ASSIGNMENT 2
Part 1

1. Representing Boolean Functions (10 points)

Give decision trees to represent the following concepts. Your decision tree must contain as few nodes as possible. You can assume A, B, and C are Boolean variables.

(a) Y = (¬A V B) ^ ¬(C ^ A)

A = 0 : 1

A = 1 :

| C = 1 : 0

| C = 0 :

| | B = 0 : 0

| | B = 1 : 1

(b) Y = (A XOR B) ^ C

C = 0 : 0

C = 1 :

| A = 1 :

| | B = 1 : 0

| | B = 0 : 1

| A = 0 :

| | B = 1 : 1

| | B = 0 : 0

(c) Y = (A V B) ^ (B V C) ^ (A V C)

A = 1 :

| B = 1 : 1

| B = 0 : 0

A = 0 :

| B = 0 : 0

| B = 1 :

| | C = 1 : 1

| | C = 0 : 0

(d) Y = (A V B) ^ ¬A ^ ¬B

Any : 0 – The output for this equation will always be 0, no matter what value combinations of A and B. Thus, on any input it should give output as 0.

2. Decision Trees (20 points)

In this question, you will use the ID3 algorithm to create a decision tree for the dataset given below. There are three Boolean attributes X1, X2, and X3 and a Boolean class attribute. Be sure to show detailed calculations for each step including entropy and information gain values. Draw a plot of the final tree that you obtain and show the class labels for the leaf nodes. Also indicate the set of instances that are associated with each leaf node.

At the top:

H(S) = Entropy = -1/2*log2(1/2) – 1/2*log2(1/2) = 1

IG(X1) = H(S) – { 1/5*(- 1/2*log2(1/2) - 1/2*log2(1/2)) +

4/5*(- 4/5*log2(4/5) – 1/5*log2(1/5)) }

= 0.27807191

IG(X2) = H(S) - { 0*(- 4/7*log2(4/7) - 3/7*log2(3/7)) +

1*(- 1/3*log2(1/3) – 2/3*log2(2/3)) }

= 0.08170417

IG(X3) = H(S) - { 0*(- 3/8*log2(3/8) - 5/8*log2(5/8)) +

1*(- 1*log2(1) – 0*log2(0)) }

= 1.0

Choose X3 as the topmost node!

Under X3:

IG(X1) = H(S) – { 0*(- 3/8*log2(3/8) - 5/8*log2(5/8)) +

1*(- 3/4*log2(3/4) – 1/4*log2(1/4)) }

= 0.954434

IG(X2) = H(S) - { 0*(- 4/7*log2(4/7) - 3/7*log2(3/7)) +

*(- 4/5*log2(0) – 1/5*log2(0)) }

= 0.954434

Second Choice: X2

Final: X1

Final Tree:

X3 = 0 :

| X2 = 0 :

| | X1 = 0 : 0

| | X1 = 1 : 1

| X2 = 1 : 1

X3 = 1 : 0