CS 4376/5376

Homework Assignment #2

Decision Theory

DUE: September 27 at 11:59 PM

The assignment should be turned in using blackboard as either a word document or pdf file. Scanned work is acceptable, as long as it is legible.

1) Suppose you are the leader of a gang of cyber-criminals that is using ransomware to extort payments from victims. You are trying to decide between three different pricing strategies. That is, how much should you charge a victim for the key to decrypt their data once you have encrypted their data? You can charge a high ($100,000), medium ($25,000), or low price ($1000). Higher prices will be accepted with a lower probability. If the victim chooses not to pay, you get a value of $0. Different types of victims will pay each price with a different probability, as shown in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Large Business | Small Business | Home User |
| High Ransom | 0.1 | 0.05 | 0 |
| Med Ransom | 0.2 | 0.3 | 0 |
| Low Ransom | 0.3 | 0.8 | 0.5 |

In addition, you know that you that your victims fall into these categories with the following probabilities:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Large Business | Small Business | Home User |
| Probability | 0.05 | 0.15 | 0.8 |

1. Write down a payoff table for this problem by calculating the expected amount of profit in each of the nine cases above (e.g., High Random and Large Business). This reflects an assumption that you will average over many victims, so the expected value is a good approximation of your payoff for each case. You will use this payoff table for the rest of the analysis below.

|  |  |  |  |
| --- | --- | --- | --- |
| Ransom vs. Business | Large | Medium | Small |
| High | 500 | 750 | 0 |
| Medium | 250 | 1125 | 0 |
| Low | 15 | 120 | 400 |

1. Calculate the expected value for each action and identify the best action according to this criterion.

E(High) = 500 + 750 => 1250

E(Med) = 1375

E(Small) = 535

Choose Medium Ransom

1. Calculate the maxmin value and the best action.

Maximize the worst-case outcome

Worst Case for High => 0

Worst Case for Med => 0

Worst Case for Low => 15

Choose Low

1. Write down the opportunity loss (regret) table for this problem

|  |  |  |  |
| --- | --- | --- | --- |
| Ransom vs. Business | Large | Medium | Small |
| High | 500 | 750 | 0 |
| Medium | 250 | 1125 | 0 |
| Low | 15 | 120 | 400 |

*Calculating regret: for each outcome, take the value of the best*

*outcome for this state and subtract the outcome for this action*

|  |  |  |  |
| --- | --- | --- | --- |
| Ransom vs. Business | Large | Medium | Small |
| High | 0 | -375 | -400 |
| Medium | -250 | 0 | -400 |
| Low | -485 | -1005 | 0 |

1. Calculate the expected regret for each action and identify the action with minimum expected regret.

|  |  |  |  |
| --- | --- | --- | --- |
| Ransom vs. Business | Large | Medium | Small |
| High | 0 | -375 | -400 |
| Medium | -250 | 0 | -400 |
| Low | -485 | -1005 | 0 |

Multiply times the probabilities:

E(High) = 0 + 0.15\*(-375)+0.8(-400)=>-376.25

E(Med) = -250(0.05) + 0 + 0.8(-400) => -332.5

E(Low) = -485(0.05) + 0.8(-1005) + 0 => -828.25

Action with minimum expected regret is Medium

1. Calculate the minmax regret and best action.

|  |  |  |  |
| --- | --- | --- | --- |
| Ransom vs. Business | Large | Medium | Small |
| High | 0 | -375 | -400 |
| Medium | -250 | 0 | -400 |
| Low | -485 | -1005 | 0 |

Maximum regret for High is -400

Maximum regret for Medium is -400

Maximum regret for Low is -1005

Choose either Med or High

1. Calculate the value of having perfect information about what type of user is compromised before you make the pricing decision.

|  |  |  |  |
| --- | --- | --- | --- |
| Ransom vs. Business | Large | Medium | Small |
| High | 500 | 750 | 0 |
| Medium | 250 | 1125 | 0 |
| Low | 15 | 120 | 400 |

