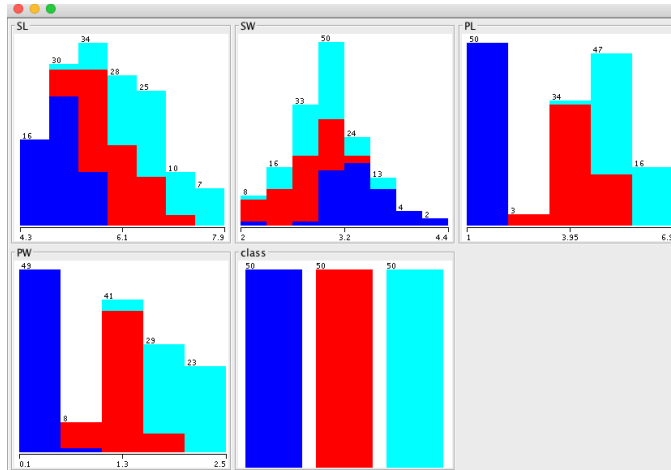
The number of instances is 150. The four attributes SL, SW, PL, PW refer to sepal length in cm, sepal width in cm, petal length in cm and petal width in cm and there are three classes: Iris Setosa(blue), Iris Versicolour(red), Iris Virginica(green).

(According to the data sets of Irvine Machine Learning Repository, the 35th sample should be:

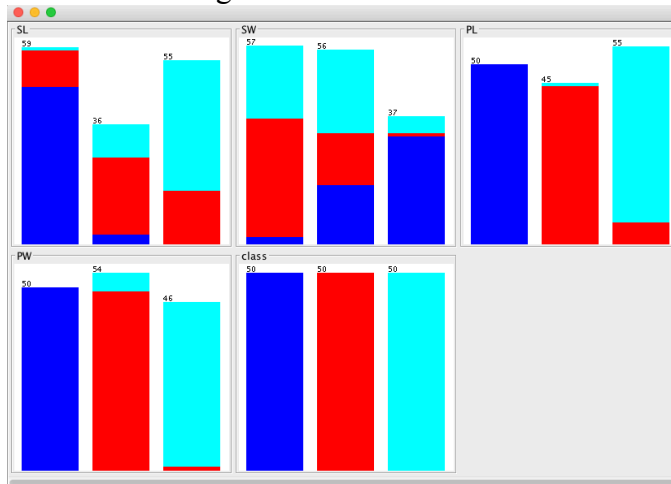
4.9,3.1,1.5,0.2,"Iris-setosa" where the error is in the fourth feature. The 38th sample:

4.9,3.6,1.4,0.1,"Iris-setosa" where the errors are in the second and third features.)

Initial data:



After discretizing:



#### (2) Trees Model

## LMT(default): Accuracy 94%

```
Logistic model tree
-----
: LM_1:1/1 (150)

Number of Leaves :    1

Size of the Tree :    1
LM_1:
Class 0 :
-1 +
[PL='(-inf-2.45)'] * 3

Class 1 :
-0.97 +
[PW='(0.8-1.75)'] * 2.69

Class 2 :
-0.86 +
[PW='(1.75-inf)'] * 2.79

Time taken to build model: 0.05 seconds

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

Correctly Classified Instances      141           94      %
Incorrectly Classified Instances     9           6      %
Kappa statistic                     0.91
Mean absolute error                  0.0771
Root mean squared error              0.1825
Relative absolute error              17.3441 %
Root relative squared error          38.7141 %
Total Number of Instances           150
```

Set ‘minNumInstances’ to 40 and ‘numBoostingIterations’ to 1, we can get the highest accuracy, which is 96%.

```
Logistic model tree
-----
: LM_1:1/1 (150)

Number of Leaves :    1

Size of the Tree :    1
LM_1:
Class 0 :
-1 +
[PL='(-inf-2.45)'] * 3

Class 1 :
-0.97 +
[PW='(0.8-1.75)'] * 2.69

Class 2 :
-0.86 +
[PW='(1.75-inf)'] * 2.79

Time taken to build model: 0.01 seconds

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

Correctly Classified Instances      144           96      %
Incorrectly Classified Instances     6           4      %
Kappa statistic                     0.94
Mean absolute error                  0.0993
Root mean squared error              0.1764
Relative absolute error              22.3352 %
Root relative squared error          37.4114 %
Total Number of Instances           150
```

