Fall 2022 - CSC545/645 Artificial Intelligence - Assignment 8

Due date: Tuesday, November 8, 2022, 10:30am. Please create a folder called assignment8 in your local working copy of the repository and place all files and folders necessary for the assignment in this folder. Once done with the assignment, add the files and folders to the repo with svn add *files, folders* and then commit with svn ci -m ‘‘SOME USEFUL MESSAGE’’ *files, folders* .

Exercise 8.1 [20 points] Read chapter ”Probabilistic Reasoning” of the textbook.

1. **You are a witness of a night-time hit-and-run accident involving a taxi in Athens. All taxis in Athens are blue or green. You swear, under oath, that the taxi was blue. Extensive testing shows that under the dim lighting conditions, discrimination between blue and green is 75% reliable.**

   (a) Is it possible to calculate the most likely color for the taxi? (Hint: distinguish carefully between the proposition that the taxi *is* blue and the proposition that it *appears* blue.)

   We can split this up into two different variables;

   T represents the color that the taxi actually is, blue or green.
   W represents the two options that the could witness says they saw;

	P(w=g | t=g) = .75	so therefore	P(w=g | t=b) = .25
also	P(w=b | t=b) = .75	so therefore	P(w=b | t=g) = .25

   In order to calculuate just a probability for a certain P(t | w) we need more information, so with just this statement we cannot conculde anything concrete.

   (b) What now, given that 9 out of 10 Athenian taxis are green?

   Now we can find P(t=b | w=b) = $\frac{P(t=b)\ n\ P(w=b)}{P(w=b)} = \frac{P(w=b \mid t=b) * P(t=b)}{P(w=b \mid t=g) * P(t=g) + P(w=b \mid t=b) * P(t=b)}$

   $= \frac{.75 * .1}{.25 * .9 + .75 * .1} = \frac{.75 * .1}{.25 * .9 + .75 * .1} = .25$

   A 25% chance the taxi is blue given the witness claimed they saw a blue taxi.

2. Text categorization is the task of assigning a given document to one of a fixed set of categories, based on the text it contains. Naive Bayes models are often used for this task. In these models, the query variable is the document category, and the "effect" variables are the presence or absence of each word in the language; the assumption is that words occur independently in documents, with frequencies determined by the document category.

(a) Explain precisely how such a model can be constructed, given as "training data" a set of documents that have been assigned to categories.

You can use the training data to generate probabilites of certain events like we did in the above problem. We would have two variables;

C represents the categories we are trying to distinguish between
$W = [w_1, w_2, \ldots, w_n]$ represents a vector of all the words in the language we are using to distinguish, which can each be true or false.

So this way we have P(C = c) where c is a specific category, which would could be calcuated just by how much of our training data is of type c. Similarly we know $P(w_i=t)$ and $P(w_i = t \mid C = c)$ by simply counting from our training data. Using these probabilites we can then calculate $P(C=c \mid W)$

(b) Explain precisely how to categorize a new document.

Since the probability of words is independent under our assumptions, this is the same as

$P(C = c \mid W)$ = $P(C = c)$ * [$P(w_i \mid C = c)$ for all w_i in W whether they are true or false]

