Introduction to Artificial Intelligence – Assignment III – Tom Goodman (1526322)

(Apologies for the lack of LaTeX formatting this assignment. I simply haven't had time).

1. **Decision Tree** [3%]

Your company wants to make predictions about whether or not a visitor to their site will make a purchase, depending on observed browsing behaviour.

You have the following attributes and their values that can be measured.

Attribute	Values
Visitor	Integer counter in chronological order
Referrer to the page	Search Engine, Banner Ad, Price Comparison website, Other
Visit frequency of page	Once, Multiple times
Page search function used	yes, no
Length of visit	$<$ 5 minutes, 5 to 10 minutes, 10 to 15 minutes, longer
Purchase made	yes, no

To train the system you have recorded information on the last 16 distinct visitors:

Visitor	Referrer	Frequency	Search	Length	Purchase
1	Search	multiple	yes	15	no
2	Search	once	no	10	yes
3	Other	multiple	yes	5	yes
4	Comparison	once	yes	15	yes
5	Banner	once	yes	10	no
6	Other	multiple	yes	10	no
7	Other	once	no	5	no
8	Banner	once	no	5	no
9	Comparison	multiple	yes	10	yes
10	Comparison	once	yes	10	yes
11	Search	once	yes	15	yes
12	Search	multiple	yes	10	no
13	Banner	once	yes	longer	yes
14	Comparison	once	no	5	yes
15	Banner	once	no	5	no
16	Comparison	multiple	no	longer	yes

Use the formulas on Entropy and Information Gain from the lecture in order to construct a decision tree. Show your working.

Referer

$$Gain(Referer) = B\left(\frac{p}{p+n}\right) - \boxed{Remainder(referer)}$$

$$-\left(\frac{p}{p+n}\log_2\frac{p}{p+n} + \left(1-\frac{p}{p+n}\right)\log_2\left(1-\frac{p}{p+n}\right)\right)$$

$$-\left(\frac{9}{16}\log_2\frac{9}{16} + \left(\frac{7}{16}\right)\log_2\left(\frac{7}{16}\right)\right)$$

$$= \underline{0.9887}.$$

$$\sum_{k=1}^{d}\frac{p_k+n_k}{p+n}\times -\left(\frac{p_k}{p_k+n_k}\log_2\frac{p_k}{p_k+n_k} + \left(1-\frac{p_k}{p_k+n_k}\right)\log_2\left(1-\frac{p_k}{p_k+n_k}\right)\right)$$

$$\sum_{k=1}^{4}\frac{p_k+n_k}{p+n}\times -\left(\frac{p_k}{p_k+n_k}\log_2\frac{p_k}{p_k+n_k} + \left(\frac{n_k}{p_k+n_k}\right)\log_2\frac{n_k}{p_k+n_k}\right)$$

$\underline{k=1}$ (search)

$$\frac{4}{16}\times -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right)$$

$$= -\frac{1}{4}\log_2\frac{1}{2} = \frac{1}{4}$$

$\underline{k=2}$ (other)

$$\frac{3}{16}\times -\left(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right)$$

$$= 0.17218$$

01-12-2015

k=3 (Comparison)

$$\frac{5}{16} \times -\left(\frac{5}{5}\log_2\frac{5}{5} + \frac{0}{5}\log_2\frac{0}{5}\right)$$

$$= \frac{5}{16}\log_2 1 = 0$$

k=4 (Banner)

$$\frac{4}{16} \times -\left(\frac{1}{4}\log_2\frac{1}{4} + \frac{3}{4}\log_2\frac{3}{4}\right)$$

$$= 0.20282$$

$$\therefore \text{Gain(Referer)} = 0.9887 - 0.625$$

$$= 0.3637.$$

$$B\left(\frac{p}{p+n}\right) = -\left(\frac{p}{p+n}\log_2\frac{p}{p+n} + \frac{n}{p+n}\log_2\frac{n}{p+n}\right)$$

$$B(q) = -\left(q\log_2 q + (1-q)\log_2(1-q)\right)$$

Gain (Frequency) 02-12-2015

0.9887 − Remainder(Frequency)

$$\sum_{k=1}^{2} -\frac{p_k+n_k}{p+n}\left(\frac{p_k}{p_k+n_k}\log_2\frac{p_k}{p_k+n_k}+\frac{n_k}{p_k+n_k}\log_2\frac{n_k}{p_k+n_k}\right)$$

k=1 (multiple)