**J48:** Accuracy 94%, unchanged when parameters are varied.

```
=== Classifier model (full training set) ===
```

```
J48 pruned tree
```

```
-----
```

```
PW = '(-inf-0.8]': Iris_setosa (50.0)
PW = '(0.8-1.75]': Iris_versicolor (54.0/5.0)
PW = '(1.75-inf)': Iris_virginica (46.0/1.0)
```

```
Number of Leaves :    3
```

```
Size of the tree :    4
```

```
Time taken to build model: 0 seconds
```

```
=== Stratified cross-validation ===
```

```
=== Summary ===
```

Correctly Classified Instances	141	94	%
Incorrectly Classified Instances	9	6	%
Kappa statistic	0.91		
Mean absolute error	0.0598		
Root mean squared error	0.193		
Relative absolute error	13.4523 %		
Root relative squared error	40.9465 %		
Total Number of Instances	150		

**SimpleCart:** Accuracy 94%, decreased when parameters change.

```
=== Classifier model (full training set) ===
```

```
CART Decision Tree
```

```
PL=('(4.75-inf)|'(2.45-4.75]')
| PW=('(1.75-inf)': Iris_virginica(45.0/1.0)
| PW!=('(1.75-inf)': Iris_versicolor(49.0/5.0)
PL!=('(4.75-inf)|'(2.45-4.75]': Iris_setosa(50.0/0.0)
```

```
Number of Leaf Nodes: 3
```

```
Size of the Tree: 5
```

```
Time taken to build model: 0.02 seconds
```

```
=== Stratified cross-validation ===
```

```
=== Summary ===
```

Correctly Classified Instances	141	94	%
Incorrectly Classified Instances	9	6	%
Kappa statistic	0.91		
Mean absolute error	0.0569		
Root mean squared error	0.1874		
Relative absolute error	12.8082 %		
Root relative squared error	39.7583 %		
Total Number of Instances	150		

**BFTree:** Accuracy 96.67% when 'minNumObj' is 3 and 'numFoldsPruning' is 2.

```

=== Classifier model (full training set) ===

Best-First Decision Tree

PL=('(4.75-inf)')|('(2.45-4.75]')
| PW=('(1.75-inf)')
| | SL=('(6.15-inf)')|('(inf-5.55]'): Iris_virginica(37.0/0.0)
| | SL=('(6.15-inf)')|('(inf-5.55]')
| | | SW=('(2.95-3.35]'): Iris_virginica(3.0/1.0)
| | | SW=('(2.95-3.35]'): Iris_virginica(5.0/0.0)
| | PW=('(1.75-inf)')
| | PL=('(4.75-inf)')
| | | SL=('(6.15-inf)')|('(inf-5.55]'): Iris_versicolor(4.0/2.0)
| | | SL=('(6.15-inf)')|('(inf-5.55]'): Iris_virginica(2.0/1.0)
| | PL=('(4.75-inf)'): Iris_versicolor(44.0/1.0)
PL=('(4.75-inf)')|('(2.45-4.75]'): Iris_setosa(50.0/0.0)

Size of the Tree: 13

Number of Leaf Nodes: 7

Time taken to build model: 0.02 seconds

```

```

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      145          96.6667 %
Incorrectly Classified Instances     5           3.3333 %
Kappa statistic                     0.95
Mean absolute error                  0.0331
Root mean squared error              0.1287
Relative absolute error              7.4556 %
Root relative squared error          27.3049 %
Total Number of Instances           150

