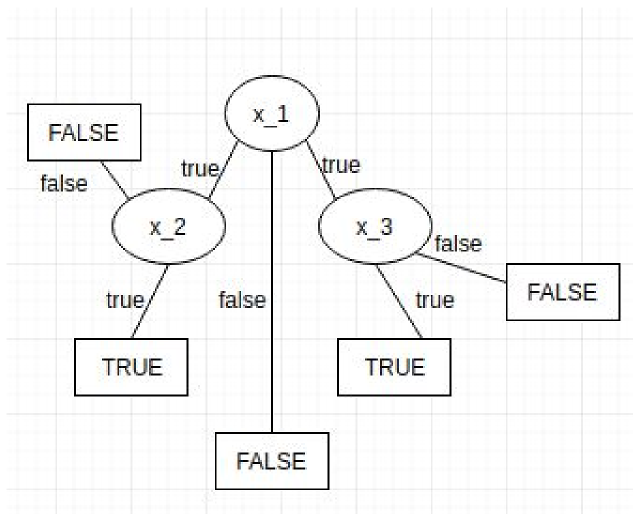
CMPU 395: Machine Learning - Assignment 01
October 10, 2018
Megan Nelson

Decision Trees

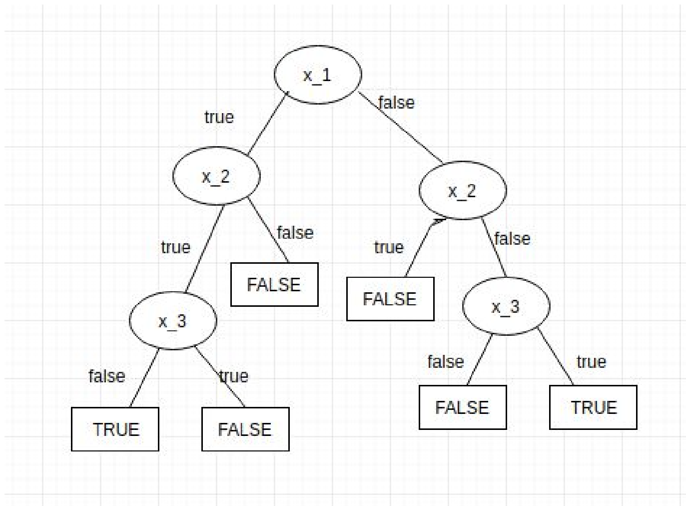
1. Decision Trees for Boolean functions:
   In order to convert a logical statement into a decision tree, it's easiest to convert the statement into a logical disjunction of conjunctions.
   a) $(x_1 \wedge x_2) \vee (x_1 \wedge x_3)$