Expected Utility High => 1250

Expected Utility Medium => 1375

Expected Utility Low => 535

EVPI = 2025 – 1375 = 650

2) A company is considering whether or not to buy cybersecurity insurance or not, and also whether or not to hire additional cybersecurity experts to harden their systems. An insurance policy that covers any possible loss up to a maximum of $5,000,000 would cost the company an annual premium of $500,000. Hiring additional cybersecurity experts would cost $1,000,000 per year. The probability of being attacked in any given year is 30%. The expected damage of an attack without hiring additional experts is $10M, while the expected loss with the additional experts is only $3M. Note that any insurance payout would be deducted from these losses.

1. Draw a complete decision tree for this problem, assuming the company makes a decision on insurance first, then whether to hire the experts, and only finds out after these decisions whether the attack happens or not.

A picture containing chart

Description automatically generated

1. Find the solution that maximizes the expected value.

When you have insurance and you hire an expert the cost is $-1.5M

When you have insurance but no expert the cost is => $-(5.5M\*0.3 + 0.5\*0.7) => $-2M

When you have no insurance and an expert cost is => $-(4M\*0.3 + 1M\*0.7) => $-1.9M

When you have no insurance and no expert the cost is $-3M

Hire the expert and buy the insurance to have the best possible expected value.

c) Look up the phrase “moral hazard” and give a 1-2 sentence description of this problem.

Since I know that the risks of keeping my current situation will be assumed by the companies that provide me insurance or that trust my judgement, should I do something about the risk or just let it happen?

1. Briefly describe how moral hazard applies to this decision problem.

Basically, the company could choose to just go with the insurance and not hire an expert to oversee the way operations are handled, so the insurance would respond for any issue.

3) You have received a security alert, and are considering two possible analysts to assign to investigate the alert. One is inexperienced, and has a salary of $25 per hour. The second is more experienced, and has a salary of $100 per hour. Both analysts will take around 2 hours to investigate. The inexperienced analyst has a higher error rate of 3% (both false positives and false negatives), while the experienced analyst has an error rate of only 1%. A false positive costs the company $500 in additional investigation time, while a false negative costs the company $50,000 in estimated costs for missing the initial signs of the attack. True positives and negatives both have a cost of 0. Based on prior experience, the probability that the alert corresponds to a real attack is 2%. Compute the expected utility for assigning each analyst to the alert. Which should you choose if you are maximizing expected utility?

False positive (tagged as attack but not really an attack)

False negative (tagged as not attack but really an attack)

True positive (tagged as attack and it is an attack)

True Negative (tagged as attack and it is not attack)

Programmer 1 => 25 dollars/hours ~ P(error rate) = 0.03 (False positive, False Negative)

Programmer 2 => 100 dollars/hour ~ P(error rate) = 0.01 (False positive, False Negative)

P(Attack|alert) = 0.02 => P(Attack ^ Alert)/P(Alert) => P(Alert|Attack)\*(P(Attack)/P(Alert)) => P(Alert|Attack)P(Attack)/ (P(Attack|Alert)\*P(Alert) + P(Attack|’Alert)\*P(‘Alert))

|  |  |  |
| --- | --- | --- |
| Programmer 1 | Positive  (Actual) | Negative  (Actual) |
| Positive | -50 | -550 |
| Negative | -50,050 | -50 |

|  |  |  |
| --- | --- | --- |
| Programmer 2 | Positive  (Actual) | Negative  (Actual) |
| Positive | -200 | -700 |
| Negative | -50,200 | -200 |

U(Programmer 1) = 0.97( -50) + 0.03(0.02\*(-50,050)+0.98(-550))

U(programmer 2) = 0.99(-200) + 0.01(0.02\*(-50,200)+0.98(-700))

U(programmer 1) => -94.7

U(programmer 2) => -214.9

Hire programmer 1

4) You are considering two different intrusion detection systems and want to use the expected value of information to figure out which is a better value for the cost. For each system you scan with the IDS, your prior probability that there is an intrusion is 5%. Your expected cost for a false negative (missing an intrusion) is $25,000. Your expected cost for a false positive (a false alarm) is $500 in time to investigate the alert.