```

**FT:** Accuracy 95.33% increased to 96% when ‘numBoostingIterations’ is set to 3.

```

=== Classifier model (full training set) ===

FT tree
-----
: FT_1:3/3 (150)

Number of Leaves :      1

Size of the Tree :      1
FT_1:
Class 0 :
-2.39 +
[PL='(-inf-2.45]'] * 5.81

Class 1 :
-1.96 +
[SW='(2.95-3.35]'] * 0.6 +
[PL='(2.45-4.75]'] * 1.28 +
[PW='(0.8-1.75]'] * 2.69

Class 2 :
-1.68 +
[SW='(-inf-2.95]'] * 0.81 +
[PL='(4.75-inf)'] * 1.35 +
[PW='(1.75-inf)'] * 2.79

Time taken to build model: 0.01 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      144          96      %
Incorrectly Classified Instances     6           4      %
Kappa statistic                     0.94
Mean absolute error                  0.0443
Root mean squared error              0.1375
Relative absolute error              9.9689 %
Root relative squared error          29.1599 %
Total Number of Instances           150

```

### (3) Rules Model

#### Decision Table: Accuracy 94%

=== Classifier model (full training set) ===

Decision Table:

Number of training instances: 150  
Number of Rules : 9  
Non matches covered by Majority class.  
Best first.  
Start set: no attributes  
Search direction: forward  
Stale search after 5 node expansions  
Total number of subsets evaluated: 17  
Merit of best subset found: 96.667  
Evaluation (for feature selection): CV (leave one out)  
Feature set: 1,3,4,5

Time taken to build model: 0 seconds

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

Correctly Classified Instances	141	94	%
Incorrectly Classified Instances	9	6	%
Kappa statistic	0.91		
Mean absolute error	0.0883		
Root mean squared error	0.1902		
Relative absolute error	19.8614 %		
Root relative squared error	40.3402 %		
Total Number of Instances	150		

#### Ridor: Accuracy 92%. Increased to 96% when 'folds' is 20 and 'minNo' is 16.

=== Classifier model (full training set) ===

Ripple Down Rule Learner(Ridor) rules

```
class = Iris_versicolor (150.0/100.0)
  Except (PL = '(-inf-2.45]') => class = Iris_setosa (48.0/0.0) [2.0/0.0]
  Except (PW = '(1.75-inf)') and (SW = '(-inf-2.95]') => class = Iris_virginica (16.0/0.0) [1.0/0.0]
  Except (PW = '(1.75-inf)') => class = Iris_virginica (27.0/1.0) [2.0/0.0]
```

Total number of rules (incl. the default rule): 4

Time taken to build model: 0 seconds

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

Correctly Classified Instances	144	96	%
Incorrectly Classified Instances	6	4	%
Kappa statistic	0.94		
Mean absolute error	0.0267		
Root mean squared error	0.1633		
Relative absolute error	6	%	
Root relative squared error	34.641	%	
Total Number of Instances	150		

### (3) Conclusion

The best approach is BFTree, in which the accuracy is 96.67%. Based on our observations, the classification of iris flowers is mainly related to PW and PL.

