### CSCI 630

HW 3-P

#### Abhishek Shah, as5553

## Question 1

**Root node (Level 0)**

Attribute number: 1

ATrue = 21

BTrue = 88

AFalse = 51

BFalse = 40

True Probability = 0.545

False Probability = 0.455

True Entropy = 0.7070040649747601

False Entropy = 0.9894340222202263

Remainder A = 0.8355096955214472

Entropy Attribute = 0.9426831892554922

------------------------------------------------

Attribute number: 2

ATrue = 55

BTrue = 60

AFalse = 17

BFalse = 68

True Probability = 0.575

False Probability = 0.425

True Entropy = 0.9986359641585718

False Entropy = 0.7219280948873623

Remainder A = 0.8810351197183077

Entropy Attribute = 0.9426831892554922

------------------------------------------------

Attribute number: 3

ATrue = 48

BTrue = 86

AFalse = 24

BFalse = 42

True Probability = 0.67

False Probability = 0.33

True Entropy = 0.9411864371816835

False Entropy = 0.9456603046006402

Remainder A = 0.9426628134299393

Entropy Attribute = 0.9426831892554922

------------------------------------------------

Attribute number: 4

ATrue = 48

BTrue = 12

AFalse = 24

BFalse = 116

True Probability = 0.3

False Probability = 0.7

True Entropy = 0.7219280948873623

False Entropy = 0.6609623351442085

Remainder A = 0.6792520630671547

Entropy Attribute = 0.9426831892554922

------------------------------------------------

Attribute number: 5

ATrue = 18

BTrue = 99

AFalse = 54

BFalse = 29

True Probability = 0.585

False Probability = 0.415

True Entropy = 0.6193821946787638

False Entropy = 0.9335289015212996

Remainder A = 0.7497530780184161

Entropy Attribute = 0.9426831892554922

------------------------------------------------

Attribute number: 6

ATrue = 18

BTrue = 69

AFalse = 54

BFalse = 59

True Probability = 0.435

False Probability = 0.565

True Entropy = 0.7355085815538398

False Entropy = 0.9985872364932767

Remainder A = 0.8841480215946216

Entropy Attribute = 0.9426831892554922

------------------------------------------------

Attribute number: 7

ATrue = 61

BTrue = 44

AFalse = 11

BFalse = 84

True Probability = 0.525

False Probability = 0.475

True Entropy = 0.9810077194625308

False Entropy = 0.5171354891896202

Remainder A = 0.7606684100828983

Entropy Attribute = 0.9426831892554922

------------------------------------------------

Attribute number: 8

ATrue = 65

BTrue = 102

AFalse = 7

BFalse = 26

True Probability = 0.835

False Probability = 0.165

True Entropy = 0.9642953539833812

False Entropy = 0.7455178428108287

Remainder A = 0.9281970646399099

Entropy Attribute = 0.9426831892554922

------------------------------------------------

A count = 72 B count = 128

Information gain = [0.10717349373404506, 0.06164806953718449, 2.0375825552920723e-05, 0.2634311261883375, 0.19293011123707615, 0.058535167660870635, 0.1820147791725939, 0.014486124615582296]

