

**Adam Yee**

**Comp 251**

**Project 1 Report: Decision Tree Implementation and Analysis**

**9/27/2011**

**Table of contents:**

- 1. Analysis with Plots**
- 2. Program output**

## Analysis and Plots

3a: Measuring Generalization

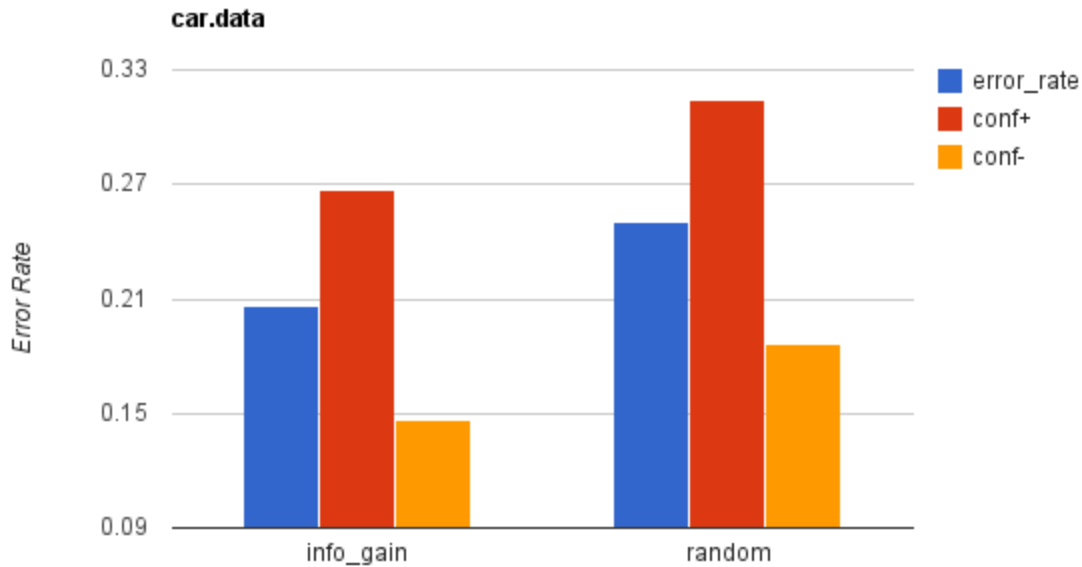
DATASET: car.data

---- Quinlan:

Averages (err,p,m): 0.206622046446 0.266846508954 0.146397583937

---- Random:

Averages (err,p,m): 0.250285298989 0.314527825783 0.186042772195



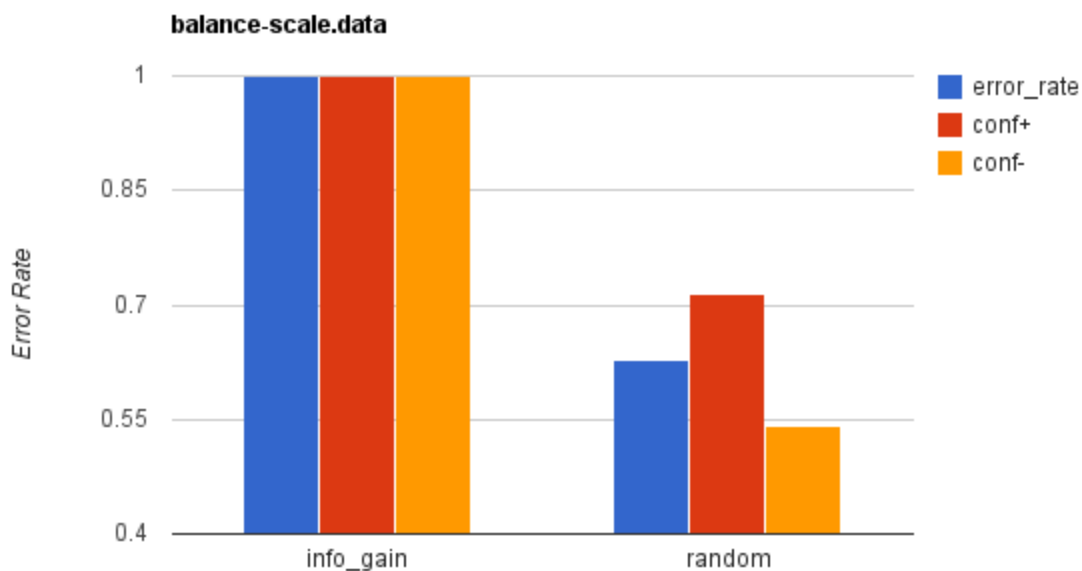
DATASET: balance-scale.data

---- Quinlan:

Averages (err,p,m): 1.0 1.0 1.0

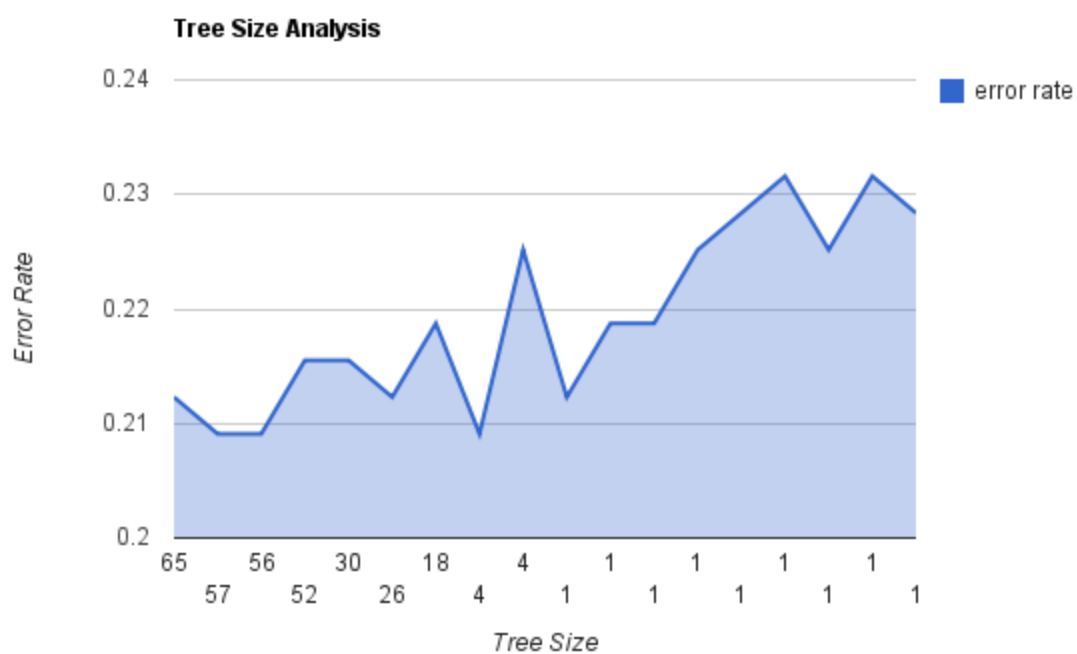
---- Random:

Averages (err,p,m): 0.627672869914 0.714824039796 0.540521700033



3b: Investigating the Correlation Between Tree Size and Tree Accuracy  
(65, 0.21221864951768488)

(57, 0.20900321543408362)  
 (56, 0.20900321543408362)  
 (52, 0.21543408360128624)  
 (4, 0.20900321543408362)  
 (1, 0.21221864951768488)  
 (30, 0.21543408360128624)  
 (1, 0.2186495176848875)  
 (18, 0.2186495176848875)  
 (1, 0.2186495176848875)  
 (26, 0.21221864951768488)  
 (1, 0.22508038585209)  
 (4, 0.22508038585209)  
 (1, 0.22829581993569126)  
 (1, 0.23151125401929262)  
 (1, 0.22508038585209)  
 (1, 0.23151125401929262)  
 (1, 0.22829581993569126)