For IDS1, P(Alert|Intrusion) = 0.98, and P(Alert|No Intrusion) = 0.1.

For IDS2, P(Alert|Intrusion) = 0.9, and P(Alert|No Intrusion) = 0.01.

IDS1 costs $500 per system, while IDS2 costs $200 per system.

Prior probability => P(Intrusion) = 0.05

1. Use Bayes rule to calculate P(Intrusion|Alert) for IDS1.
   1. A picture containing drawing

      Description automatically generated
   2. P(Intrusion|Alert) = P(Alert|Intrusion) \* P(Intrusion) / (P(Alert|Intrusion)\*P(Intrusion) + P(Alert|No Intrusion)\*(1-P(Intrusion)) = > (0.98\*0.05)/((0.98\*0.05)+(0.1\*0.95)) ~ 0.34027
2. Use Bayes rule to calculate P(Intrusion|Alert) for IDS2.
   1. A picture containing drawing

      Description automatically generated
   2. P(Intrusion|Alert) = P(Alert|Intrusion) \* P(Intrusion) / (P(Alert|Intrusion)\*P(Intrusion) + P(Alert|No Intrusion)\*(1-P(Intrusion)) = > (0.9\*0.05)/((0.9\*0.05)+(0.01\*0.95)) ~ 0.8256
3. What is the expected value of the information provided by IDS1, compared to missing all of the intrusions?
   1. Expected Value of Missing = 0.05\*25000 => -1250
   2. Expected Value of IDS1 = 0.05\*(0.02\*25000)+0.95\*(0.1\*500) => -72.5
4. What is the expected value of the information provided by IDS2, compared to missing all of the intrusions?
   1. Expected Value of Missing = 0.05\*25000 => -1250
   2. Expected Value of IDS2 = 0.05\*(0.1\*25000)+0.95\*(0.01\*500) => 129.75
5. Accounting for all costs, which IDS should you buy?
   1. IDS 2 because it has the best expected value between

5) a) Write pseudocode for a **recursive** method that will calculate the expected value of a decision tree, assuming that all decisions are made by maximizing the expected utility. Your method should return a double value representing the expected utility, and take in as a parameter a reference to the root node in the tree.

Assume that you have a DecisionTreeNode class already implemented the has the following functionality (you do not need to provide pseudocode for this class). First, there are three types of nodes: Leaf, Chance, or Decision. The getType() method returns this type.

A Leaf node has no children, but has a utility value that can be accessed by calling the getUtility() method on the node. A Chance node represents a random outcome that is chosen by nature according to a known probability distribution. It has two methods, getChildren() which returns a list of DecisionTreeNode objects that are the children of this node, and getProbabilities() that returns a list of the probabilities of each child being selected (in the same order as the getChildren() method). Finally, a Decision node has a list of children that can be accessed by calling the getChildren() method. Calling the getUtility() method on any node that is not a Leaf results in an error.

def Utility(root):

if getType(root) == Leaf:

return getUtility(root)

else if getType(root) == Decision:

max = -inf

for child in getChildren(root):

u = Utility(child)

if u > max:

max = u

return max

else if getType(root) == Chance:

max = -inf;

children = getChildren(root);

probabilities = getProbabilities(root);

for i in range(len(children)):

u = getUtility(children[i]) \* probabilities[i];

if u > max:

max = u

return max

b) Briefly describe the changes you would need to make to your code to minimize maximum regret for the decision tree instead.

At every branch that is immediately before the leaves we would compute the best outcome and then from all the leaves apply a subtraction by to the best outcome by the element of a leaf under consideration. After that we would override the value at the leaf with the computed result, then we would select the highest regret among the leaves, and at every other level we would return the minimum of calling recursively the same algorithm.