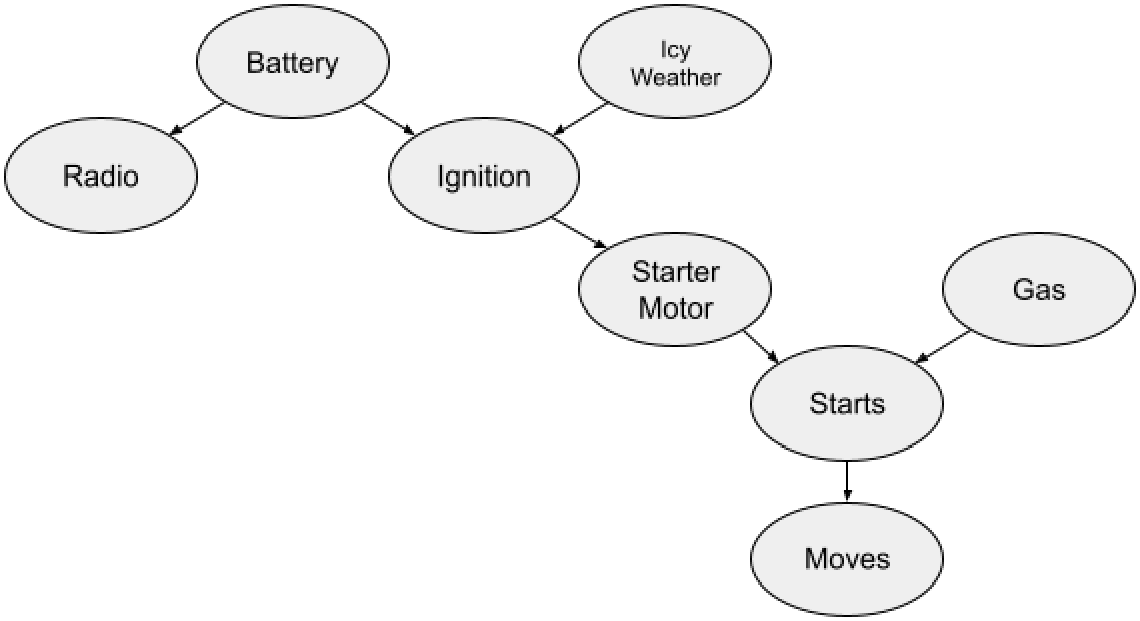
(c) Is the independence assumption reasonable? Discuss.

I would not think that the assumtion is very reasonable. A tax document, for example, probably has a lot of financial jargon that appears together, just a recipt would contain a long list of prices and the term 'sales tax' or 'tip' just as a recipe probabily contains temepratures, times and 'whisk', 'bake', or 'mix' much more than the average document. Given we see one of these words, the probabiliy that other similar words exist should be much higher.

[6 points]

3. Consider the network for car diagnosis shown in figure below.

(a) Extend the network with the Boolean variables *IcyWeather* and *StarterMotor*.



(b) Give reasonable conditional probability tables for all the nodes.

P(Battery) = .98

P(Icy Weather) = .15

P(Gas) = .95

P(Radio) = (B=0): 0 B(B=1): .99

P(Ignition) = (B=0, IW=0): 0 (B=0, IW=1): 0, (B=1, IW = 0): .95, (B=1, IW = 1): .9

P(Starter Motor) = (I = 0): 0, (I = 1): .98

P(Starts) = (SM=0, G=0): 0, (SM=0, G=1): 0, (SM=1, G=0): 0, (SM=1, G=1): .99

P(Moves) = (S=0): 0, (S=1): .95

(c) How many independent values are contained in the joint probability distribution for eight Boolean nodes, assuming no conditional independence relations hold among them?

The entire joint distrbution would be 2^8 -1 spots, or 255 values.

(d) How many independent probability values do your network tables contain?

My network contains 15 values (Battery, IW and Gas are totally independent (+3), then each other node is based on one other node, with the exception of starts which is based on two (+8) (+4)

(e) The conditional distribution for *Starts* could be described as a noisy-AND distribution. Define this family in general and relate it to the noisy-OR distribution.

This describes how you can infer the output of something based on its parents and/or *something else*, that something else being given the name noise. As my probabilities show, not all of them are exactly equal to 1. The probability that the radio starts given the battery works its so close to 1, but there can always be a reason it doesnt actually work, whether its the motherboard or the weather or a radioactive particle flipping a transistor. We compensate for all these randim variables we could never enumerate to exhaustion (not to mention it would massively complicate the model) by adding some “noise”.

In our specific example of the Starts, we can get a slightly better model that predicts Starts based on Stater Motor AND Gas AND *noise*. This makes it a noisy-AND distribution that is highest when the starter motor and gas are true and the noise is high, which we can say is high when its a newer or highly reliable car, or low if the car is old or has a history of starting issues.

In contrast, we can make an OR system work almost the same way, since (A n B) = !A u !B, so we could calcuate start via 1 - !(Starts) !(Gas) !(noise), which would effectively be the same value.

[10 points]

References

[Pea88] Judea Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Elsevier, 1988.