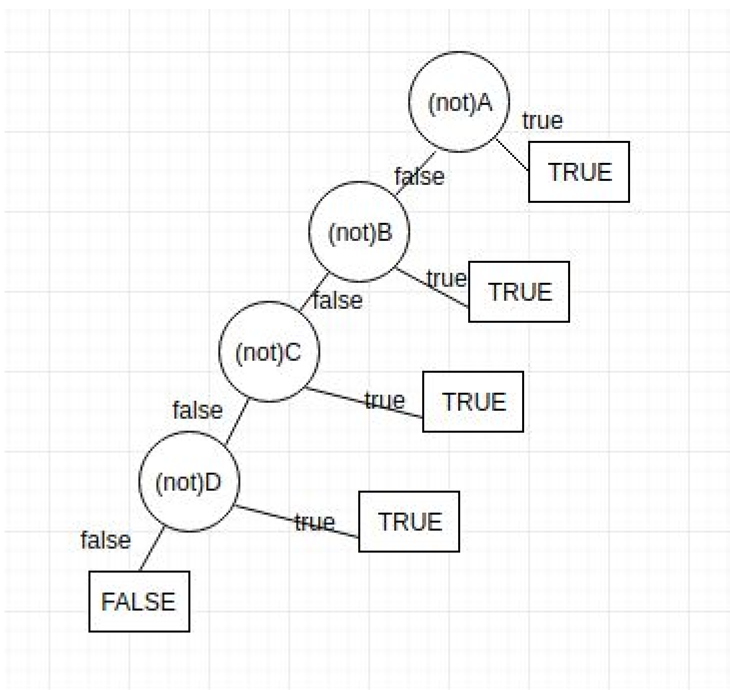
   b) $(x_1 \wedge x_2) \bigoplus x_3$

   In order to convert to DT, first convert the statement into a logical disjunction of conjunctions.
   Conversion from XOR to DNF (disjunctive normal form):
   $a(not)b \Rightarrow (a \wedge (notb)) \vee ((not)a \wedge b)$
   Then:
   $((x_1 \wedge x_2) \wedge (not)x_3) \vee ((not)(x_1 \wedge x_2) \wedge x_3)))$

   $(x_1 \wedge x_2 \wedge (not)x_3) \vee ((not)x_1 \wedge (not)x_2 \wedge x_3)$



c) $(not)A \vee (not)B \vee (not)C \vee (not)D$



2. Will aliens invade the earth?

a) There are 2^9 = 512 possible different mappings of the four features to the boolean outputs since each way to fill the 9 different label slots is a different function.

b) Entropy of the labels

Invade: 5/9 examples
Not Invade: 4/9 examples

$$H(Invade?) = -(5/9)log_2(5/9) - (4/9)log_2(4/9)$$

= ~ .991

c) Information Gain of the features

<u>Technology</u>

$$EE = (3/9) * .918 + (6/9) * .918 = .918$$

IG: .991 - .918 = **.073**

Technology = Yes: 3/9

T=Yes && Invade=Yes: 1/3, T=Yes && Invade=No: ⅔

$$H(Tech - Yes) = -(1/3)log_2(1/3) - (2/3)log_2(2/3)$$

= ~.918

Technology = No: 6/9

$$H(Tech - No) = -(4/6)log_2(4/6) - (2/6)log_2(2/6)$$

= ~.918

<u>Environment:</u>

$$EE = (5/9) * .721 + (4/9) * .811 = .761$$

.991 - .761 = **.23**

Environment = Yes: 5/9

$$H(Env - Yes) = -(4/5)log_2(4/5) - (1/5)log_2(1/5)$$

=~.721

Environment = No: 4/9

$$H(Env - No) = -(1/4)log_2(1/4) - (3/4)log_2(3/4)$$

= ~.811

<u>Human:</u>

$$EE = (4/9) * .811 + (4/9) * 0 + (1/9) * 0 = .360$$

.991 - .360 = **.631**

Human = Like: 4/9

= ~.811

Human = Not Care: 4/9

= 0

Human = Hate: 1/9

= 0

Distance:

$$EE = (2/9) * 1 + (1/9) * 0 + (3/9) * .918 + (3/9) * .918 = .834$$

.991 - .834 = **.157**

Distance = 1: 2/9

= 1

Distance = 2: 1/9

= 0

Distance = 3: 3/9

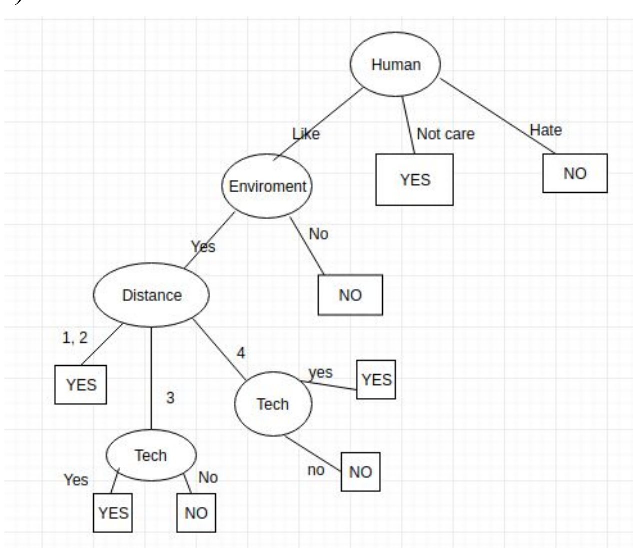
=~.918

Distance = 4: 3/9

=~.918

d) I will use the human feature as the attribute for tree root because it has the most information gain out of all the other features, meaning that the human feature best classifies the set of instances into pure subsets.

e)



f) First instance: Yes
Second instance: No
Third instance: Yes

My decision tree was accurate for ⅔ of the new cases.