Maximum of information gain = 0.2634311261883375

**4 th attribute will be the root node which is nothin but level 0.**

------------------------------------------------

**True node split (Level 1)**

Attribute number: 1

ATrue = 15

BTrue = 9

AFalse = 33

BFalse = 3

True Probability = 0.4

False Probability = 0.6

True Entropy = 0.9544340029249649

False Entropy = 0.41381685030363374

Remainder A = 0.6300637113521662

Entropy Attribute = 0.7219280948873623

------------------------------------------------

Attribute number: 2

ATrue = 37

BTrue = 4

AFalse = 11

BFalse = 8

True Probability = 0.6833333333333333

False Probability = 0.31666666666666665

True Entropy = 0.4612160405139095

False Entropy = 0.9819407868640977

Remainder A = 0.6261122101914691

Entropy Attribute = 0.7219280948873623

------------------------------------------------

Attribute number: 3

ATrue = 34

BTrue = 6

AFalse = 14

BFalse = 6

True Probability = 0.6666666666666666

False Probability = 0.3333333333333333

True Entropy = 0.6098403047164004

False Entropy = 0.8812908992306927

Remainder A = 0.7003238362211645

Entropy Attribute = 0.7219280948873623

------------------------------------------------

Attribute number: 5

ATrue = 11

BTrue = 9

AFalse = 37

BFalse = 3

True Probability = 0.3333333333333333

False Probability = 0.6666666666666666

True Entropy = 0.9927744539878084

False Entropy = 0.38431154412649704

Remainder A = 0.5871325140802675

Entropy Attribute = 0.7219280948873623

------------------------------------------------

Attribute number: 6

ATrue = 8

BTrue = 9

AFalse = 40

BFalse = 3

True Probability = 0.2833333333333333

False Probability = 0.7166666666666667

True Entropy = 0.9975025463691153

False Entropy = 0.3650551896402848

Remainder A = 0.54424860738012

Entropy Attribute = 0.7219280948873623

------------------------------------------------

Attribute number: 7

ATrue = 41

BTrue = 4

AFalse = 7

BFalse = 8

True Probability = 0.75

False Probability = 0.25

True Entropy = 0.4327501587887541

False Entropy = 0.9967916319816366

Remainder A = 0.5737605270869748

Entropy Attribute = 0.7219280948873623

------------------------------------------------

Attribute number: 8

ATrue = 45

BTrue = 11

AFalse = 3

BFalse = 1

True Probability = 0.9333333333333333

False Probability = 0.06666666666666667

True Entropy = 0.7147274731317892

False Entropy = 0.8112781244591328

Remainder A = 0.7211641832202788

Entropy Attribute = 0.7219280948873623

------------------------------------------------

A True count = 48 B True count = 12

Information gain = [0.10717349373404506, 0.06164806953718449, 2.0375825552920723e-05, 0.2634311261883375, 0.19293011123707615, 0.058535167660870635, 0.1820147791725939, 0.014486124615582296]

