NUMB3RS Activity: Are You Sure?
Episode: "Soft Target"

Topic: Mathematical Prediction **Grade Level**: 9 - 12
Objective: Introduce Decision Trees
Time: 20-30 minutes

Introduction

In "Soft Target," Charlie uses Linear Discriminant Analysis (LDA), a multiple regression method, to predict the next terrorist attack. This method, like many others used to predict events, may use a tree diagram to visualize possibilities. The use of tree diagrams is seen in probability.

Students are familiar with finding the probability of tossing a head or tail in a simple coin toss experiment. The probability of flipping a head, for example, is $\frac{1}{2}$. If you ask students the probability of tossing the coin twice and getting a head both times, they may list the four possible outcomes, TT, TH, HT, HH, each with probability $\frac{1}{4}$. If a coin is tossed eight times, a tree with all outcomes is tedious to construct.

The Coin Toss Tree diagram on the next page shows the outcomes and probabilities for two coin tosses. Because each toss is independent, the probability of tossing two heads is found by multiplying the probabilities along the two far left branches, leading to HH.

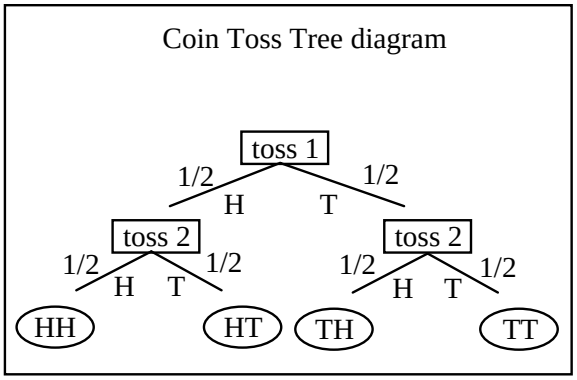
$$P(two\ heads) = \frac{1}{2} \bullet \frac{1}{2} = \frac{1}{4}$$

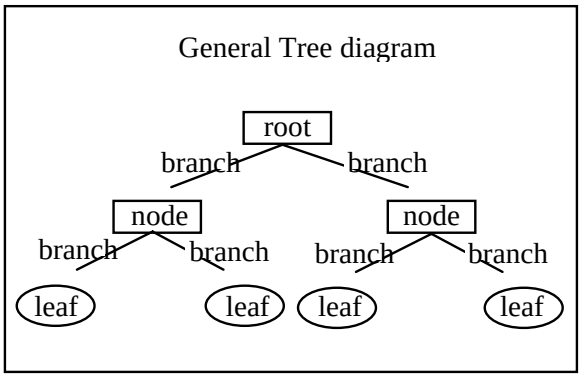
The tree diagram allows exploration of conditional probability concepts using such questions as—What is the probability of tossing a head given that your first toss was a head? In the tree where a head shows on the first toss, you can go directly to toss 2 and see that the probability is $\frac{1}{2}$.

Depending on what information you are given or assuming, you can focus on different parts of the tree diagram. In the given case, we ignored the outcomes with the first toss being tails. Considering certain parts of the tree diagram, given various information, is used in predicting events such as the next terrorist attack.

The Coin Toss Tree diagram can be generalized as shown in the General Tree diagram. All nodes, including the root, describe an event or attribute. The branches contain outcomes, probabilities, or both. Each leaf shows the final outcome of traversing along a path leading to that leaf. Beginning at the root and moving along a path to a single leaf describes a series of events with a particular result.

Terry Souhrada (retired), The University of Montana, Missoula, MT





Discuss with Students

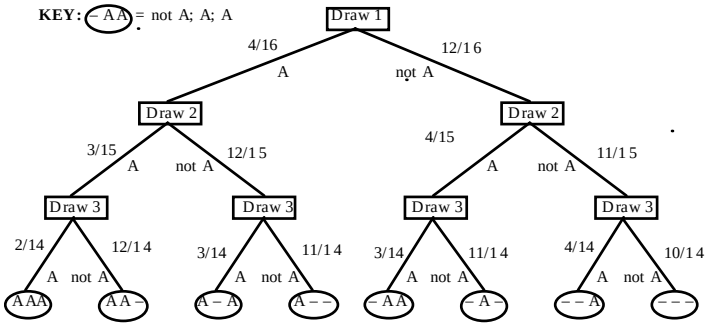
Discuss the Coin Toss Tree diagram and some of the probabilities associated with the branches. Then say to students: in "Soft Target," you are given information that a terrorist is planning to bomb two targets. There are five government buildings and three public buildings, for a total of eight possible targets.