3. Majority error

a) Calculate the Majority Error measure, calculate information gain for the four features.

ME(Invade?) = 1 - 5/9 = 4/9

Technology
- ME(T-yes) = 1 - ⅔ = 1/3
- ME(T-No) = 1 - 4/6 = ⅓
- Avg(ME) = 1/3
- Info Gain: 4/9-⅓ = **.111**

Environment
- ME(E-Yes) = 1 - ⅘ = 1/5
- ME(E-No) = 1 - ¾ = ¼
- Avg(ME) = .225
- Info Gain: 4/9 - .225 = **.219**

Human
- ME(H-Like) = 1 - ¾ = 1/4
- ME(H-Not Care) = 1 - 4/4 = 0
- ME(H-Hate) = 1 - 1/1 = 0
- Avg(ME) = .083
- Info Gain: 4/9 - .083 = **.361**

Distance
- ME(D-1) = 1 - ½ = ½
- ME(D-2) = 1 - 1/1 = 0
- ME(D-3) = 1 - ⅔ = ⅓
- ME(D-4) = 1 - ⅔ = ⅓
- Avg(ME) = .291
- Info Gain: 4/9 - 2.91 = **.153**

b) The root of the tree should still be the human attribute. Both majority error and entropy produced the same precedence of root ordering in determining which attribute classifies the set S of instances most purely (as in - the human attribute classifies the majority of S, followed by environment, distance, and technology, which created the same result as using entropy).

Linear Classifiers

1. Linear classifier: a vector of k weights $<w_1 \ldots w_k>$ and a threshold Y with the following rule:

   If $w_1(x_1) + \ldots + w_k(x_k) > Y$ then Y = true else false

   A data table can be converted to this form.
   Hypothesis: $(x_1) + (x_2) + (x_3) + (x_4) >= 2$ then $y=1$, else $y=-1$

2. Three out of thirteen instances are misclassified when using the classifier.

3. New hypothesis: $(x_1) + (x_4) >= 1$ then $y=1$ else $y=-1$
   If both x_1 and x_4 are 0, then $y= -1$

Decision Lists (Extra Credit)

1. Show that 1-decision lists are linearly separable functions.
   Aka: find a w and b such that the decision list will return 1 if, and only if, $w^T(X)>= b$

   Linear separability: two sets of positive and negative example points are considered linearly separable if there exists one line in the plane that classifies one half-space as the set of ‘positive examples’ and the other half-space as one that contains only ‘negative examples’

   The only two possible functions that can be 1-decision lists are the NOT function and the decision itself being classed as true or false. Ie,

   x examples map directly to y output (share the same attribute value) (EXAMPLE 1)

x_1	y
1	1
0	0

NOT function (EXAMPLE 2)

x_1	y
0	1
1	0

Linear decision boundary function:
$Z = (w^T)x + b$
Y (output) → 1 if $z >= 0$
0 if $z<0$

Wrapping bias term into weight, and an extra dummy feature 1 into x n-dimensional vector, and setting threshold z to zero:

$w_0x_0 + \ldots + w_n(x_d) = 0$

FOR THE NOT FUNCTION:

$X_1 = 0$ and $y=1$, need a function that satisfies: $(w_1)(x_1) + b = b > 0$
(true for any **$b > 0$**, set $b=1$)
Second training case:
$X_1 = 1, y = 0$
$w_1(1) + b = \mathbf{w_1 + 1 < 0}$

We can satisfy this inequality for NOT functions: $w_1 = -2, b = 1$
Therefore the one decision list (NOT function) is linearly separable.

Now, let's check EXAMPLE 1 table (the only other example of a one-decision list).

For the first instance of the table:
$x_1(w_1) + b > 0$, $w_1 + b > 0$
For the second instance:'
$x_1(w_1) + b < 0$, $b < 0$
We can satisfy this inequality: **$w_1 = 0$, and $b=-1$**
Therefore, one decisions lists of any type are linearly separable since the weight and bias term inequality can be satisfied with every example of a one decision list.