## 2. Program Output

Program (Python script): [decisiontree.py](#)

COMP 251 Project 1, Decision Trees

Press '9' to exit or end sessions during inputs

Available data-sets:

- car.data        - balance-scale.data

Splitting method:

- quinlan        - random        - prune

Data-set choice > car.data

Splitting choice > quinlan

Session 1

Options:

1) dumpTree   2) treeSize    3) Categorize test\_set

choose one: 1

Node safety (N=1) (D=False)

    Node persons (N=2) (D=False)

        leaf unacc

        Node buying (N=3) (D=False)

        Node maint (N=8) (D=False)

    Node buying (N=12) (D=False)

        Node lug\_boot (N=13) (D=False)

        Node maint (N=17) (D=False)

        Node doors (N=19) (D=False)

        leaf unacc

    leaf unacc

Options:

1) dumpTree   2) treeSize    3) Categorize test\_set

choose one: 2

Interior: 19

Leaves: 53

Total: 72

Options:

1) dumpTree   2) treeSize    3) Categorize test\_set

choose one: 3

Error rate: 0.244186046512

Options:

1) dumpTree   2) treeSize    3) Categorize test\_set

choose one: 9

ending session

Available data-sets:

- car.data        - balance-scale.data

Splitting method:

- quinlan        - random        - prune

Data-set choice > car.data

Splitting choice > prune

Pruning, please wait...

...

...

...

Pruning complete

## Session 2

Options:

1) dumpTree 2) treeSize 3) Categorize test\_set

choose one: 2

Interior: 9

Leaves: 25

Total: 34

## Program (Python script): [run-analysis.py](#)

Usage: python3 run-analysis.py 1 0 0

Where the #'s enable (1/0) analysis of:

generalization

tree size regarding pruning

ROC

adam@adam-laptop-2:~/workspace/COMP 251 Project 1 Decision Trees/src\$ python3 run-analysis.py 1 1 0

COMP 251 Project 1, Decision Trees

Running run-analysis.py

DATASET: car.data

---- Quinlan:

(err,p,m): 0.206581482608 0.266506754659 0.146656210558

---- Random:

calculating...

Averages (err,p,m): 0.251767005825 0.316130426334 0.187403585317

DATASET: balance-scale.data

---- Quinlan:

(err,p,m): 1.0 1.0 1.0

---- Random:

calculating...

Averages (err,p,m): 0.594504652436 0.689829471053 0.499179833819

Tree size analysis

Pruning 1 node(s), please wait...

Pruning 2 node(s), please wait...

Pruning 3 node(s), please wait...

Pruning 4 node(s), please wait...

Pruning 5 node(s), please wait...

Pruning 6 node(s), please wait...

Pruning 7 node(s), please wait...

Pruning 8 node(s), please wait...

Pruning 9 node(s), please wait...

Pruning 10 node(s), please wait...

Pruning 11 node(s), please wait...

Pruning 12 node(s), please wait...

Pruning 13 node(s), please wait...

Pruning 14 node(s), please wait...

Pruning 15 node(s), please wait...

Pruning 16 node(s), please wait...

Pruning 17 node(s), please wait...

Pruning 18 node(s), please wait...

(1, 0.23151125401929262)

(52, 0.23151125401929262)

(49, 0.23151125401929262)

(59, 0.23151125401929262)

(46, 0.23472668810289388)  
(37, 0.23151125401929262)  
(1, 0.23151125401929262)  
(30, 0.23151125401929262)  
(1, 0.23151125401929262)  
(22, 0.23794212218649524)  
(1, 0.23472668810289388)  
(1, 0.23151125401929262)  
(1, 0.23794212218649524)  
(15, 0.24437299035369775)  
(19, 0.24437299035369775)  
(1, 0.2411575562700965)  
(11, 0.2411575562700965)  
(25, 0.2411575562700965)

Analysis Complete