6x, 3y, 3n,

$$\frac{-6}{16}\left(\frac{3}{6}\log_2\frac{3}{6}+\frac{3}{6}\log_2\frac{3}{6}\right)$$

$$=\frac{3}{8}$$

k=2 (once)

10x, 6y, 4n

$$\frac{-10}{16}\left(\frac{6}{10}\log_2\frac{6}{10}+\frac{4}{10}\log_2\frac{4}{10}\right)$$

$$=0.6068$$

$$\therefore \text{Gain (Frequency)} = 0.9887 - 0.9818$$

$$= 6.856\times10^{-3}$$

$\boxed{\text{Length}}$ 02-12-2015

$$Gain(Length) = 0.9887 - Remainder(Length)$$

$$\sum_{k=1}^{4} -\frac{p_k + n_k}{p+n}\left(\frac{p_k}{p_k+n_k}\log_2\frac{p_k}{p_k+n_k} + \frac{n_k}{p_k+n_k}\log_2\frac{n_k}{p_k+n_k}\right)$$

k=1 (15)

3, 2y, 1n

$$\frac{-3}{16}\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right)$$
$$= 0.1722$$

k=2 (10)

6, 3y, 3n.

$$\frac{-6}{16}\left(\log_2\frac{1}{2}\right)$$
$$= \frac{3}{8}$$

k=3 (5)

~~3y~~ 5, 2y, 3n.

$$\frac{-5}{16}\left(\frac{2}{5}\log_2\frac{2}{5} + \frac{3}{5}\log_2\frac{3}{5}\right)$$
$$= 0.3034.$$

k=4 (longer)

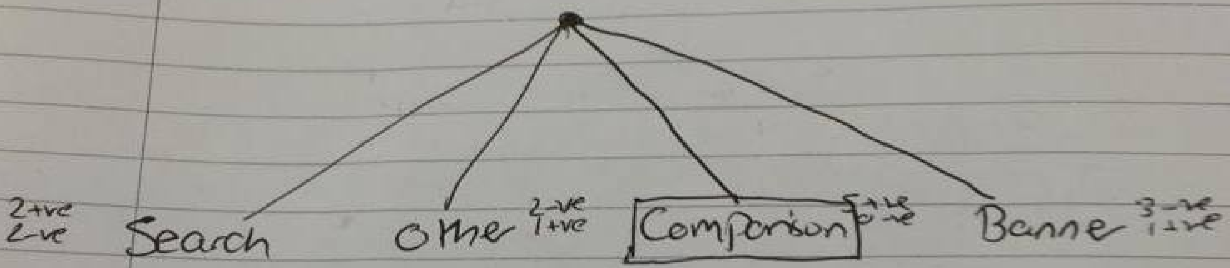
2, 2y, 0n

$$\frac{-2}{16}\left(\frac{2}{2}\log_2 1 + \frac{0}{2}\log_2\frac{0}{2}\right) = 0$$

6

$\therefore$ Gain (Length) $= 0.9887 - 0.8506$
$= 0.1381$.

Hence, since the information gain is highest for Referer, we should split by that.



$$H(\text{Purchase}) = -\left(\frac{p}{p+n}\log_2\frac{p}{p+n} + \frac{n}{p+n}\log_2\frac{n}{p+n}\right)$$

4+ve 7-ve

$$= -\left(\frac{4}{11}\log_2\frac{4}{11} + \frac{7}{11}\log_2\frac{7}{11}\right)$$

$$= 0.9457$$

02-12-2015

$H(\text{search}) =$
$-(\log_2 \frac{1}{2}) = 1$

$H(\text{Other}) =$
$-(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}) = 0.918$

$H(\text{Banner}) =$
$-(\frac{1}{4}\log_2\frac{1}{4} + \frac{3}{4}\log_2\frac{3}{4}) = 0.811$

