第二次编程作业

1. 数据集描述

1. Title: 1984 United States Congressional Voting Records Database

2. Source Information:

(a) Source: Congressional Quarterly Almanac, 98th Congress,

2nd session 1984, Volume XL: Congressional Quarterly Inc.

Washington, D.C., 1985.

(b) Donor: Jeff Schlimmer (Jeffrey.Schlimmer@a.gp.cs.cmu.edu)

(c) Date: 27 April 1987

3. Past Usage

- Publications

1. Schlimmer, J. C. (1987). Concept acquisition through

representational adjustment. Doctoral dissertation, Department of

Information and Computer Science, University of California, Irvine, CA.

-- Results: about 90%-95% accuracy appears to be STAGGER's asymptote

- Predicted attribute: party affiliation (2 classes)

4. Relevant Information:

This data set includes votes for each of the U.S. House of

Representatives Congressmen on the 16 key votes identified by the

CQA. The CQA lists nine different types of votes: voted for, paired

for, and announced for (these three simplified to yea), voted

against, paired against, and announced against (these three

simplified to nay), voted present, voted present to avoid conflict

of interest, and did not vote or otherwise make a position known

(these three simplified to an unknown disposition).

5. Number of Instances: 435 (267 democrats, 168 republicans)

6. Number of Attributes: 16 + class name = 17 (all Boolean valued)

7. Attribute Information:

1. Class Name: 2 (democrat, republican)

2. handicapped-infants: 2 (y,n)

3. water-project-cost-sharing: 2 (y,n)

4. adoption-of-the-budget-resolution: 2 (y,n)

5. physician-fee-freeze: 2 (y,n)

6. el-salvador-aid: 2 (y,n)

7. religious-groups-in-schools: 2 (y,n)

8. anti-satellite-test-ban: 2 (y,n)

9. aid-to-nicaraguan-contras: 2 (y,n)

10. mx-missile: 2 (y,n)

11. immigration: 2 (y,n)

12. synfuels-corporation-cutback: 2 (y,n)

13. education-spending: 2 (y,n)

14. superfund-right-to-sue: 2 (y,n)

15. crime: 2 (y,n)

16. duty-free-exports: 2 (y,n)

17. export-administration-act-south-africa: 2 (y,n)

8. Missing Attribute Values: Denoted by "?"

NOTE: It is important to recognize that "?" in this database does

not mean that the value of the attribute is unknown. It

means simply, that the value is not "yea" or "nay" (see

"Relevant Information" section above).

Attribute: #Missing Values:

1: 0

2: 0

3: 12

4: 48

5: 11

6: 11

7: 15

8: 11

9: 14

10: 15

11: 22

12: 7

13: 21

14: 31

15: 25

16: 17

17: 28

9. Class Distribution: (2 classes)

1. 45.2 percent are democrat

2. 54.8 percent are republican

Class predictiveness and predictability: Pr(C|A=V) and Pr(A=V|C)

Attribute 1: (A = handicapped-infants)

0.91; 1.21 (C=democrat; V=y)

0.09; 0.10 (C=republican; V=y)

0.43; 0.38 (C=democrat; V=n)

0.57; 0.41 (C=republican; V=n)

0.75; 0.03 (C=democrat; V=?)

0.25; 0.01 (C=republican; V=?)

Attribute 2: (A = water-project-cost-sharing)

0.62; 0.45 (C=democrat; V=y)

0.38; 0.23 (C=republican; V=y)

0.62; 0.45 (C=democrat; V=n)

0.38; 0.23 (C=republican; V=n)

0.58; 0.10 (C=democrat; V=?)

0.42; 0.06 (C=republican; V=?)

Attribute 3: (A = adoption-of-the-budget-resolution)

0.91; 0.87 (C=democrat; V=y)

0.09; 0.07 (C=republican; V=y)

0.17; 0.11 (C=democrat; V=n)

0.83; 0.44 (C=republican; V=n)

0.64; 0.03 (C=democrat; V=?)

0.36; 0.01 (C=republican; V=?)

Attribute 4: (A = physician-fee-freeze)

0.08; 0.05 (C=democrat; V=y)

0.92; 0.50 (C=republican; V=y)

0.99; 0.92 (C=democrat; V=n)

0.01; 0.01 (C=republican; V=n)

0.73; 0.03 (C=democrat; V=?)

0.27; 0.01 (C=republican; V=?)

Attribute 5: (A = el-salvador-aid)

0.26; 0.21 (C=democrat; V=y)

0.74; 0.48 (C=republican; V=y)

0.96; 0.75 (C=democrat; V=n)

0.04; 0.02 (C=republican; V=n)

0.80; 0.04 (C=democrat; V=?)

