Introduction to AI 2021 Final Study Guide

Question 1: Planning

Consider the following planning domain and problem:

Description of problem and domain

This will be something like, but definitely different than: "We live in a world with three types of things: Cargos, Airports, Planes. We have two Cargo objects, c1 and c2, two plane objects, p1 and p2, and two airports, MSY and ATL."

Initial State: ? *// Some arrangement of the stuff outlined in the problem description.*

Action Descriptions: *You will be given the following for some number of actions:*

Action: ? *//The name and parameters of an action you can take.*

Precondition: ? *//The thing(s) which must be true before you can take this action.*

Effect: ? *// The thing(s) which are now true after taking this action.*

Goal: ? *// specifying certain things you wish to be true in the final state.*

All of the above will be given to you. You then are responsible for filling in the remaining blanks.

Give the shortest plan which will satisfy the goal. If your plan is fewer than ? steps, leave the unused spaces blank:

1. _____
2. _____

…

?. _____

(the idea of the above is that you will just be creating the plan by hand, using your smart human brain. You don't need to worry about deriving it using any of the algorithms we covered in class).

Suppose we use the HSP heuristic to estimate how many steps will be required to solve this problem. For each literal below, fill in the cost estimated by HSP. "Initial Values" are the costs before considering any steps. "After First Iteration" are the costs after considering every possible step once, and "After Second Iteration" are the costs after considering every possible step twice.

Literal	Initial Value	After First iteration	After Second Iteration

Based on those cost values, how would HSP estimate the cost of the following expressions? (Please use the additive version of HSP discussed in class – there are other versions some of you used in your programming assignment. That's great, but additive here please!)

Expression_one Estimated Cost: _____

Expression_two Estimated Cost: _____

Expression_three Estimated cost: _____

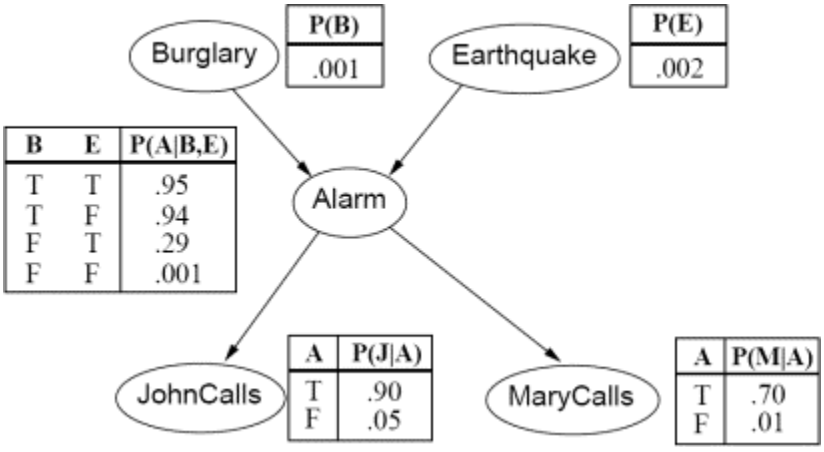
Question 2: Bayesian Networks and Probabilistic Inference

Consider the following problem:

This will be the description of the problem. It will be something like (but definitely different than) this: "Suppose you have two well meaning neighbors, John and Mary. They have both agreed to call you when they hear your security alarm go off. Your security alarm is meant to go off when there is a burglary, but sometimes it goes off when there is an earthquake. But sometimes your neighbors don't call you when it does go off, and other times they call you even when it hasn't gone off."

Now, suppose the following Bayesian Network is constructed to represent this problem:

You will be given a full Bayesian network, and all of the conditional probability tables for each node. So something like…



Use the above network to calculate the following probabilities. Round your answers to 2 decimal places or give them as percentages:

P(*?*) = _____ English Description of what the probability is asking for

Things you should feel confident about before tackling this question:

To be able to go in the 'causal' direction (e.g., given the values for variables of a parent node, what is the probability of variables being true or false for a child node).

To be able to go in the 'diagnostic' direction (e.g., given the values for variables of a child node, what is the probability of variables being true or false for a parent node).

Question 3: Decision Tree Induction

Consider the data set below:

You will be presented with a table of examples. The index, attributes, and class label for each example will be clearly marked.

