**Decision tree implementation**

**Arguments:**

Training\_data\_filename(space)Test\_data\_filename(space)depth

E.g.: monk1.txt monk1test.txt 2

**Training data filenames:** monk1.txt, monk2.txt, monk3.txt

**Test data filename:** monk1test.txt, monk2test.txt, monk3test.txt

**Important note: No error checking has been provided in case:**

1. Wrong filename is provided as input or spelling error.
2. Extra spacing or character in between the arguments.
3. Incorrect no. of arguments is provided.
4. Incorrect ordering of arguments.

Question1:

Assumption:

The Root node is the node considered at depth 0. Following is the accuracy obtained at various depths for the given datasets. Methodology used is: training on a particular dataset and testing on the corresponding test dataset. Eg: Decision tree if trained on monk1.txt has been tested on monk1test.txt (applicable for all datasets). All values are in %. Fully trained refers to the depth where the entire decision tree is trained and the accuracy does not increase beyond that point.

|  |  |  |  |
| --- | --- | --- | --- |
| **Depth** | **monk1test.txt** | **monk2test.txt** | **monk3test.txt** |
| **0** | **75** | **59.72** | **80.55** |
| **1** | **72** | **60.64** | **97.22** |
| **2** | **74.30** | **61.80** | **100** |
| **3** | **75.23** | **61.34** | **95.83 (Fully trained)** |
| **4** | **79.62** | **65.27** |  |
| **5** | **80.09 (Fully trained)** | **67.36 (Fully trained)** |  |
| **Average** |  |  |  |

Question2:

I have created a confusion matrix and the learned tree for each of the datasets for each of their corresponding levels. I would like to mention that my root node is depth 0.

monk1.txt:

Learned decision tree (depth 1 and 2):

**Depth 1:**

####### LEVEL 0 #######

a5(0)

1234

####### LEVEL 1 #######

a4(2) a6(3) a1(4)

123 12 123

Explanation:

a5 is my root feature and since it is not obtained by splitting (obtained through prior entropy) I have denoted its parent as 0. The format of printing is: feature\_name (parent\_attribute it has split on).

For example: feature a4 is splitting on attribute 2 of feature a5.

Confusion matrix for depth 1 for monk1test.txt:

|  |  |  |
| --- | --- | --- |
| n = 432 | Predicted 0 | Predicted 1 |
| Actual 0 | 192 | 24 |
| Actual 1 | 96 | 120 |

**Depth 2:**

####### LEVEL 0 #######

a5(0)

1234

####### LEVEL 1 #######

a4(2) a6(3) a1(4)

123 12 123

####### LEVEL 2 #######

a1(1) a2(2) a3(3) a4(1) a3(2) a2(1) a2(2) a2(3)

123 123 12 123 12 123 123 23

Confusion matrix for depth 2 for monk1test.txt:

|  |  |  |
| --- | --- | --- |
| n = 432 | Predicted 0 | Predicted 1 |
| Actual 0 | 150 | 66 |
| Actual 1 | 45 | 171 |

monk2.txt:

Learned decision tree (depth 1 and 2):

**Depth 1:**

####### LEVEL 0 #######

a5(0)

1234

####### LEVEL 1 #######

a3(1) a3(2) a3(3) a2(4)

12 12 12 123

Confusion matrix for depth 1 for monk2test.txt:

|  |  |  |
| --- | --- | --- |
| n = 432 | Predicted 0 | Predicted 1 |
| Actual 0 | 222 | 68 |
| Actual 1 | 102 | 40 |

**Depth 2:**

####### LEVEL 0 #######

a5(0)

1234

####### LEVEL 1 #######

a3(1) a3(2) a3(3) a2(4)

12 12 12 123

####### LEVEL 2 #######

a6(1) a4(2) a4(1) a2(2) a6(1) a4(2) a6(1) a1(2) a3(3)

12 123 123 123 12 123 12 123 12

Confusion matrix for depth 1 for monk2test.txt:

|  |  |  |
| --- | --- | --- |
| n = 432 | Predicted 0 | Predicted 1 |
| Actual 0 | 187 | 103 |
| Actual 1 | 62 | 80 |

monk3.txt:

Learned decision tree (depth 1 and 2):

**Depth 1:**

####### LEVEL 0 #######

a2(0)

123

####### LEVEL 1 #######

a5(1) a5(2) a4(3)

1234 1234 123

Confusion matrix for depth 1 for monk3test.txt:

|  |  |  |
| --- | --- | --- |
| n = 432 | Predicted 0 | Predicted 1 |
| Actual 0 | 204 | 0 |
| Actual 1 | 12 | 216 |

**Depth 2:**

####### LEVEL 0 #######

a2(0)

123

####### LEVEL 1 #######

a5(1) a5(2) a4(3)