SEARCH

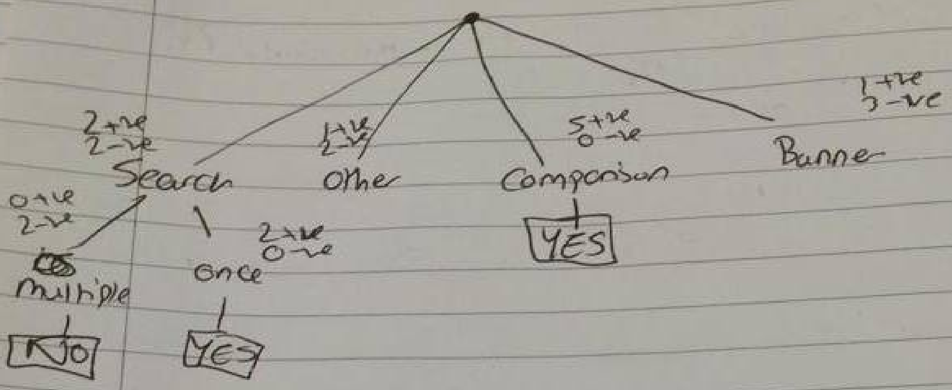
- Gain (Frequency) = 1 − Remainder (Frequency)

k=1 (multiple)
2, 0y, 2n
∴ = 0 (because $\log_2 1$)

k=2 (once)
2, 2y, 0n
∴ = 0 (because $\log_2 1$)

Gain (Frequency) = 1 − 0
= 1

Hence, we split by frequency, as it completely categorises the 'Search' branch



$\boxed{\text{OTHER}}$ 1y, 2n

Gain (Frequency) = 0.918 − Reminder (Frequency)

k=1 (multiple) 2, 1y, 1n.

$$= \frac{-2}{3}\left(\log_2 \frac{1}{2}\right) = \frac{2}{3}$$

k=2 (once) 1, 0y, 1n.

$$= 0 \quad \left(\text{because } \log_2 1\right)$$

$$\text{Gain(Frequency)} = 0.918 - \left(\frac{2}{3}\right)$$
$$= \underline{0.251}$$

02-12-2015

Gain (Search) $= 0.918 -$ Remainder (Search)

3,

k = 1 (yes)

2, 1p, 1n.

$$-\frac{2}{3}\left(\log_2 \frac{1}{2}\right) = \frac{2}{3}$$

k = 2 (no)

1, 1p 0n.

$= 0$ (Because $\log_2 1$)

Gain (Search) $= 0.918 - \frac{2}{3}$

$= 0.251$

Gain (Length) =

k = 1 (15)

0, 0p, 0n.

$= 0$

k = 2 (10)

1, 0y 1n.

$= 0$ (Because $\log_2 1$)

k = 3 (5) 2, 1y, 1n.

$$= -\frac{2}{3}\left(\log_2 \frac{1}{2}\right) = \frac{2}{3}.$$

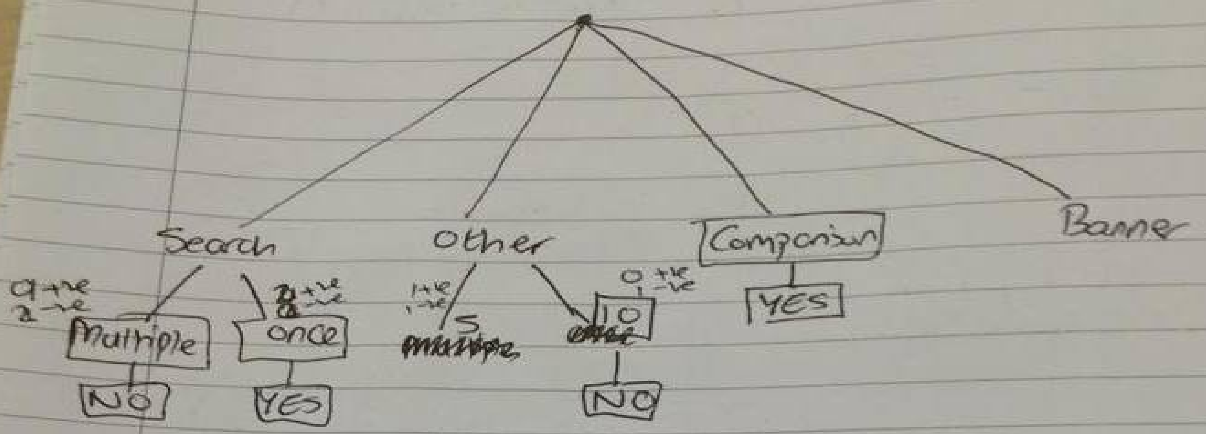
k = 4 (longer)

0, 0y 0n.