Calculate the following values. Round your answers to two decimal places:

Entropy of the data set before any splitting: _____

You'll then be asked the following for some number of attributes:

Consider the ??? attribute as the root of the tree:

Entropy of having the ??? attribute: _____

Entropy of not having the ??? attribute: _____

Weighted average entropy of children: _____

Information gain for ??? attribute: _____

Based on these calculations, which attribute should be the root? _____

You may find the following equations and formulae helpful to you in solving this question:

The entropy of a variable V with possible values $\{v_1, v_2, \dots, v_n\}$ can be expressed in bits as:

$$entropy(V) = \sum_{v \in V} -P(v) \log_2\big(P(v)\big)$$

You can calculate the base 2 logarithm of any number x using this formula:

$$\log_2(x) = \frac{\log_{10}(x)}{\log_{10}(2)}$$

The base 10 logarithm is $\log_{10}(10^x) = x$ (i.e., what power do you raise 10 by to get the input. E.g., $\log_{10}(100) = \log_{10}(10^2) = 2$ (you raise 10 to the second power to get 100).

Here are some common base 2 logarithm values:

$$\log_2\left(\frac{1}{4}\right) = -2.00 \qquad \log_2\left(\frac{1}{3}\right) = -1.58 \qquad \log_2\left(\frac{1}{2}\right) = -1.00$$

$$\log_2\left(\frac{2}{3}\right) = -0.58 \qquad \log_2\left(\frac{3}{4}\right) = -0.42 \qquad \log_2(1) = 0.00$$

Question 4: Unsupervised Learning

Consider the data set below:

You will be given a table that has some number of examples in it. The examples will each have an index and two attributes, but will have no class label.

Use the k-Means (Lloyd's) algorithm to cluster these observations into k = 2 clusters.

Assume the observations are initially classified randomly as follows:

Cluster 1: ??? *// List of examples from the table in cluster 1, given to you.*

Cluster 2: ??? *// The remaining examples, also given to you.*

Calculate the centroids of the clusters, rounding to 2 decimal places:

Cluster 1: [X = _____, Y = _____]

Cluster 2: [X = _____, Y = _____]

How will the observations be reclassified given the above centroids?

Cluster 1: _____ *// Now *you* list off all the examples in cluster 1.*

Cluster 2: _____ *// and you list off all the examples in cluster 2.*

Recalculate the centroids of the clusters, rounding to 2 decimal places:

Cluster 1: [X = _____, Y = _____]

Cluster 2: [X = _____, Y = _____]

Has the algorithm converged, or does it need additional iterations? _____

You may find the following formula for measuring Euclidean distance between two n-ary vectors *X* and *Y* (i.e, "The Distance Formula") helpful:

$$d(X,Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \cdots + (X_n - Y_n)^2}$$

Question 5: Lazy Learning

Consider the data set below:

You will be given a table that has some number of examples in it. The examples will each have an index, some number of attributes, and a class label.

Use the k Nearest Neighbor classification technique to classify the following new observation according to Euclidian distance:

You will be given "one new row" for the above table, i.e., a new example, but it will notably be missing a value for the class label.

For reference, the Euclidean distance between two n-ary vectors *X* and *Y* is:

$$d(X,Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \cdots + (X_n - Y_n)^2}$$

List the 3 nearest neighbors for the new observations, and their distances from the new observation (rounded to 2 decimal places).

Nearest neighbor: _____ Distance: _____

Second nearest neighbor: _____ Distance: _____

Third nearest neighbor: _____ Distance: _____

Based on simple majority vote, how would 3NN classify this new observation? _____

Would this classification be different if you used the weighted-voting mechanism we described in class? _____

Extra Credit:

There will be some number of questions directly pertaining to material covered in the first half of the semester. Given that it is extra credit, I am going to leave the format of these questions a surprise; you should be focusing your studying efforts on the previous question types!

And, of course, the most important study tip:

Before the exam, do your best to get a good sleep the night before, and eat well before taking the test!

I'll be there in the "room" with you, so you can ask questions if you have any! But, with this study guide in hand, I'm hoping that nothing on the exam should be too surprising!

It has been a pleasure working with you all this semester! I hope you have a great (and safe) summer vacation! And to those of you graduating: congratulations! Please, go out there and make the world a better place!

The world could really use your help.