Jiedong Tong

Z5176253

Question 1

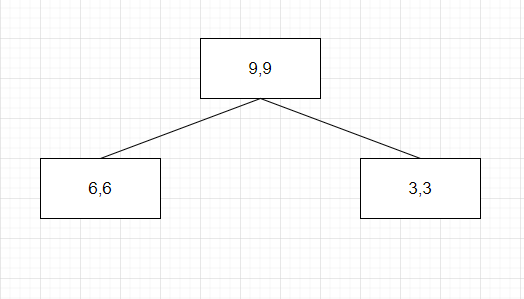
Because we want to use the results from a), b) and c) to compare with the optimal tree, we must find the optimal tree first.

According to what we have learned, we need to maximize the information gain to get the optimal tree.

Proceed as follows:

At first, calculating information gain for all attributes will determine the first node of optimal tree.

1. Split Author:



Parents (9,9) is 9 skips and 9 reads.

Split by known or unknown.

Entropy(parent) = = -(9/18)\*log(9/18) – (9/18) \*log(9/18) = 1

Entropy(known) = = -(6/12)\*log(6/12) – (6/12) \*log(6/12) = 1

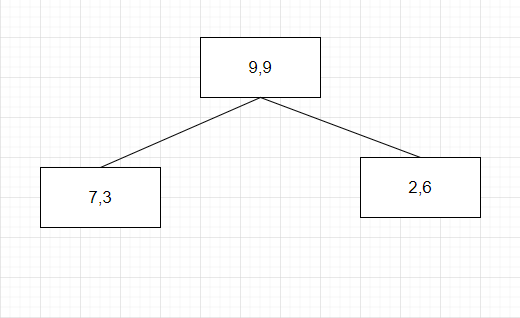
Entropy(unknown) = = -(3/6)\*log(3/6) – (3/6) \*log(3/6) = 1

Average Entropy = 1\*12/18 + 1\*6/18 = 1

Information gain = 1 – 1 = 0

Jiedong Tong

Z5176253

1. Split Thread:

Parents (9,9) is 9 skips and 9 reads.

Split by new or followup.

Entropy(parent) = = -(9/18)\*log(9/18) – (9/18) \*log(9/18) = 1

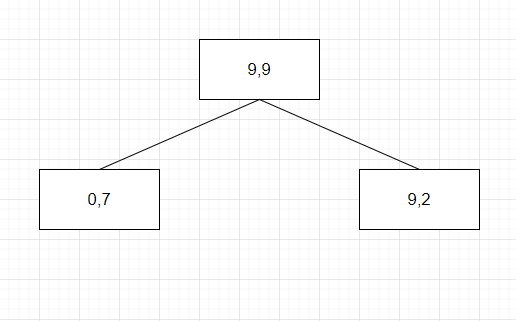
Entropy(new) = = -(7/10)\*log(7/10) – (3/10) \*log(3/10) = 0.88129

Entropy(followup) = = -(2/8)\*log(2/8) – (6/8) \*log(6/8) = 0.81128

Average Entropy = 0.88129\*10/18 + 0.81128\*8/18 = 0.85017

Information gain = 1 – 0.85017 = 0.14983

1. Split Length:



Parents (9,9) is 9 skips and 9 reads.

Split by long or short.

Jiedong Tong

Z5176253

Entropy(parent) = = -(9/18) \*log(9/18) – (9/18) \*log(9/18) = 1

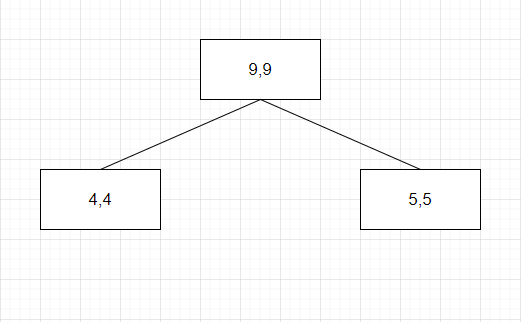
Entropy(long) = = -(0/7) \*log(0/7) – (7/7) \*log(7/7) = 0

Entropy(short) = = -(9/11) \*log(9/11) – (2/11) \*log(2/11) = 0.68404

Average Entropy = 0\*7/18 + 0.68404\*11/18 = 0.41802

Information gain = 1 – 0.41802 = 0.58198

1. Split Where\_read:



Parents (9,9) is 9 skips and 9 reads.

Split by home or work.

Entropy(parent) = = -(9/18) \*log(9/18) – (9/18) \*log(9/18) = 1

Entropy(home) = = -(4/8) \*log(4/8) – (4/8) \*log(4/8) = 1

Entropy(work) = = -(5/10) \*log(5/10) – (5/10) \*log(5/10) = 1

Average Entropy = 1\*8/18 +1\*10/18 = 1

Information gain = 1 – 1 = 0

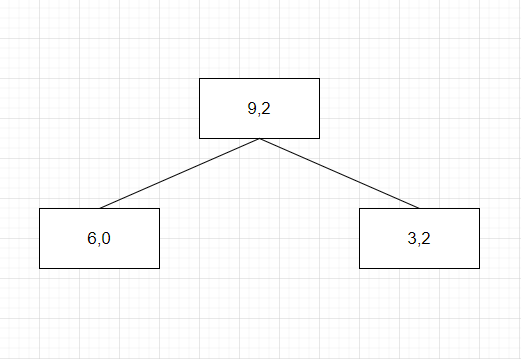
Because of 0. 68404 > 0.14983 > 0 = 0, we should split on length first.