1234 1234 123

####### LEVEL 2 #######

a4(3) a1(2) a3(3) a5(1)

123 123 12 1234

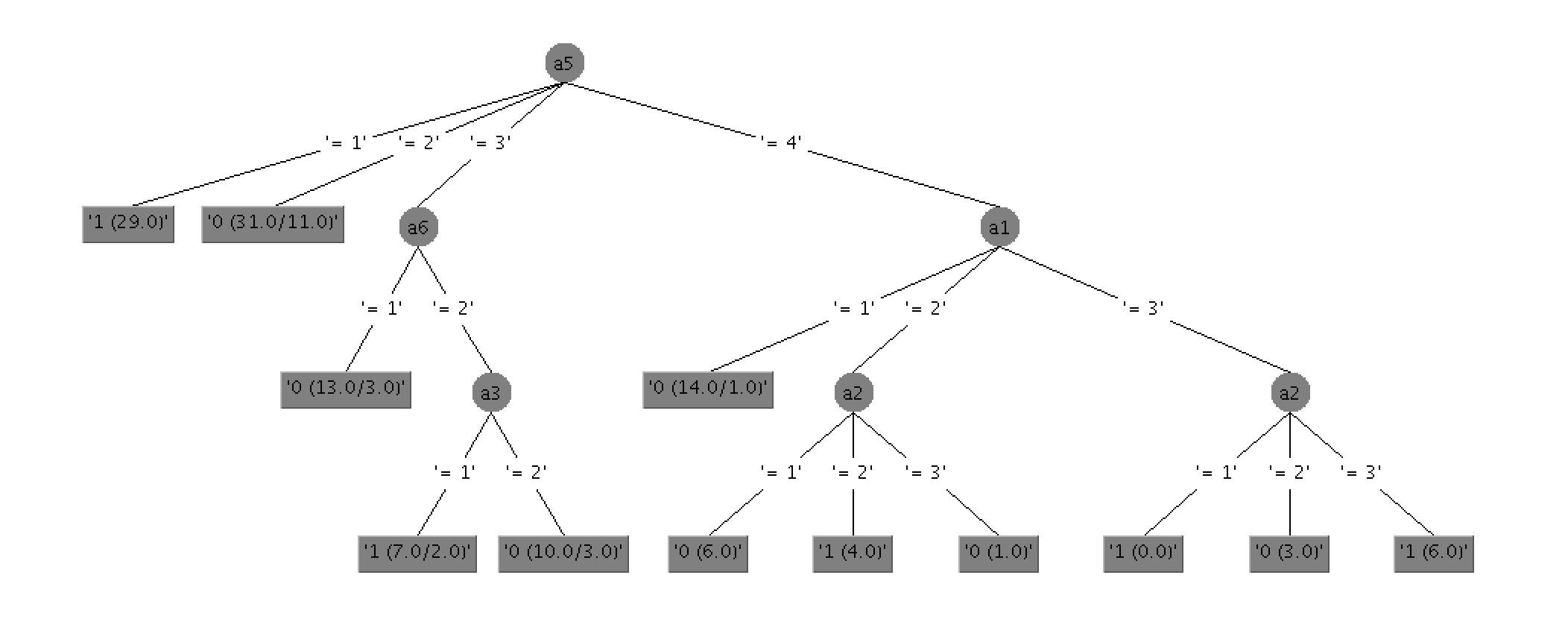
Confusion matrix for depth 2 for monk3test.txt:

|  |  |  |
| --- | --- | --- |
| n = 432 | Predicted 0 | Predicted 1 |
| Actual 0 | 204 | 0 |
| Actual 1 | 0 | 228 |

Question 3:

monk1.txt:

Learned decision tree of weka:



Confusion matrix:

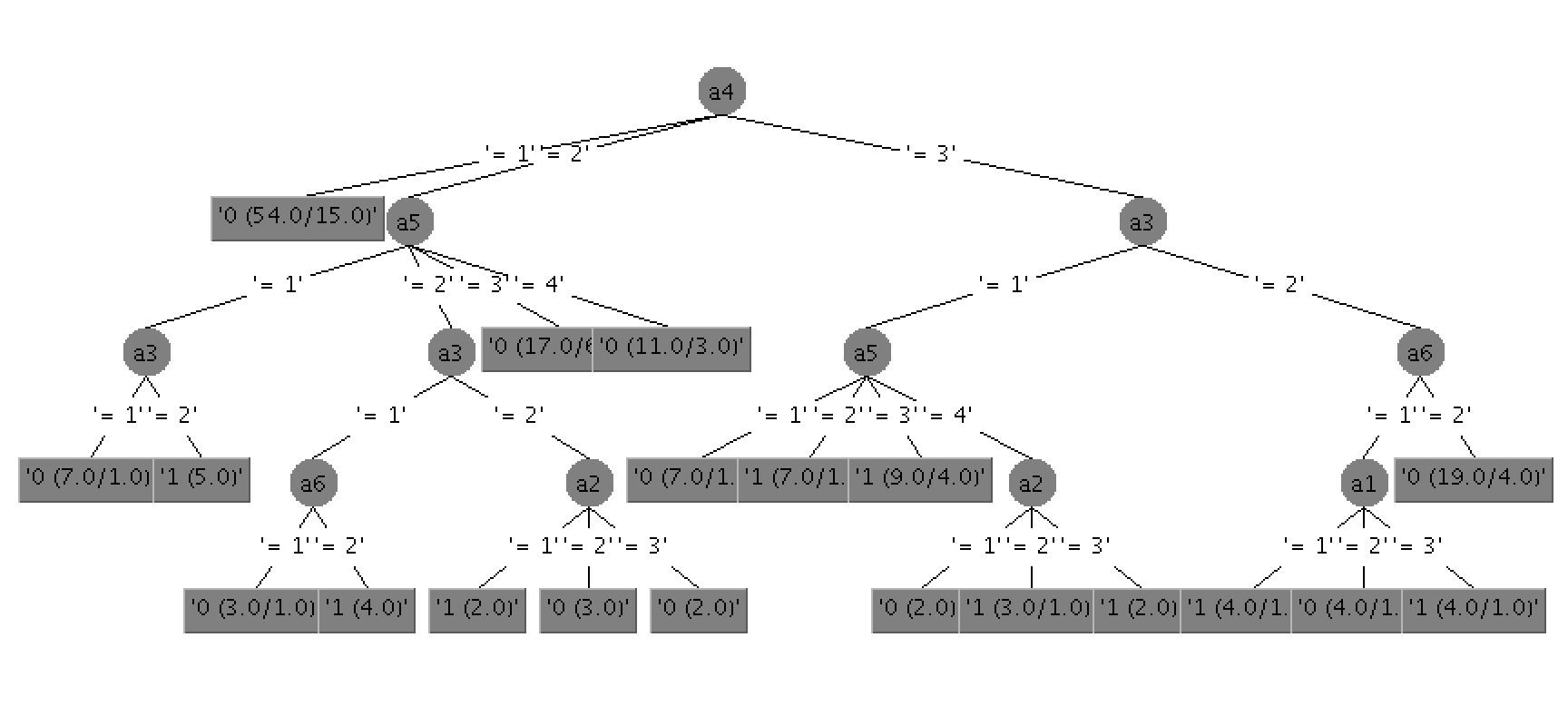
a b <-- classified as

186 30 | a = 0

75 141 | b = 1

monk2.txt:

Learned decision tree of weka:



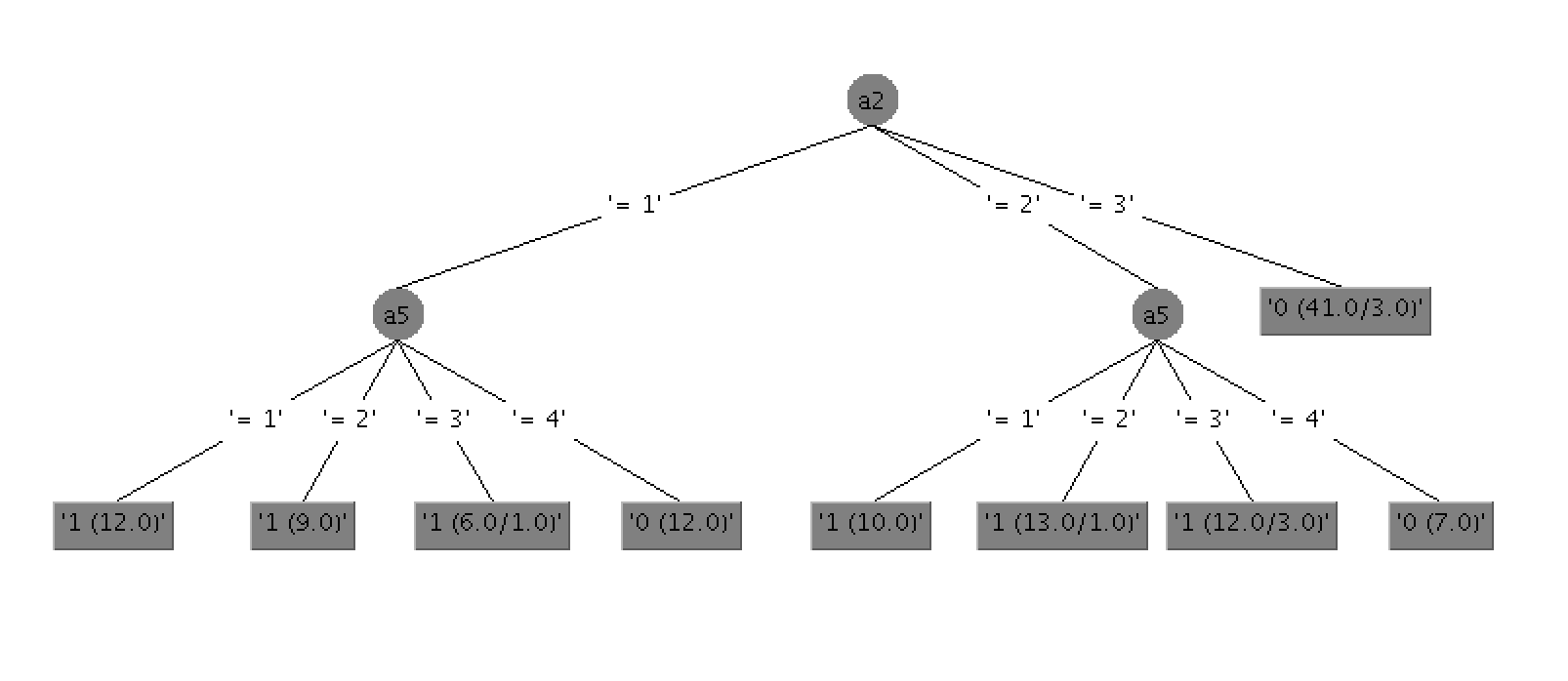
a b <-- classified as

233 57 | a = 0

94 48 | b = 1

monk3.txt:

Learned decision tree of weka:



a b <-- classified as

204 0 | a = 0

12 216 | b = 1

Question 4:

owndataset.txt:

Learned decision tree (depth 1 and 2):

**Depth 1:**

####### LEVEL 0 #######

a13(0)

01

####### LEVEL 1 #######

a11(0) a16(1)

01 01

Confusion matrix for depth 1 for owndataset\_test.txt:

|  |  |  |
| --- | --- | --- |
| n = 187 | Predicted 0 | Predicted 1 |
| Actual 0 | 12 | 3 |
| Actual 1 | 51 | 121 |

**Depth 2:**

####### LEVEL 0 #######

a13(0)

01

####### LEVEL 1 #######

a11(0) a16(1)

01 01

####### LEVEL 2 #######

a16(0) a3(1) a8(0)

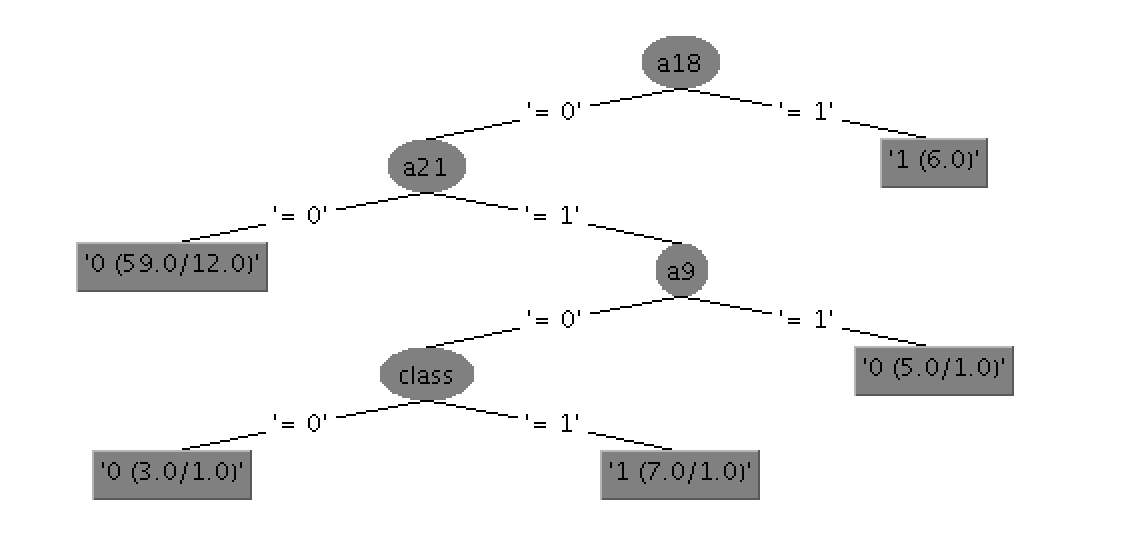
01 01 01

Confusion matrix for depth 2 for owndataset\_test.txt:

|  |  |  |
| --- | --- | --- |
| n = 187 | Predicted 0 | Predicted 1 |
| Actual 0 | 11 | 4 |
| Actual 1 | 50 | 122 |

WEKA ANALYSIS:

Learned decision tree:



Confusion matrix generated by weka:

a b <-- classified as

10 5 | a = 0

40 132 | b = 1