1. What is the probability that the first target is a government building?
2. Given that the first target is a public building, what is the probability that the second target will also be a public building? Explain.

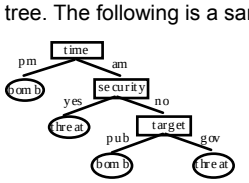
Discuss with Students answers: 1. 5/8 **2.** the probability of the second target being a public building given that the first was a public building is 2/7; the probability of either of the government buildings being bombed is near 0.

Student page answers:

1. a.



b. approximate probabilities AAA, (0.007);AA –, (0.043); A – A, (0.043); A – –, (0.157); – AA, (0.043); – A –, (0.157); – – A, (0.157); – – –, (0.393); **c.** $P(AAA) + P(AA–) + P(A–A) + P(–AA) \approx 0.136$ **2. a**. only a threat; **b.** security; Regardless of the security, the tree always gives correct prediction when the target type is a government building and the time of day is known; **c.** target type; **d.** no; **e.** about 91%, answers will vary, most will probably answer yes; **3. a.** strategies will vary**; b-d**. answers will vary based on the student's tree. The following is a sample tree that correctly predicts the given information.



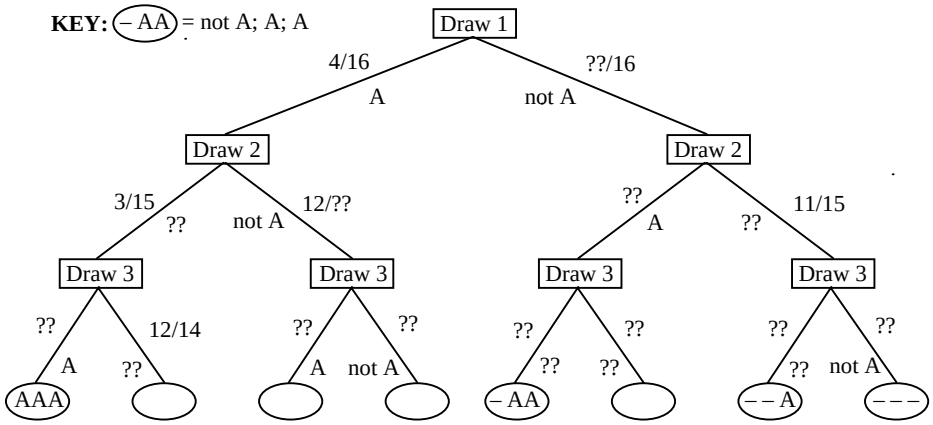
Terry Souhrada (retired), The University of Montana, Missoula, MT

Name _____ Date _____

NUMB3RS Activity: Are You Sure?

Charlie introduces Don to decision trees by using an analogy with cards. He shows Don how you can use a tree diagram to calculate the probability of certain outcomes. He starts with a stack of 16 cards—4 Jacks, 4 Queens, 4 Kings, and 4 Aces. He takes 3 cards out of a pile, one at a time, and notes whether the card is an Ace or not. After he removes a card, he does not replace it in the stack.

Charlie points out that the probability of drawing an Ace changes after each step and depends on what cards were previously drawn. His given tree for drawing the cards is partially complete.



1. a. Fill in the missing probabilities in the spaces provided on the branches of the tree.

b. Find the probabilities for each final outcome (leaf). Write your answers in the space provided beneath each leaf. Round to the nearest thousandths.

c. What is the probability of drawing at least two Aces in three draws?