0.20; 0.01 (C=republican; V=?)

Attribute 6: (A = religious-groups-in-schools)

0.45; 0.46 (C=democrat; V=y)

0.55; 0.46 (C=republican; V=y)

0.89; 0.51 (C=democrat; V=n)

0.11; 0.05 (C=republican; V=n)

0.82; 0.03 (C=democrat; V=?)

0.18; 0.01 (C=republican; V=?)

Attribute 7: (A = anti-satellite-test-ban)

0.84; 0.75 (C=democrat; V=y)

0.16; 0.12 (C=republican; V=y)

0.32; 0.22 (C=democrat; V=n)

0.68; 0.38 (C=republican; V=n)

0.57; 0.03 (C=democrat; V=?)

0.43; 0.02 (C=republican; V=?)

Attribute 8: (A = aid-to-nicaraguan-contras)

0.90; 0.82 (C=democrat; V=y)

0.10; 0.07 (C=republican; V=y)

0.25; 0.17 (C=democrat; V=n)

0.75; 0.41 (C=republican; V=n)

0.27; 0.01 (C=democrat; V=?)

0.73; 0.03 (C=republican; V=?)

Attribute 9: (A = mx-missile)

0.91; 0.70 (C=democrat; V=y)

0.09; 0.06 (C=republican; V=y)

0.29; 0.22 (C=democrat; V=n)

0.71; 0.45 (C=republican; V=n)

0.86; 0.07 (C=democrat; V=?)

0.14; 0.01 (C=republican; V=?)

Attribute 10: (A = immigration)

0.57; 0.46 (C=democrat; V=y)

0.43; 0.28 (C=republican; V=y)

0.66; 0.52 (C=democrat; V=n)

0.34; 0.23 (C=republican; V=n)

0.57; 0.01 (C=democrat; V=?)

0.43; 0.01 (C=republican; V=?)

Attribute 11: (A = synfuels-corporation-cutback)

0.86; 0.48 (C=democrat; V=y)

0.14; 0.06 (C=republican; V=y)

0.48; 0.47 (C=democrat; V=n)

0.52; 0.43 (C=republican; V=n)

0.57; 0.04 (C=democrat; V=?)

0.43; 0.03 (C=republican; V=?)

Attribute 12: (A = education-spending)

0.21; 0.13 (C=democrat; V=y)

0.79; 0.42 (C=republican; V=y)

0.91; 0.80 (C=democrat; V=n)

0.09; 0.06 (C=republican; V=n)

0.58; 0.07 (C=democrat; V=?)

0.42; 0.04 (C=republican; V=?)

Attribute 13: (A = superfund-right-to-sue)

0.35; 0.27 (C=democrat; V=y)

0.65; 0.42 (C=republican; V=y)

0.89; 0.67 (C=democrat; V=n)

0.11; 0.07 (C=republican; V=n)

0.60; 0.06 (C=democrat; V=?)

0.40; 0.03 (C=republican; V=?)

Attribute 14: (A = crime)

0.36; 0.34 (C=democrat; V=y)

0.64; 0.49 (C=republican; V=y)

0.98; 0.63 (C=democrat; V=n)

0.02; 0.01 (C=republican; V=n)

0.59; 0.04 (C=democrat; V=?)

0.41; 0.02 (C=republican; V=?)

Attribute 15: (A = duty-free-exports)

0.92; 0.60 (C=democrat; V=y)

0.08; 0.04 (C=republican; V=y)

0.39; 0.34 (C=democrat; V=n)

0.61; 0.44 (C=republican; V=n)

0.57; 0.06 (C=democrat; V=?)

0.43; 0.04 (C=republican; V=?)

Attribute 16: (A = export-administration-act-south-africa)

0.64; 0.65 (C=democrat; V=y)

0.36; 0.30 (C=republican; V=y)

0.19; 0.04 (C=democrat; V=n)

0.81; 0.15 (C=republican; V=n)

0.79; 0.31 (C=democrat; V=?)

0.21; 0.07 (C=republican; V=?)

二 作业要求

以此数据集为背景，实现决策树算法，请自行分割训练和测试集来进行算法准确性判断。每个小组完成一份实验报告（描述实验过程，并将准确率记录其中），与源代码压缩打包后以【必（选）修班\_第x（数字）组\_第二次小组作业】命名,邮件名同上。[于12月21日前发送到ssAI2014@163.com](mailto:于12月21日前发送到ssAI2014@163.com)

（注：不按要求命名的邮件一律不予以接受）