When  $PW < 0.8$ , the flower probably is Iris Setosa,

PW between 0.8 and 1.75, Iris Versicolour

PW  $> 1.75$  Iris Virginica

And when  $PL < 2.45$ , Iris Setosa

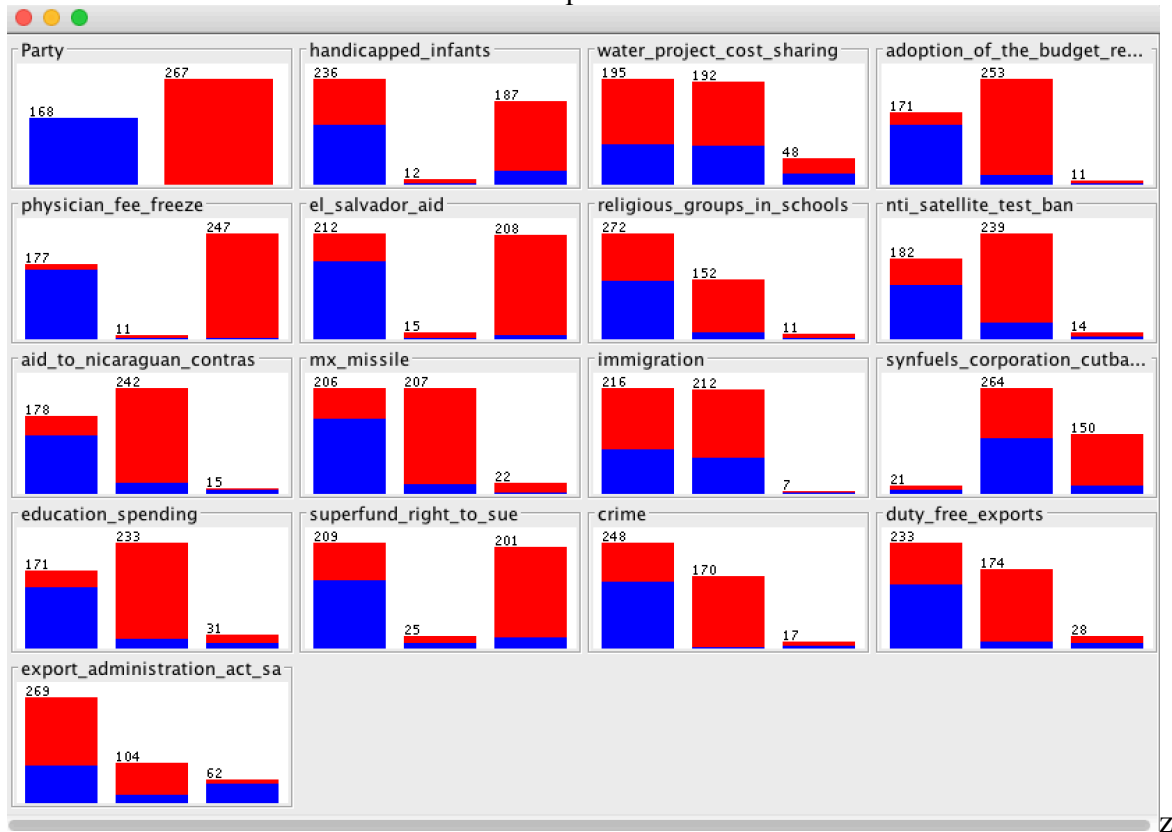
PL between 2.45 and 4.75, Iris Versicolour

PL  $> 4.75$ , Iris Virginica.

## 2. Congressional Voting Records

### (1) Introduction of data

Data set before and after discretization are quite same. The overview chart is as followed:



### (2) Tree Model

J48: Accuracy 94.9425%, increased to 95.6322% when 'minNumObj' is set to 6.

=== Classifier model (full training set) ===

J48 pruned tree

```
physician_fee_freeze = y
| synfuels_corporation_cutback = w: republican (7.0)
| synfuels_corporation_cutback = n: republican (138.0/3.0)
| synfuels_corporation_cutback = y
| | mx_missile = n
| | | adoption_of_the_budget_resolution = n: republican (21.0/3.0)
| | | adoption_of_the_budget_resolution = y
| | | | water_project_cost_sharing = y: democrat (4.0)
| | | | water_project_cost_sharing = n: republican (2.0)
| | | | water_project_cost_sharing = w: democrat (0.0)
| | | adoption_of_the_budget_resolution = w: republican (0.0)
| | mx_missile = y: democrat (5.0/1.0)
| | mx_missile = w: republican (0.0)
physician_fee_freeze = w: democrat (11.0/3.0)
physician_fee_freeze = n: democrat (247.0/2.0)
```

Number of Leaves : 11

Size of the tree : 16

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	413	94.9425 %
Incorrectly Classified Instances	22	5.0575 %
Kappa statistic	0.894	
Mean absolute error	0.068	
Root mean squared error	0.2051	
Relative absolute error	14.3367 %	
Root relative squared error	42.1278 %	
Total Number of Instances	435	

**LMT: Accuracy 96.7816%, getting worse when changing value**

Logistic model tree

: LM\_1:20/20 (435)