Secondly, calculating information gain of all attributes except to determine the next node to split.

1. Split Author:

Jiedong Tong

Z5176253



Parents (9,2) is 9 reads and 2 skips.

Split by known or unknown.

Entropy(parent) = = -(9/11) \*log(9/11) – (2/11) \*log(2/11) = 0.68404

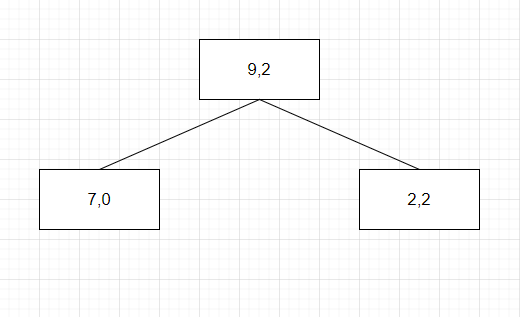
Entropy(known) = = -(6/6) \*log(6/6) – (0/6) \*log(0/6) = 0

Entropy(unknown) = = -(3/5)log(3/5) – (2/5)log(2/5) = 0.97095

Average Entropy = 0\*6/11 + 0.97095\*5/11 = 0.44134

Information gain = 0.68404 – 0.44134 = 0.24270

1. Split Thread:



Parents (9,2) is 9 reads and 2 skips.

Split by new or followup.

Entropy(parent) = = -(9/11) \*log(9/11) – (2/11) \*log(2/11) = 0.68404

Entropy(new) = = -(7/7) \*log(7/7) – (0/7) \*log(0/7) = 0

Entropy(followup) = = -(2/4) \*log(2/4) – (2/4) \*log(2/4) = 1

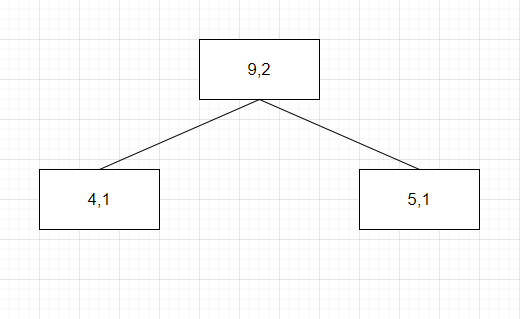
Average Entropy = 0\*7/11 + 1\*4/11 = 0.36364

Jiedong Tong

Z5176253

Information gain = 0.68404 – 0.36364 = 0.32040

1. Split where\_read:



Parents (9,2) is 9 reads and 2 skips.

Split by new or followup.

Entropy(parent) = = -(9/11) \*log(9/11) – (2/11) \*log(2/11) = 0.68404

Entropy(home) = = -(4/5) \*log(4/5) – (1/5) \*log(1/5) = 0.72193

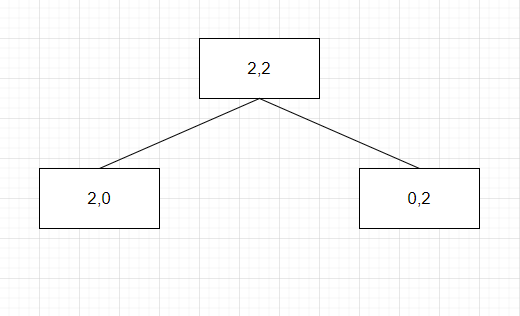
Entropy(work) = = -(5/6) \*log(5/6) – (1/6) \*log(1/6) = 0.65002

Average Entropy = 0.72193\*5/11 + 0.65002\*6/11 = 0.68271

Information gain = 0.68404 – 0.68271 = 0.00133

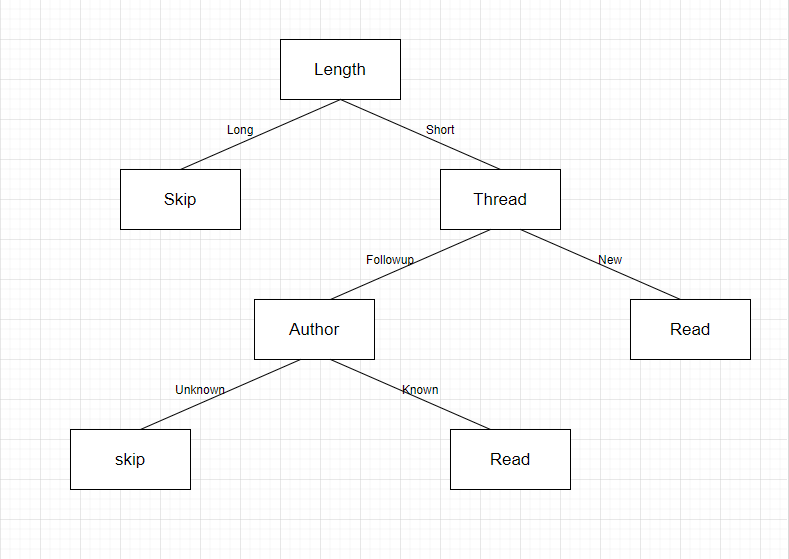
Because of 0.32040 > 0.24270 > 0.00133, we should split on Thread next.