$= 0$

11 Gain (Length) $= 0.918 - \frac{2}{3} = 0.251$

Because all of the gains are the same, we will split by Frequency.



$\boxed{\text{Banner}}$

~~Gain (Frequency) = 0.811 - Remainder (Frequency)~~

~~k = 1 (multiple)~~

~~0, 0y, 0n~~

~~= 0~~

k = 2 (

Gain (Length) = 0.811 - ~~Gain~~ Remainder (Length)

k = 1 (15)

0, 0y, 0n

= 0

02-12-2015

K=2 (10)

1, 0y, 1n

= 0 (Because $\log_2 1$)

K=3 (5)

2, 0y, 2n

= 0 (Because $\log_2 1$)

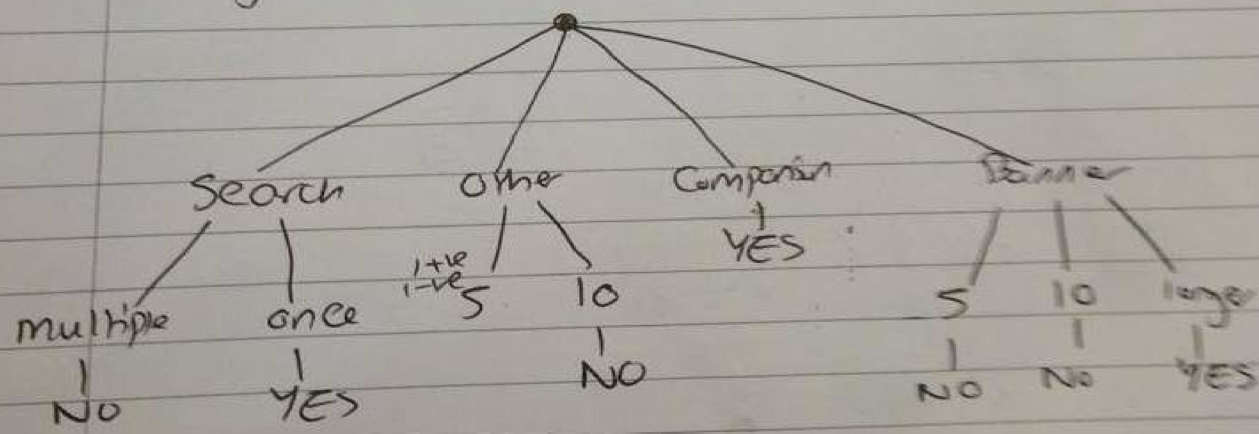
K=4 (longer)

1, 1y, 0n

= 0 (Because $\log_2 1$)

∴ Gain (Length) = 0.811 − 0
= 0.811

Hence, since the Information gain is equal to the entropy (there is no remainder), we have should split on Length.



5 | Other

Entropy (~~[illegible]~~ 5)

2, 1 +ve, 1 -ve

$= -\left(\log_2 \frac{1}{2}\right) = 1$

Gain (Frequency) = 1 − Remainder (Frequency)

k = 1 (multiple)