Number of Leaves : 1

Size of the Tree : 1

LM\_1:

Class 0 :

0.31 +

[water\_project\_cost\_sharing=y] \* -0.51 +  
[adoption\_of\_the\_budget\_resolution=n] \* 0.38 +  
[adoption\_of\_the\_budget\_resolution=y] \* -0.85 +  
[physician\_fee\_freeze=y] \* 1.8 +  
[physician\_fee\_freeze=n] \* -1.58 +  
[nti\_satellite\_test\_ban=n] \* -0.38 +  
[nti\_satellite\_test\_ban=y] \* 0.32 +  
[aid\_to\_nicaraguan\_contras=w] \* 0.44 +  
[mx\_missile=y] \* -0.54 +  
[immigration=n] \* -0.59 +  
[synfuels\_corporation\_cutback=n] \* 0.71 +  
[synfuels\_corporation\_cutback=y] \* -0.8 +  
[education\_spending=n] \* -0.58 +  
[education\_spending=w] \* 0.35 +  
[superfund\_right\_to\_sue=w] \* -0.46 +  
[duty\_free\_exports=y] \* -0.41 +  
[export\_administration\_act\_sa=y] \* 0.41

```

Class 1 :
-0.31 +
[water_project_cost_sharing=y] * 0.51 +
[adoption_of_the_budget_resolution=n] * -0.38 +
[adoption_of_the_budget_resolution=y] * 0.85 +
[physician_fee_freeze=y] * -1.8 +
[physician_fee_freeze=n] * 1.58 +
[nti_satellite_test_ban=n] * 0.38 +
[nti_satellite_test_ban=y] * -0.32 +
[aid_to_nicaraguan_contras=w] * -0.44 +
[mx_missile=y] * 0.54 +
[immigration=n] * 0.59 +
[synfuels_corporation_cutback=n] * -0.71 +
[synfuels_corporation_cutback=y] * 0.8 +
[education_spending=n] * 0.58 +
[education_spending=w] * -0.35 +
[superfund_right_to_sue=w] * 0.46 +
[duty_free_exports=y] * 0.41 +
[export_administration_act_sa=y] * -0.41

```

Time taken to build model: 0.48 seconds

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

Correctly Classified Instances	421	96.7816 %
Incorrectly Classified Instances	14	3.2184 %
Kappa statistic	0.9324	
Mean absolute error	0.0556	
Root mean squared error	0.1698	
Relative absolute error	11.7247 %	
Root relative squared error	34.8824 %	
Total Number of Instances	435	

**SimpleCart:** Accuracy 95.6322%, no difference after changing value



=== Classifier model (full training set) ===

CART Decision Tree

```
physician_fee_freeze=(n)|(w)
| adoption_of_the_budget_resolution=(y)|(n): democrat(247.0/2.0)
| adoption_of_the_budget_resolution!=(y)|(n)
| | mx_missile=(n)|(y): democrat(6.0/1.0)
| | mx_missile!=(n)|(y): republican(2.0/0.0)
physician_fee_freeze!=(n)|(w)
| synfuels_corporation_cutback=(y)
| | adoption_of_the_budget_resolution=(w)|(y)
| | | nti_satellite_test_ban=(n)|(w): democrat(6.0/0.0)
| | | nti_satellite_test_ban!=(n)|(w): republican(3.0/0.0)
| | adoption_of_the_budget_resolution!=(w)|(y)
| | | el_salvador_aid=(n): democrat(2.0/0.0)
| | | el_salvador_aid!=(n): republican(18.0/3.0)
| synfuels_corporation_cutback!=(y): republican(142.0/3.0)
```

Number of Leaf Nodes: 8

Size of the Tree: 15

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	416	95.6322 %
Incorrectly Classified Instances	19	4.3678 %
Kappa statistic	0.9088	
Mean absolute error	0.0801	
Root mean squared error	0.2027	
Relative absolute error	16.8954 %	
Root relative squared error	41.6347 %	
Total Number of Instances	435	