Thirdly, we will split on Author. (after we split on author we will get a clear classification, that’s enough)



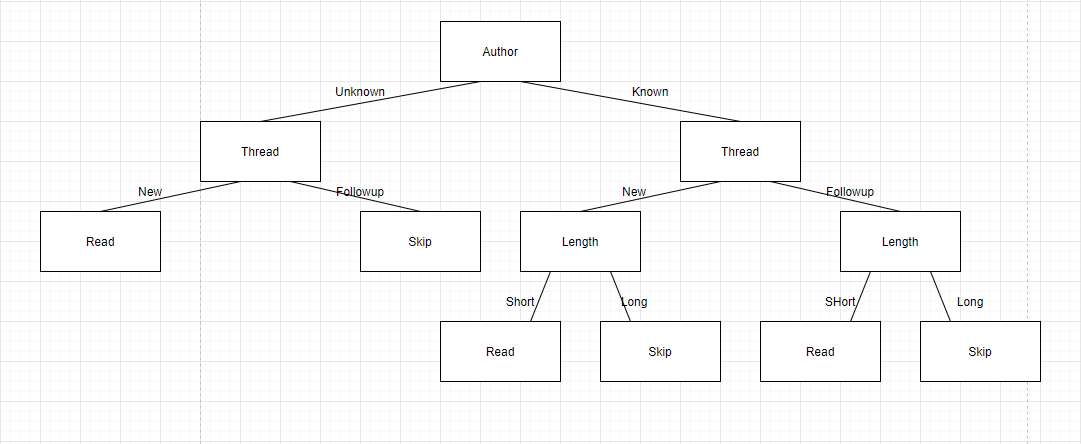
Jiedong Tong

Z5176253

Then we can get optimal decision tree as below:

a)

Tree when the features are in the order [Author, Thread, Length, WhereRead] show as below:



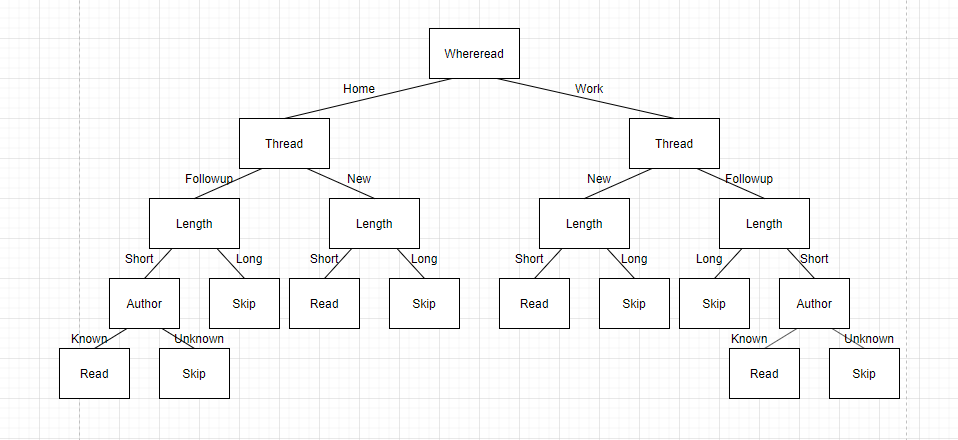
This tree represent a different function than that found with the maximum information gain split. Because if we try to determine Example e19, the optimal tree return “skip” but this tree return “read”. So the function for this tree is different from optimal tree’s.

Jiedong Tong

Z5176253

b)

Tree when the features are in the order [WhereRead, Thread,Length, Author] show as below:



This tree doesn’t represent a different function than that found with the maximum information gain split. Because it will get same results for more examples. However, it is different from the one given for the preceding part.

c)

No trees here correctly classifies training sets but represents a different function than those found by the preceding algorithms.

We can use logical expression to prove this conclusion.

About optimal tree, the logical expression is:

((known ∨ new) ∧ short)

About tree in a), the logical expression is:

(known ∧ new ∧ short) ∨ (known ∧ (¬new) ∧ short) ∨ (¬known ∧ new)

* (known ∧ short) ∨ (¬known ∧ new)

Jiedong Tong

Z5176253

About tree in b), the logical expression is:

((known ∨ new) ∧ short)

It’s obviously there are 3 features and can combined into 6 different decision trees.

However some combinations will represent same thing in logical expression. After we remove those same decision tree functions there are only 2 possible decision tree functions. [Author, Thread, Length], [Thread, Length, Author]. Then we can find out there are no tree correctly classifies the training examples but examples but represents a different function than those found by the preceding algorithms.

Question2

I used Jupyter Notebook to finish this question, please check it in another file named COMP3411 assignment2 - Jupyter Notebook.pdf.