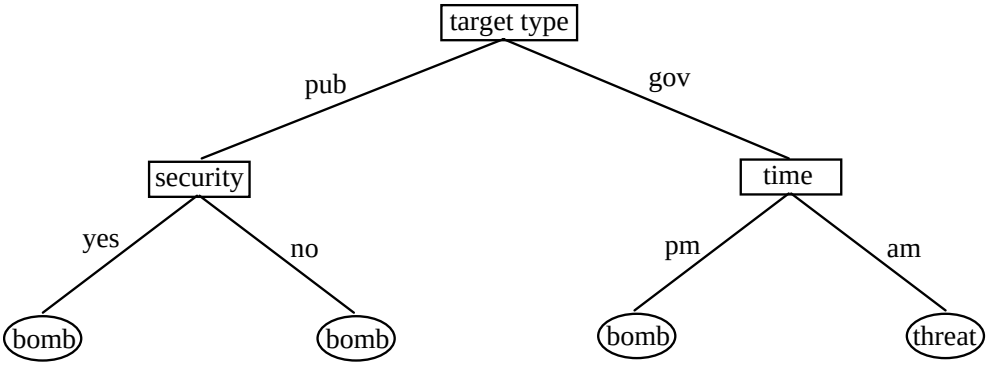
Charlie wants to make a tree diagram from data about the last 11 bomb threats to show the final results. The table shows whether there was security at the location of the threat, the time of day of the threat (morning or afternoon), the type of building (public or government), and the outcome (whether there really was a bomb or if it was a false threat).

	security	**time of day**	**target type**	***outcome***
1	yes	afternoon	public	bomb
2	yes	afternoon	government	bomb
3	no	afternoon	government	bomb
4	yes	morning	government	false
5	no	morning	public	bomb

Terry Souhrada (retired), The University of Montana, Missoula, MT

	security	time of day	target type	*outcome*
6	yes	morning	government	false
7	yes	afternoon	public	bomb
8	no	afternoon	public	bomb
9	no	morning	government	false
10	no	morning	public	bomb
11	yes	morning	public	false

Charlie used the information to create the decision tree below. There are no probabilities assigned to the branches, only choices. Charlie tests the tree by following each path using the information in the table to see if the tree leads to the known final outcome.



2. **a**. The FBI receives a tip that a bomb is going to be placed at a government building in the morning. Based on the tree, do you think there really is a bomb?

b. When the target type is a government building what piece of information appears of little importance?

c. What piece of information appears to be the most important?

d. Test the decision tree using the information in the table. Does the outcome in all the leaves match the outcomes in the table?

e. What percentage of the outcomes does the tree correctly predict? Would you use this tree to make a decision?

3. In information theory, it is known that larger decision trees are typically less accurate than smaller ones.

a. Create a more accurate tree in which every decision in the table is correctly predicted.

b. Use the data in the table to "test" your tree from **part a** to see how accurately it predicts the known events.

c. What percentage of the time is your tree accurate?

d. Suppose Charlie used your tree to make a report to the FBI. What information would you have him include in the report based on your tree?

Terry Souhrada (retired), The University of Montana, Missoula, MT

The goal of this activity is to give your students a short and simple snapshot into a very extensive math topic. TI and NCTM encourage you and your students to learn more about this topic using the extensions provided below and through your own independent research.

Extensions

Decision trees and the theory behind them are not only used to make predictions but are used heavily in learning machines and artificial intelligence. Existing algorithms are used to create highly accurate decision trees. Most of them use **entropy** as the basis for deciding the order in which the tree nodes appear. Entropy is a measure of the confusion found in a set of information. It represents the lack of clear information about the situation being modeled. The lower the entropy, the more clear information we gain. There are many web sites that discuss these topics. Some have applets that allow you to create decision trees based on any given information. One such applet can be found and downloaded from the web page

http://www.cs.ubc.ca/labs/lci/CIspace/Version4/dTree/

- Enter the information from the table into the applet to see if the applet produces the same tree that you found above.
- Use the applet to create a decision tree to predict the outcome of an upcoming sporting event and give a report on the results. Remember, the more known information you have to work with the better the tree will predict the outcome.

Using entropy to create decision trees does not come without some shortcomings. Investigate them and explore other methods for creating a decision tree, such as **gain ratio**, **information pruning**, and the **gini coefficient**. Compare the result using these methods to the trees you created earlier.

Other Resources

Quantum Information Theory: Entropy
http://members.aol.com/jmtsgibbs/entropy.htm

Decision Trees
http://www.aaai.org/AITopics/html/trees.html

Decision Tree Learning
http://www2.cs.uregina.ca/~hamilton/courses/831/notes/ml/dtrees/4_dtrees1.html

Terry Souhrada (retired), The University of Montana, Missoula, MT