**BFTree:** Accuracy is 95.6322% when 'minNumObj' is set to 5.

=== Classifier model (full training set) ===

Best-First Decision Tree

```
physician_fee_freeze=(n)|(w)
| adoption_of_the_budget_resolution=(y)|(n)
| | adoption_of_the_budget_resolution=(y)|(w): democrat(224.0/0.0)
| | adoption_of_the_budget_resolution!=(y)|(w): democrat(23.0/2.0)
| adoption_of_the_budget_resolution!=(y)|(n): democrat(6.0/3.0)
physician_fee_freeze!=(n)|(w)
| synfuels_corporation_cutback=(y)
| | adoption_of_the_budget_resolution=(w)|(y): democrat(6.0/3.0)
| | adoption_of_the_budget_resolution!=(w)|(y): republican(18.0/5.0)
| synfuels_corporation_cutback!=(y)
| | duty_free_exports=(y): republican(11.0/2.0)
| | duty_free_exports!=(y)
| | | adoption_of_the_budget_resolution=(y): republican(13.0/1.0)
| | | adoption_of_the_budget_resolution!=(y): republican(118.0/0.0)
```

Size of the Tree: 15

Number of Leaf Nodes: 8

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	416	95.6322 %
Incorrectly Classified Instances	19	4.3678 %
Kappa statistic	0.9084	
Mean absolute error	0.0765	
Root mean squared error	0.2003	
Relative absolute error	16.1375 %	
Root relative squared error	41.1418 %	
Total Number of Instances	435	

**FT:** Accuracy 96.7816%, getting worse after changing parameters.

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

Correctly Classified Instances	421	96.7816 %
Incorrectly Classified Instances	14	3.2184 %
Kappa statistic	0.9323	
Mean absolute error	0.0395	
Root mean squared error	0.1731	
Relative absolute error	8.3238 %	
Root relative squared error	35.5574 %	
Total Number of Instances	435	

### (3) Rules Model

**Ridor:** Accuracy reaches 95.6322% when 'folds' is 9 and 'minNo' is 10.

---

```
=== Classifier model (full training set) ===
```

```
Ripple Down Rule Learner(Ridor) rules
```

```
Party = democrat (435.0/168.0)  
  Except (physician_fee_freeze = y) and (synfuels_corporation_cutback = n) => Party = republican (119.0/2.0) [19.0/1.0]  
  Except (physician_fee_freeze = y) and (immigration = y) => Party = republican (17.0/3.0) [1.0/0.0]  
  Except (physician_fee_freeze = y) => Party = republican (20.0/8.0) [1.0/0.0]
```

```
Total number of rules (incl. the default rule): 4
```

```
Time taken to build model: 0 seconds
```

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

Correctly Classified Instances	416	95.6322 %
Incorrectly Classified Instances	19	4.3678 %
Kappa statistic	0.9082	
Mean absolute error	0.0437	
Root mean squared error	0.209	
Relative absolute error	9.21 %	
Root relative squared error	42.9237 %	
Total Number of Instances	435	

**ConjunctiveRule:** Accuracy 95.6322% when 'folds' is set to 10.

=== Classifier model (full training set) ===

Single conjunctive rule learner:

(physician\_fee\_freeze = y) => Party = republican

Class distributions:

Covered by the rule:

republican	democrat
0.929936	0.070064

Not covered by the rule:

republican	democrat
0.021277	0.978723

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	416	95.6322 %
Incorrectly Classified Instances	19	4.3678 %
Kappa statistic	0.9088	
Mean absolute error	0.0828	
Root mean squared error	0.2036	
Relative absolute error	17.4641 %	
Root relative squared error	41.8232 %	
Total Number of Instances	435	

#### (4) Conclusion

The best approaches are LMT and FT, in which the accuracy is 96.7816%. We can tell from the above analysis that the voting result has the most to do with ‘physician-fee-freeze’ and ‘adoption-of-the-budget-resolution’. When the former is ‘n’ or the latter is ‘y’, the result is highly likely to be democrat.