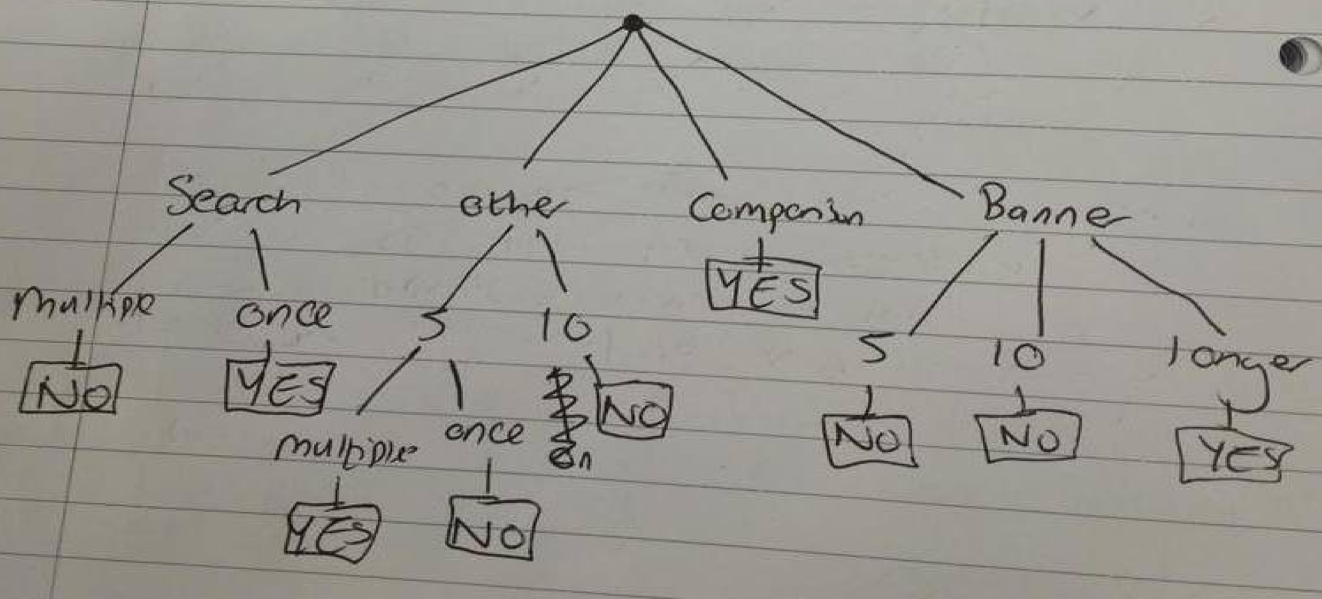
1, 1 +ve, 0 -ve
$= 0$ (Because $\log_2 1$)

k = 2 (once)

1, 0 +ve, 1 -ve
$= 0$ (Because $\log_2 1$)

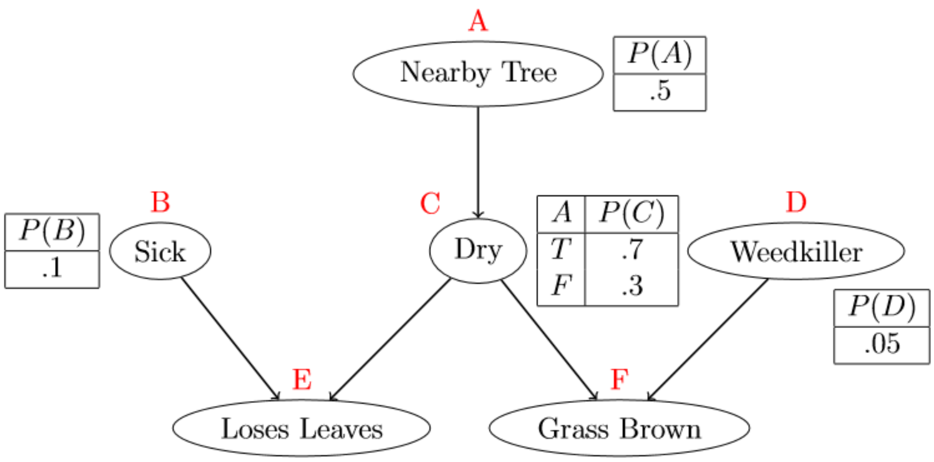
∴ Gain (Frequency) = 1

Since there is no remainder, we should split by frequency.



2. **Bayesian Network** [3%]

Consider the following Bayesian Network about unhealthy plants in your garden.



B	C	$P(E)$
T	T	.9
T	F	.8
F	T	.6
F	F	.05

C	D	$P(F)$
T	T	.99
T	F	.9
F	T	.98
F	F	.1

You can observe that the grass is brown. What is the probability distribution for the plant loosing leaves?

$< 0.2939, 0.2303 >$. Hence $\alpha = 1.9077$ and $\mathbb{P}(E) =< .56, .44 >$.

The important part is that you show your working!