Maximum of information gain = 0.17767948750724227

**6 th attribute will be the node of True split of the root node which will be at level 2.**

------------------------------------------------

**False node split (Level 1)**

Attribute number: 1

ATrue = 6

BTrue = 79

AFalse = 18

BFalse = 37

True Probability = 0.6071428571428571

False Probability = 0.39285714285714285

True Entropy = 0.3681150054280774

False Entropy = 0.9121156307204276

Remainder A = 0.5818295367929293

Entropy Attribute = 0.6609623351442085

------------------------------------------------

Attribute number: 2

ATrue = 18

BTrue = 56

AFalse = 6

BFalse = 60

True Probability = 0.5285714285714286

False Probability = 0.4714285714285714

True Entropy = 0.8003922080453356

False Entropy = 0.4394969869215134

Remainder A = 0.6302558895155337

Entropy Attribute = 0.6609623351442085

------------------------------------------------

Attribute number: 3

ATrue = 14

BTrue = 80

AFalse = 10

BFalse = 36

True Probability = 0.6714285714285714

False Probability = 0.32857142857142857

True Entropy = 0.6071716548713029

False Entropy = 0.7553754125614288

Remainder A = 0.6558671752552013

Entropy Attribute = 0.6609623351442085

------------------------------------------------

Attribute number: 5

ATrue = 7

BTrue = 90

AFalse = 17

BFalse = 26

True Probability = 0.6928571428571428

False Probability = 0.30714285714285716

True Entropy = 0.37395136668121065

False Entropy = 0.9681647320759548

Remainder A = 0.5564597574810249

Entropy Attribute = 0.6609623351442085

------------------------------------------------

Attribute number: 6

ATrue = 10

BTrue = 60

AFalse = 14

BFalse = 56

True Probability = 0.5

False Probability = 0.5

True Entropy = 0.5916727785823275

False Entropy = 0.7219280948873623

Remainder A = 0.6568004367348449

Entropy Attribute = 0.6609623351442085

------------------------------------------------

Attribute number: 7

ATrue = 20

BTrue = 40

AFalse = 4

BFalse = 76

True Probability = 0.42857142857142855

False Probability = 0.5714285714285714

True Entropy = 0.9182958340544896

False Entropy = 0.28639695711595625

Remainder A = 0.5572107615181847

Entropy Attribute = 0.6609623351442085

------------------------------------------------

Attribute number: 8

ATrue = 20

BTrue = 91

AFalse = 4

BFalse = 25

True Probability = 0.7928571428571428

False Probability = 0.20714285714285716

True Entropy = 0.680471054135434

False Entropy = 0.5787946246321198

Remainder A = 0.6594095080240332

Entropy Attribute = 0.6609623351442085

------------------------------------------------

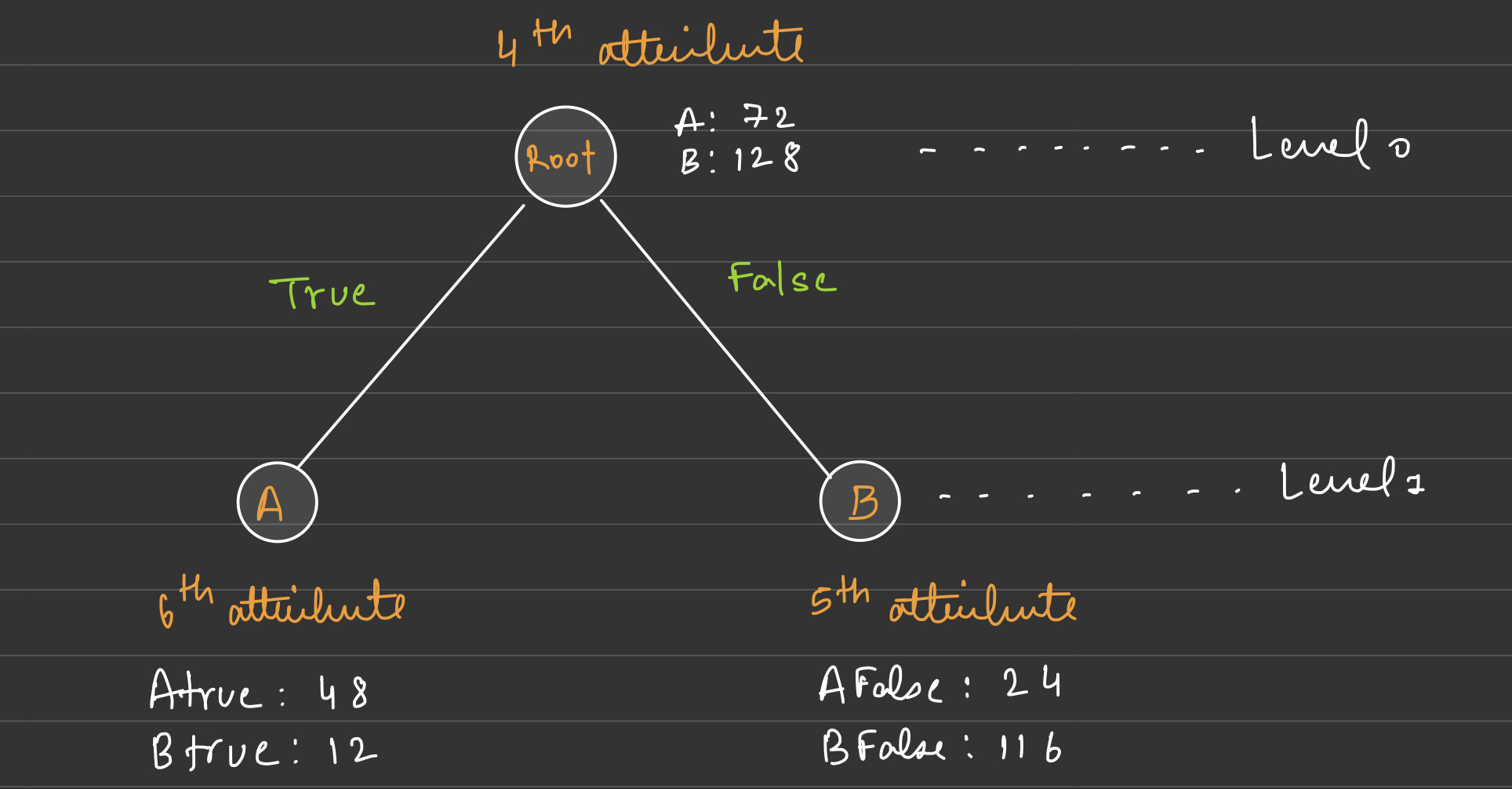
A False count = 24 B False count = 116

Information gain = [0.10717349373404506, 0.06164806953718449, 2.0375825552920723e-05, 0.2634311261883375, 0.19293011123707615, 0.058535167660870635, 0.1820147791725939, 0.014486124615582296]

Maximum of information gain = 0.1045025776631836

**5 th attribute will be the node of False split of the root node which will be at level 2.**

## Question 2



As we can see above, 4th Attribute becomes the first stump of our decision tree.

First stump is nothing but the root of our decision tree.

Consider two splits (True split, False split) for our root node.

Node A (A True count = 48 B True count = 12) will be the node of True split as A True count is maximum for that split.

Node B (A False count = 24 B False count = 116) will be the node of False split as B False count is maximum for that split.

**3.** Initial weight

There are altogether 200 data samples provided.

**Initial weight** = 1/200

= **0.005**

**1.** Error rate for first stump

**Error** = Incorrect number data of A \* Initial weight + Incorrect number data of B \* Initial weight

= 12 \* 0.005 + 24 \* 0.005

=  **0.18**

**2.** Hypothesis weight

**Hypothesis weight** = ln((1- error)/error)

= ln((1-0.18)/0.18)

= **1.52**

**4.** New weights of each example correctly classified

Incorrectly classified weight remains unchanged and it is nothing but the initial weight.

**Incorrectly classified weight** = Initial weight = **0.005**

**Correctly classified weight** = Initial weight \* (error/ (1-error))

= 0.005 \* (0.18/(1-0.18))

**= 0.0010975**