# 

# I526 Applied Machine Learning

## Weka assignment 1

### Created by:

### Himanshu Goyal (hgoyal)

### Rituraj Singh (singhri)

### Use Weka’s default decision tree (J48) algorithm on this training set to learn a decision tree. Report the tree and the confusion matrix on the test set.

Solution 1:

1. Training set: Monk\_train1

Test set: Monk\_test1

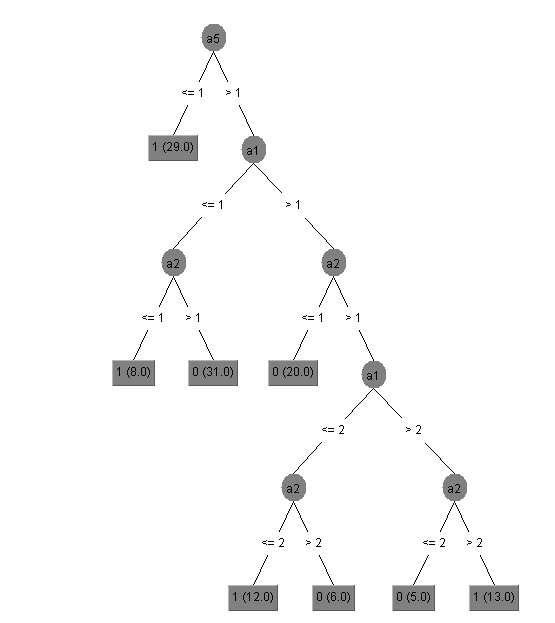
**Confusion matrix:**

|  |  |  |
| --- | --- | --- |
| **a** | **b** |  |
| 216 | 0 | **a=0** |
| 0 | 216 | **b=1** |

**Accuracy**: 100%

**J48 pruned tree:**

* + Number of Leaves: 8
  + Size of the tree: 15



1. Training set: Monk\_train2

Test set: Monk\_test2

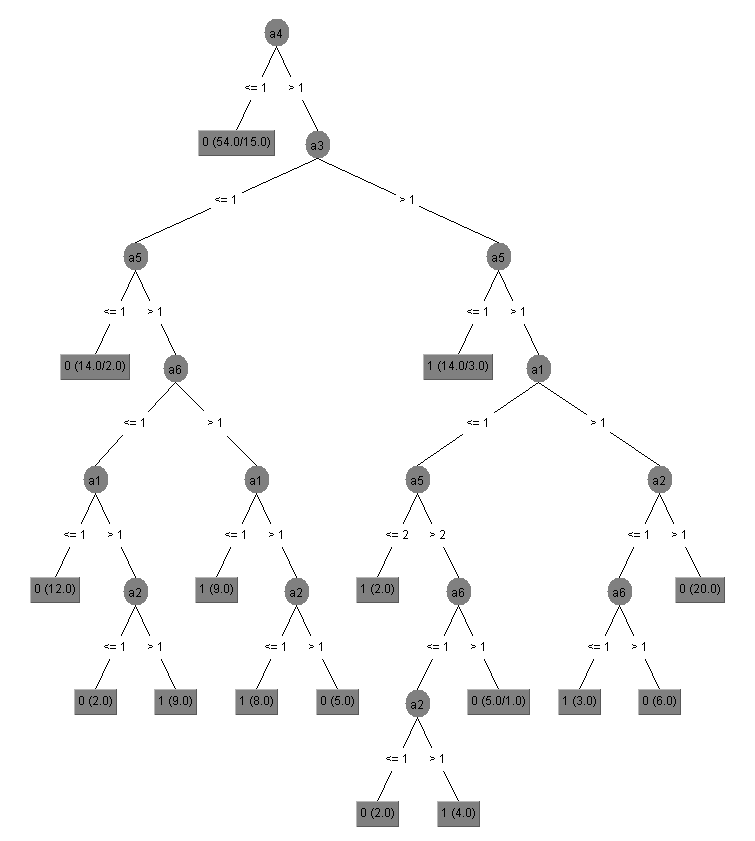
**Confusion matrix:**

|  |  |  |
| --- | --- | --- |
| **a** | **b** |  |
| 252 | 32 | **a=0** |
| 52 | 90 | **b=1** |

**Accuracy**: 80.55%

**J48 pruned tree**:

* + Number of Leaves: 16
  + Size of the tree: 31



1. Training set: Monk\_train3

Test set: Monk\_test3

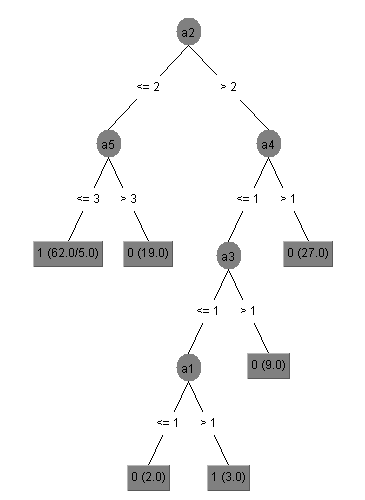
**Confusion matrix:**

|  |  |  |
| --- | --- | --- |
| **a** | **b** |  |
| 192 | 12 | **a=0** |
| 8 | 220 | **b=1** |

**Accuracy**: 95.37%

**J48 pruned tree**:

* + Number of Leaves: 6
  + Size of the tree: 11



**OVERALL AVG ACCURACY: (100+80.55+95.37)/3=91.97**

**RANGE: 19.45**

**VARIANCE: V=Σ(x-x')2/n = 68.82 AND STD. DEVIATION: SD = √V = 8.3**

1. Use Logistic Regression in Weka on the same data set.

Solution 2: Logistic

A) Training set: Monk\_train1

Test set: Monk\_test1

**Confusion matrix:**

|  |  |  |
| --- | --- | --- |
| **a** | **b** |  |
| 137 | 79 | **a=0** |
| 69 | 147 | **b=1** |

**Accuracy**: 65.74%

B) Training set: Monk\_train2

Test set: Monk\_test2

**Confusion matrix:**

|  |  |  |
| --- | --- | --- |
| **a** | **b** |  |
| 255 | 35 | **a=0** |
| 125 | 17 | **b=1** |

**Accuracy**: 62.96%

1. Training set: Monk\_train3

Test set: Monk\_test3

**Confusion matrix:**

|  |  |  |
| --- | --- | --- |
| **a** | **b** |  |
| 172 | 32 | **a=0** |
| 48 | 180 | **b=1** |

**Accuracy**: 81.48%

**OVERALL AVG ACCURACY: (65.74+62.96+81.48)/3=70.06**

**RANGE: 15.74**

**VARIANCE: V=Σ(x-x')2/n = 66.5 AND STD. DEVIATION: SD = √V = 8.15**

### Repeat steps 1 and 2 with your “own” data set and report the confusion matrices.

Solution 3:

J-48: (80:20 split)

**Confusion Matrix:**

|  |  |  |
| --- | --- | --- |
| **a** | **b** |  |
| 32 | 40 | **a=0** |
| 10 | 118 | **b=1** |

**ACCURACY**: 75%

**J48 pruned tree**:

* Number of Leaves: 93
* Size of the tree: 124



**Logistic**:

**Confusion matrix:**

|  |  |  |
| --- | --- | --- |
| **a** | **b** |  |
| 147 | 153 | **a=0** |
| 96 | 604 | **b=1** |

**ACCURACY**:75.1%