02-12-2015

2.) $\langle 0.2939, 0.2303 \rangle$

hence, $\alpha = 1.9077$ and $\mathbb{P}(E) = \langle .56, .44 \rangle$

$$\sum_A \sum_B \sum_C \sum_D \mathbb{P}(a, b, c, d, E, F)$$

$$\sum_a \sum_b \sum_c \sum_d P(a)P(b)P(d)\,\mathbb{P}(c|a)\,\mathbb{P}(e|b,c)\,\mathbb{P}(f|c,d)$$

$$= \sum_b \sum_c \sum_d P(A)P(b)P(d)\,\mathbb{P}(c|A)\,\mathbb{P}(e|b,c)\,\mathbb{P}(f|c,d)$$

$$+ \sum_b \sum_c \sum_d P(\neg A)P(b)P(d)\,\mathbb{P}(c|\neg A)\,\mathbb{P}(e|b,c)\,\mathbb{P}(f|c,d)$$

$$= \sum_c \sum_d P(A)P(B)P(d)\,\mathbb{P}(c|A)\,\mathbb{P}(e|B,c)\,\mathbb{P}(f|c,d)$$
$$+ \sum_c \sum_d P(A)P(\neg B)P(d)\,\mathbb{P}(c|A)\,\mathbb{P}(e|\neg B,c)\,\mathbb{P}(f|c,d)$$
$$+ \sum_c \sum_d P(\neg A)P(B)P(d)\,\mathbb{P}(c|\neg A)\,\mathbb{P}(e|B,c)\,\mathbb{P}(f|c,d)$$
$$+ \sum_c \sum_d P(\neg A)P(\neg B)P(d)\,\mathbb{P}(c|\neg A)\,\mathbb{P}(e|\neg B,c)\,\mathbb{P}(f|c,d)$$

$$= \sum_d P(A)P(B)P(d)\,\mathbb{P}(C|A)\,\mathbb{P}(e|B,C)\,\mathbb{P}(f|C,d)$$
$$+ \sum_d P(A)P(B)P(d)\,\mathbb{P}(\neg C|A)\,\mathbb{P}(e|B,\neg C)\,\mathbb{P}(f|\neg C,d)$$
$$+ \sum_d P(A)P(\neg B)P(d)\,\mathbb{P}(C|A)\,\mathbb{P}(e|\neg B,C)\,\mathbb{P}(f|C,d)$$
$$+ \sum_d P(A)P(\neg B)P(d)\,\mathbb{P}(\neg C|A)\,\mathbb{P}(e|\neg B,\neg C)\,\mathbb{P}(f|\neg C,d)$$
$$+ \sum_d P(\neg A)P(B)P(d)\,\mathbb{P}(C|\neg A)\,\mathbb{P}(e|B,\neg C)\,\mathbb{P}(f|C,d)$$
$$+ \sum_d P(\neg A)P(\neg B)P(d)\,\mathbb{P}(C|\neg A)\,\mathbb{P}(e|\neg B,\neg C)\,\mathbb{P}(f|\neg C,d)$$
$$+ \sum_d P(\neg A)P(\neg B)P(d)\,\mathbb{P}(C|\neg A)\,\mathbb{P}(e|\neg B,C)\,\mathbb{P}(f|\neg C,d)$$
$$+ \sum_d P(\neg A)P(\neg B)P(d)\,\mathbb{P}(\neg C|\neg A)\,\mathbb{P}(e|\neg B,\neg C)\,\mathbb{P}(f|\neg C,d)$$

OVERLEAF →

$$P(A)\,P(B)\,P(D)\,P(C|A)\,P(e|B,C)\,P(f|C,D)$$
$$+\,P(A)\,P(B)\,P(\neg D)\,P(C|A)\,P(e|B,C)\,P(f|C,\neg D)$$
$$+\,P(\neg A)\,P(B)\,P(D)\,P(C|\neg A)\,P(e|B,C)\,P(f|C,D)$$
$$+\,P(\neg A)\,P(B)\,P(\neg D)\,P(C|\neg A)\,P(e|B,C)\,P(f|C,\neg D)$$
$$+\,P(A)\,P(\neg B)\,P(D)\,P(C|A)\,P(e|\neg B,C)\,P(f|C,D)$$
$$+\,P(A)\,P(\neg B)\,P(\neg D)\,P(C|A)\,P(e|\neg B,C)\,P(f,C,\neg D)$$
$$+\,P(\neg A)\,P(\neg B)\,P(D)\,P(C|\neg A)\,P(e|\neg B,C)\,P(f|C,D)$$
$$+\,P(\neg A)\,P(\neg B)\,P(\neg D)\,P(C|\neg A)\,P(e|\neg B,C)\,P(f|C,\neg D)$$
$$+\,P(A)\,P(B)\,P(D)\,P(\neg C|A)\,P(e|B,\neg C)\,P(f|\neg C,D)$$
$$+\,P(A)\,P(B)\,P(\neg D)\,P(\neg C|A)\,P(e|B,\neg C)\,P(f|C,\neg D)$$
$$+\,P(\neg A)\,P(B)\,P(D)\,P(\neg C|\neg A)\,P(e|B,\neg C)\,P(f|\neg C,D)$$
$$+\,P(\neg A)\,P(B)\,P(\neg D)\,P(\neg C|\neg A)\,P(e|B,\neg C)\,P(f|\neg C,\neg D)$$
$$+\,P(\neg A)\,P(\neg B)\,P(D)\,P(C|\neg A)\,P(e|\neg B,C)\,P(f|C,D)$$
$$+\,P(\neg A)\,P(\neg B)\,P(\neg D)\,P(C|\neg A)\,P(e|\neg B,C)\,P(f|C,\neg D)$$
$$+\,P(\neg A)\,P(\neg B)\,P(D)\,P(\neg C|\neg A)\,P(e|\neg B,\neg C)\,P(f|\neg C,D)$$
$$+\,P(\neg A)\,P(\neg B)\,P(\neg D)\,P(\neg C|\neg A)\,P(e|\neg B,\neg C)\,P(f|\neg C,\neg D)$$

$\cancel{P(A) = .5}$ $\quad$ $\cancel{P(B) = .1}$

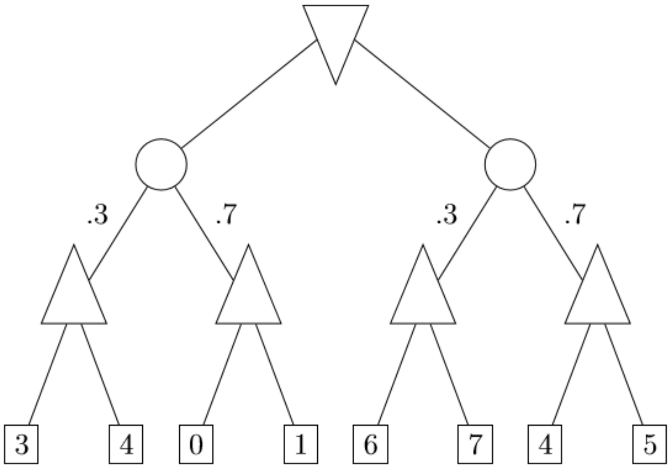
$\cancel{P(C|A) = .7}$ $\quad$ $\cancel{P(D) = .05}$

Hey! I assumed that this was enough to prove it. It didn't make sense to spend another 80000 hours subbing all the values in. Hope that's alright!!!

Thanks! :)

3. **Expected MiniMax** [2%]

Consider the following MiniMax tree with chance nodes.



Downward triangles are Min nodes, upward triangles are Max nodes, circles represent chance nodes, and squares are leafs with utility values. The probabilities given on the outgoing edges from the chance nodes correspond to the likelihood that a particular child node will be selected.

In expected MiniMax the utility value propagation works as follows:

Min Node Takes the minimum of the values of its children.

Max Node Takes the maximum of the values of its children.

Chance Node Takes the sum of the values of its children weighted by the probabilities given on the edges, or

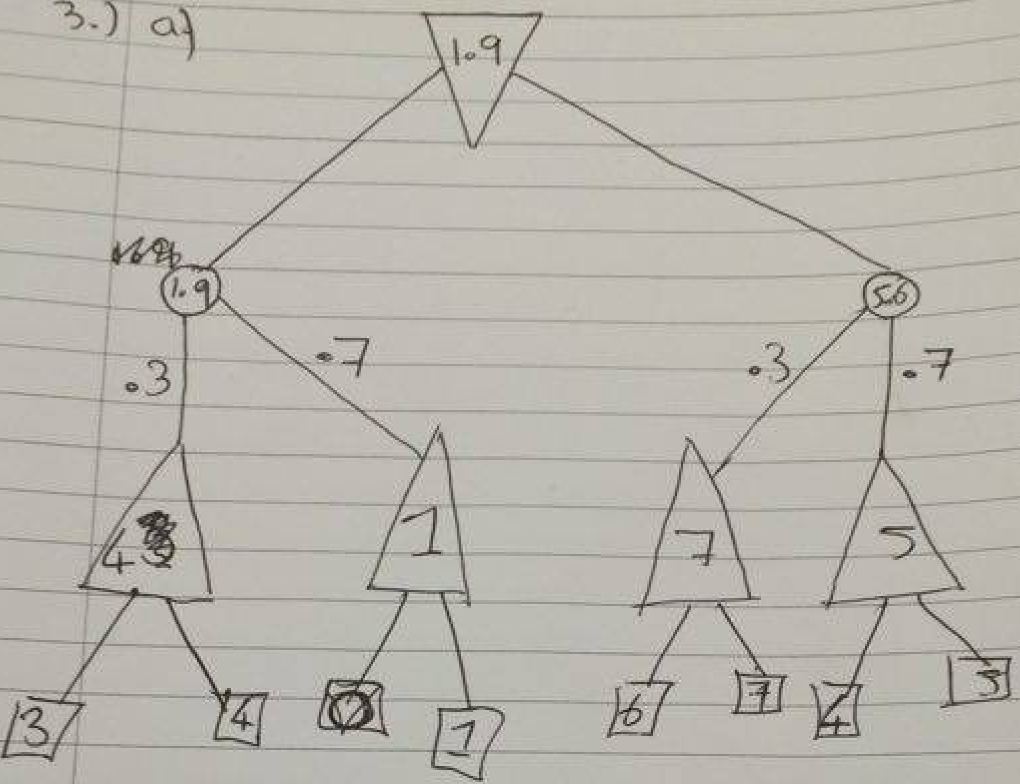
$$\sum_{c \in Children} P(c) \cdot value(c)$$

See also AI:AMA for more details.

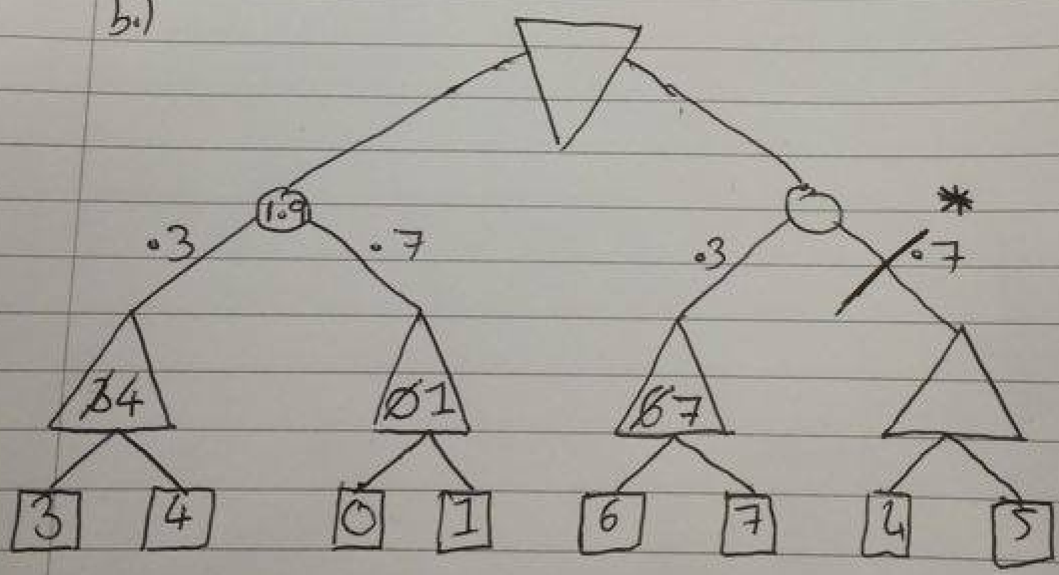
(a) Fill in the utility value for each node computed by the Expected MiniMax algorithm.

(b) Consider that Expected MiniMax uses DFS left to right and that all utility values are non-negative. When using α-β-pruning, is there any part of the tree that does not need to be expanded? Justify your answer.

3.) a)



b.)



* This can be pruned, because $.3 \times 7 > 1.9$, meaning that no matter the value of the max node below, the min node at the top would NEVER take the value